



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE TECNOLOGIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA E CIÊNCIA DE MATERIAIS**  
**DOUTORADO EM ENGENHARIA E CIÊNCIA DE MATERIAIS**

**RAPHAELLA HERMONT FONSECA MURTA**

**PREVISÃO DAS PROPRIEDADES MECÂNICAS DE VERGALHÕES DE AÇO**  
**UTILIZADOS NA CONSTRUÇÃO CIVIL**

**FORTALEZA**

**2022**

RAPHAELLA HERMONT FONSECA MURTA

PREVISÃO DAS PROPRIEDADES MECÂNICAS DE VERGALHÕES DE AÇO  
UTILIZADOS NA CONSTRUÇÃO CIVIL

Tese apresentada ao Curso de Doutorado em Engenharia e Ciência de Materiais do Programa de Pós-Graduação em Engenharia e Ciência de Materiais do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Engenharia e Ciência de Materiais. Área de Concentração: PROPRIEDADES FÍSICAS E MECÂNICAS DOS MATERIAIS

Orientador: Prof. Dr. Elineudo Pinho de Moura

Co-Orientador: Prof. Dr. Guilherme de Alencar Barreto

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

H472p Hermont Fonseca Murta, Raphaella.  
Previsão das propriedades mecânicas de vergalhões de aço utilizados na construção civil  
/ Raphaella Hermont Fonseca Murta. – 2022.  
85 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de  
Pós-Graduação em Engenharia e Ciência de Materiais, Fortaleza, 2022.

Orientação: Prof. Dr. Elineudo Pinho de Moura.

Coorientação: Prof. Dr. Guilherme de Alencar Barreto.

1. Vergalhões de aço. 2. Propriedades Mecânicas. 3. Análise de Regressão Linear. 4.  
Máquina de Aprendizagem Mínima. 5. Rede Neural Artificial. I. Título.

CDD 620.11

---

RAPHAELLA HERMONT FONSECA MURTA

PREVISÃO DAS PROPRIEDADES MECÂNICAS DE VERGALHÕES DE  
AÇOUTILIZADOS NA CONSTRUÇÃO CIVIL

Tese apresentada ao Curso de Doutorado em Engenharia e Ciência de Materiais do Programa de Pós-Graduação em Engenharia e Ciência de Materiais do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutora em Engenharia e Ciência de Materiais.

Área de Concentração: PROPRIEDADES FÍSICAS E MECÂNICAS DOS MATERIAIS

Aprovada em:

BANCA EXAMINADORA

---

Prof. Dr. Elineudo Pinho de Moura (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Guilherme de Alencar Barreto (Co-Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Hamilton Ferreira Gomes de Abreu  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Paulo Rangel Rios  
Universidade Federal Fluminense (UFF)

---

Prof. Dr. André Luiz Vasconcellos da Costa e Silva  
Universidade Federal Fluminense (UFF)



A Deus.

A minha família.

## **AGRADECIMENTOS**

Aos meus pais, Ana Lúcia Fonseca e Breno Hermont Murta, pelo amor, incentivo e apoio incondicional.

Ao meu marido, Otilio Freire, pelo companheirismo e suporte para que eu pudesse desenvolver este projeto.

A minha família, em especial Arthur Murta e Ariana Salim, pelo apoio e companheirismo.

Ao meu orientador, Professor Dr. Elineudo Pinho de Moura, pela oportunidade e apoio no desenvolvimento deste trabalho.

Ao meu co-orientador, Professor Dr. Guilherme de Alencar Barreto, por toda a ajuda no desenvolvimento deste trabalho.

Aos meus professores, em especial Hamilton Ferreira e Francisco Marcondes, pelo incentivo e colaboração.

Aos integrantes do Centro de Ensaios não Destrutivos - CENDE/UFC, Aldecira Gadelha, Pedro Paulo e Fabrício Braga, pela participação na elaboração desta pesquisa.

Aos meus colegas de trabalho, em especial Alex Nascimento, pela contribuição no desenvolvimento dessa tese.

Aos meus amigos, Ana Beatriz, Rafael Silva, Gabi Tomaz, Isabelle Maria, Larissa Almeida, Bárbara Quaiato, Adila Kelley, Nadja Nogueira e Thiago Andrade, por se fazerem presentes na minha vida.

Aos meus colegas de curso, em especial Úrsula Cid, Vinicius Oliveira, Giovani Gonçalves, Pablo Leão e Natan Lima, pela amizade, paciência e convivência durante o doutorado.

A todos que direta ou indiretamente fizeram parte da minha formação, o meu muito obrigada.

”Uma teoria matemática não deve ser considerada completa, até que você a tenha deixado tão clara, de modo que pode ser explicada para o primeiro homem que você encontrar na rua.”

(DAVID HILBERT)

## RESUMO

As propriedades mecânicas do aço dependem fortemente de sua composição química e dos parâmetros empregados durante o processamento termomecânico a qual foi submetido. Entender como cada variável afeta tais propriedades é indispensável à obtenção de produtos siderúrgicos de elevada qualidade com menor custo. Entretanto, o grande número de variáveis envolvidas dificulta esta tarefa. Ferramentas estatísticas combinadas com algoritmos de previsões podem ser utilizadas na identificação dos parâmetros mais relevantes e na obtenção de uma função matemática que consiga descrever adequadamente as propriedades mecânicas do aço a partir dos pares de entrada-saída selecionados. No presente trabalho, informações sobre a composição química e as variáveis do processamento termomecânico foram coletadas em uma siderúrgica e utilizadas na previsão de propriedades mecânicas de vergalhões de aço submetidos a tratamento térmico usando os métodos de estimação: Análise de Regressão Linear (ARL), Máquina de Aprendizagem Mínima (MAM), Rede Neural Artificial (RNA), Máquina de Vetores de Suporte (MVS) e Máquina de Vetores de Suporte por Mínimos Quadrados (MVSMQ). O coeficiente de determinação foi calculado entre os valores observados e preditos para as seguintes propriedades mecânicas: Limite de Escoamento (LE), Limite de Resistência à Tração (LRT), Razão LRT/LE e o Alongamento Percentual (AP). Os resultados estimados pelos algoritmos se mostraram promissores, indicando que MAM e MVSMQ podem ser úteis na avaliação e escolha dos parâmetros mais adequados a serem utilizados durante o processo de produção do aço.

**Palavras-chave:** Vergalhões de aço; Propriedades Mecânicas; Análise de Regressão Linear; Máquina de Aprendizagem Mínima; Rede Neural Artificial; Máquina de Vetores de Suporte; Máquina de Vetores de Suporte por Mínimos Quadrados.

## ABSTRACT

The mechanical properties of steel depend strongly on its chemical composition and the parameters used during submitted thermomechanical processing. Understanding how each variable affects such properties is indispensable for obtaining high-quality steel products at a lower cost. However, the large number of variables involved in the manufacturing process makes this task difficult. It is possible to use statistical tools combined with predictive modeling to identify the most relevant parameters and to obtain a mathematical model that can adequately describe the mechanical properties of the rebar from the selected input-output pairs. In the present work, information about the chemical composition and the variables of thermomechanical processing were collected at steel industry and used to predict the mechanical properties of steel rebar submitted to heat treatment using the methods of estimation: Linear Regression Analysis(LRA), Minimal Learning Machine (MLM), Artificial Neural Networks (ANN), Support-Vector Machine (SVM) and Least-Squares Support-Vector Machines (LSSVM). The determination coefficient was calculated between the observed and predicted values for mechanical properties: Yield Strength (YS), Ultimate Tensile Strength (UTS), UTS/YS ratio, and Percent Elongation (PE). The results estimated by the algorithms were promising, indicating that MLM and LSSVM can be useful in evaluating and choosing the most adequate parameters to be used during the steel production process.

**Keywords:** Steel Rebar; Mechanical Properties; Linear Regression Analysis; Minimal Learning Machine; Artificial Neural Networks; Support-Vector Machine; Least-Squares Support-Vector Machines.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Diagrama de fases ferro-carbono . . . . .	22
Figura 2 – Diagrama de fases de aços hipoeutetóides . . . . .	23
Figura 3 – Diagrama de transformação tempo-temperatura para o aço 1050 . . . . .	24
Figura 4 – Influência do teor de carbono sobre as propriedades mecânicas do aço carbono esfriados lentamente . . . . .	25
Figura 5 – Recristalização durante a laminação a quente . . . . .	29
Figura 6 – Representação da laminação a quente . . . . .	30
Figura 7 – Curvas de variação da temperatura da barra para diferentes posições de raio durante processo de tratamento térmico superficial . . . . .	31
Figura 8 – Diagrama tensão-deformação mostrando a deformação elástica e plástica. O limite de escoamento está indicado pela tensão $\sigma_y$ . . . . .	33
Figura 9 – Diagrama tensão-deformação até o ponto de fratura, F. LRT é o Limite de Resistência à Tração e está indicado pelo ponto M. . . . .	34
Figura 10 – Modelo matemático de um neurônio artificial . . . . .	43
Figura 11 – Função de Limiar . . . . .	44
Figura 12 – Função Linear por Partes . . . . .	44
Figura 13 – Tangente hiperbólica . . . . .	45
Figura 14 – Função sigmóide logística para diferentes valores de $c$ . . . . .	46
Figura 15 – Modelo matemático da rede perceptron de múltiplas camadas . . . . .	47
Figura 16 – Representação gráfica do subajuste, sobreajuste e ajuste ideal de um modelo	48
Figura 17 – Regra da parada antecipada baseada na validação cruzada . . . . .	49
Figura 18 – Exemplo de um problema não linearmente separável . . . . .	51
Figura 19 – Exemplo do mapeamento no $\mathbb{R}^3$ . . . . .	52
Figura 20 – Previsão das propriedades mecânicas pela ARL. . . . .	64
Figura 21 – Previsão da limite de escoamento pela MAM . . . . .	65
Figura 22 – Previsão da limite de resistência à tração pela MAM . . . . .	65
Figura 23 – Previsão da Razão LRT/LE pela MAM . . . . .	66
Figura 24 – Previsão do alongamento percentual pelo MAM . . . . .	66
Figura 25 – Previsão das propriedades mecânicas pela MAM após 100 execuções de treino/teste. . . . .	67
Figura 26 – Previsão das propriedades mecânicas pela RNA. . . . .	73

Figura 27 – Previsão das propriedades mecânicas pela MVS. . . . .	74
Figura 28 – Previsão das propriedades mecânicas pela MVSMQ. . . . .	76
Figura 29 – Comparação geral do desempenho dos modelos avaliados em termos dos valores médios de $R^2$ . . . . .	77
Figura 30 – Comparação geral do desempenho dos modelos avaliados em termos dos valores máximos de $R^2$ . . . . .	78

## LISTA DE TABELAS

Tabela 1 – Exemplo de funções de kernel. . . . .	54
Tabela 2 – Parâmetros da laminação do tratamento térmico do vergalhão do aço. . . . .	59
Tabela 3 – Variáveis de entrada relacionadas à composição química [%].. . . .	60
Tabela 4 – Propriedades mecânicas do vergalhão previstas pelos modelos. . . . .	60
Tabela 5 – Análise de regressão linear para as propriedades mecânicas do vergalhão após 1000 execuções de treino/teste. . . . .	62
Tabela 6 – Estatísticas descritivas do desempenho da MAM após 100 execuções de treinamento/teste para dois valores diferentes do hiperparâmetro $K$ (número de pontos de referência). Os melhores resultados são mostrados em negrito. . . . .	67
Tabela 7 – Sumário dos resultados obtidos pela rede neural artificial para o LE classificado pelo $R^2$ máximo após 100 execuções de treino/teste. . . . .	69
Tabela 8 – Sumário dos resultados obtidos pela rede neural artificial para o LRT classificado pelo $R^2$ máximo após 100 execuções de treino/teste. . . . .	69
Tabela 9 – Sumário dos resultados obtidos pela rede neural artificial para a razão do LRT/LE classificado pelo $R^2$ máximo após 100 execuções de treino/teste. . . . .	69
Tabela 10 – Sumário dos resultados obtidos pela rede neural artificial para o AP classificado pelo $R^2$ máximo após 100 execuções de treino/teste. . . . .	69
Tabela 11 – Sumário dos resultados obtidos pela rede neural artificial para o limite de escoamento classificado pelo $R^2$ médio após 100 execuções de treino/teste. . . . .	70
Tabela 12 – Sumário dos resultados obtidos pela rede neural artificial para o Limite de Resistência à tração classificado pelo $R^2$ médio após 100 execuções de treino/teste. . . . .	70
Tabela 13 – Sumário dos resultados obtidos pela rede neural artificial para a razão do LRT/LE classificado pelo $R^2$ médio após 100 execuções de treino/teste, . . . . .	71
Tabela 14 – Sumário dos resultados obtidos pela rede neural artificial para o alongamento percentual classificado pelo $R^2$ médio após 100 execuções de treino/teste. . . . .	71
Tabela 15 – Estatísticas descritivas do desempenho da MVS após 100 execuções de treino/teste. . . . .	72
Tabela 16 – Estatísticas descritivas do desempenho da MVSMQ após 100 execuções de treinamento/teste. . . . .	75



Tabela 17 – Porcentagem dos dados de teste que apresentaram um erro médio menor do que 5% e 10% . . . . .	79
---	----

## LISTA DE ALGORITMOS

Algoritmo 1 – Análise de Regressão Linear + Teste de Significância . . . . .	40
Algoritmo 2 – Máquina de Aprendizagem Mínima . . . . .	42
Algoritmo 3 – Redes Neurais Artificiais . . . . .	50
Algoritmo 4 – Máquina de Vetores de Suporte . . . . .	54
Algoritmo 5 – Máquina de Vetores de Suporte por Mínimos Quadrados . . . . .	57

## LISTA DE ABREVIATURAS E SIGLAS

<i>R</i> <sup>2</sup>	Coeficiente de determinação
AP	Alongamento Percentual
ARL	Análise de Regressão Linear
CCC	Cúbica de Corpo Centrado
CFC	Cúbica de Face Centrada
LE	Limite de Escoamento
LRT	Limite de Resistência à Tração
MAM	Máquina de Aprendizagem Mínima
MVS	Máquina de Vetores de Suporte
MVSMQ	Máquina de Vetores de Suporte por Mínimos Quadrados
RNA	Rede Neural Artificial
Sgm	Sigmoide Logística
Tanh	Tangente Hiperbólica

## LISTA DE SÍMBOLOS

$a_{jj}$	j-ésimo elemento diagonal da matriz $(X'X)^{-1}$
a,b e c	Hiperparâmetros do modelo
Al	Alumínio
$\alpha$	Ferrita
$\alpha_{ap}$	Taxa de aprendizado
$\alpha_n, \alpha_n^*$	Multiplicadores positivos de Lagrange alpha
$\eta_n, \eta_n^*$	Multiplicadores positivos de Lagrange eta
$b$	Matriz dos coeficientes da regressão linear
$\beta_n$	n-ésimo coeficiente da regressão linear
C	Carbono
Cr	Cromo
Cu	Cobre
$c_0$	Comprimento de referência do material
$c_f$	Comprimento final do material
$\delta_k$	Gradiente Local
$\delta_k(y)$	Matriz de distância de y aos pontos de referência k
$\Delta_k(x)$	Matriz de distância de x aos pontos de referência k
E	Módulo de Elasticidade
$E_m$	Número de épocas máximo
$\varepsilon$	Deformação
$\varepsilon_m$	Matriz dos erros associados a cada dado estimado
Fe	Ferro
$\gamma$	Austenita
$J(y)$	Função objetivo
K	Função de Kernel
Mn	Manganês

Mo	Molibdênio
Nb	Nióbio
Ni	Níquel
O	Oxigênio
P	Fósforo
$\phi$	Função de ativação
$\phi'_k$	Derivada da função de ativação
$\Phi(x)$	Função de transformação de $\mathbb{R}^m \rightarrow \mathbb{R}^q$ ( $q > n$ )
S	Enxofre
Si	Silício
$\sigma$	Tensão
$\sigma^2$	Variância do erro
$s^2$	Variância amostral do erro
$\sigma_y$	Tensão Elástica
$t$	Distribuição da variável aleatória
Ti	Titânio
V	Vanádio
$w$	Pesos sinápticos
$X$	Matriz com a primeira coluna igual a 1 e as outras com os dados da regressão
$X_{test}$	Matriz com os dados de teste
$X_{tre}$	Matriz com os dados de treinamento
$\xi_n, \xi_n^*$	Variáveis de folga
$y_{ip}$	Resposta de previsão do i-ésimo dado
$y_p$	Matriz das variáveis preditas
$Y_{pred}$	Matriz com os valores previstos
$y_k^o$	Saída observada
$W^T$	Coefficientes da regressão linear da MVS
Zn	Zinco

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>18</b>
<b>1.1</b>	<b>Objetivo geral</b>	<b>19</b>
<b>1.2</b>	<b>Objetivos específicos</b>	<b>19</b>
<b>1.3</b>	<b>Produção científica</b>	<b>20</b>
<b>1.4</b>	<b>Organização do restante da tese</b>	<b>20</b>
<b>2</b>	<b>FABRICAÇÃO E PROPRIEDADES MECÂNICAS DO AÇO CARBONO</b>	<b>21</b>
<b>2.1</b>	<b>Aço Carbono</b>	<b>21</b>
<b>2.2</b>	<b>Elementos presentes no aço</b>	<b>24</b>
<b>2.2.1</b>	<i>Carbono</i>	<b>25</b>
<b>2.2.2</b>	<i>Manganês</i>	<b>26</b>
<b>2.2.3</b>	<i>Níquel</i>	<b>26</b>
<b>2.2.4</b>	<i>Molibdênio</i>	<b>26</b>
<b>2.2.5</b>	<i>Nióbio</i>	<b>26</b>
<b>2.2.6</b>	<i>Fósforo</i>	<b>27</b>
<b>2.2.7</b>	<i>Silício</i>	<b>27</b>
<b>2.2.8</b>	<i>Cobre</i>	<b>27</b>
<b>2.2.9</b>	<i>Vanádio</i>	<b>27</b>
<b>2.2.10</b>	<i>Enxofre</i>	<b>28</b>
<b>2.2.11</b>	<i>Cromo</i>	<b>28</b>
<b>2.2.12</b>	<i>Estanho</i>	<b>28</b>
<b>2.3</b>	<b>Processo de laminação</b>	<b>28</b>
<b>2.4</b>	<b>Tratamento térmico de vergalhões de aço</b>	<b>30</b>
<b>2.5</b>	<b>Propriedades mecânicas</b>	<b>32</b>
<b>2.5.1</b>	<i>Limite de Escoamento</i>	<b>32</b>
<b>2.5.2</b>	<i>Limite de Resistência à Tração</i>	<b>33</b>
<b>2.5.3</b>	<i>Razão LRT/LE</i>	<b>34</b>
<b>2.5.4</b>	<i>Alongamento Percentual</i>	<b>35</b>
<b>2.6</b>	<b>Resumo do Capítulo</b>	<b>35</b>
<b>3</b>	<b>MÉTODOS DE ESTIMAÇÃO E CONJUNTO DE DADOS</b>	<b>36</b>
<b>3.1</b>	<b>Métodos de estimação</b>	<b>36</b>

3.1.1	<i>Análise de Regressão Linear</i> . . . . .	36
3.1.2	<i>Máquina de Aprendizagem Mínima</i> . . . . .	39
3.1.3	<i>Redes Neurais Artificiais</i> . . . . .	41
3.1.4	<i>Máquinas de Vetores de Suporte</i> . . . . .	49
3.1.5	<i>Máquinas de Vetores de Suporte por Mínimos Quadrados</i> . . . . .	55
3.2	<b>Coefficiente de determinação - <math>R^2</math></b> . . . . .	57
3.3	<b>Conjunto de dados</b> . . . . .	58
3.4	<b>Resumo do Capítulo</b> . . . . .	60
4	<b>RESULTADOS E DISCUSSÕES DAS SIMULAÇÕES NUMÉRICAS</b> .	61
4.1	<b>Análise de Regressão Linear</b> . . . . .	61
4.2	<b>Máquina de Aprendizagem Mínima</b> . . . . .	63
4.3	<b>Rede neural de múltiplas camadas</b> . . . . .	66
4.4	<b>Máquinas de Vetores de Suporte</b> . . . . .	72
4.5	<b>Máquinas de Vetores de Suporte por Mínimos Quadrados</b> . . . . .	75
4.6	<b>Considerações Finais dos Resultados</b> . . . . .	77
4.7	<b>Resumo do Capítulo</b> . . . . .	79
5	<b>CONCLUSÃO</b> . . . . .	80
	<b>REFERÊNCIAS</b> . . . . .	82

## 1 INTRODUÇÃO

O funcionamento do mundo moderno tal qual conhecemos é extremamente dependente do aço. O aço está presente de diversas formas no nosso dia a dia: no setor automobilístico, na construção civil, em utensílios domésticos, entre outras. Cerca de 1.878 milhões de toneladas de aço foram produzidas no mundo em 2020 movimentando bilhões de dólares na economia. (ASSOCIATION, 2021)

A produção do aço pode ocorrer através de duas rotas: a integrada e a semi-integrada. As usinas integradas operam em três fases básicas: redução, refino e lingotamento, tendo como o seu principal insumo o minério de ferro. Já as usinas semi-integradas operam em duas fases básicas: refino e lingotamento, e utilizam sucata metálica, ferro gusa ou ferro esponja como insumo básico. (ARAÚJO, 2021)

Em usinas siderúrgicas semi-integradas o processo de refino ocorre em fornos que usam a eletricidade como fonte de energia para a fusão e refino do metal. Na grande maioria dessas usinas os Fornos Elétricos a Arco (FEAs) são utilizados para fundir sucatas de aço e transformá-las em aço líquido (YILDIRIM; PREZZI, 2017). Em seguida, o aço líquido é vazado para o forno panela onde ocorre a adição de elementos de liga e remoção de impurezas. Neste processo, o controle da composição química do aço recebe atenção especial para que os produtos possam atingir as propriedades mecânicas desejadas no final do processo de fabricação (SAMPAIO *et al.*, 2009). O aço é então solidificado em equipamentos de lingotamento contínuo para produzir produtos semi-acabados, tais como tarugos, blocos, lingotes e placas.

Algumas siderúrgicas possuem ainda o processo de laminação para produzirem produtos acabados. Nesta etapa o aço é processado por equipamentos chamados laminadores que reduzem a seção transversal por compressão do metal e promovem a recristalização do material, exercendo um grande impacto na resistência e elasticidade do produto laminado (BOËR *et al.*, 2012). Em algumas usinas, após o processo de laminação pode ocorrer um tratamento térmico adicional com o objetivo de melhorar as propriedades mecânicas do aço. Além da composição química, outros fatores de extrema importância para as propriedades mecânicas finais do aço são os parâmetros de processo da laminação e do tratamento térmico ao qual será submetido (ARSHADI *et al.*, 2020). Desta forma, a escolha adequada da matéria-prima e dos parâmetros do processo de laminação são muito importantes para obter boas propriedades mecânicas do aço. Portanto, o controle dos insumos aliado aos parâmetros apropriados que proporcionam aços de maior qualidade e menor custo é primordial para uma indústria siderúrgica se manter competitiva



(WORRELL, 2011).

Modelos matemáticos estão sendo utilizados para prever propriedades mecânicas de produtos siderúrgicos a partir da composição química do aço e dos parâmetros do processo. A análise de regressão tem sido um método comum e eficaz para alcançar esse objetivo (GIGOVIĆ-GEKIĆ *et al.*, 2014). A aplicação de métodos de estimação provocou um grande avanço na formulação de modelos de predição mais robustos (JUUTILAINEN *et al.*, 2003). Ghaisari *et al.* (2012) previram as propriedades mecânicas de produtos laminados a frio a partir de uma ampla variedade de parâmetros de todas as etapas de produção por redes neurais artificiais. Yang *et al.* (2009) desenvolveram a rede neural Bayesiana para prever a resistência à tração de produtos laminados a quente a partir de composições de aço e variáveis críticas do processo de laminação. Belayadi e Bourahla (2019) usaram um modelo de rede neural para prever e calcular as propriedades mecânicas das ligas Al-Zn a partir de sua composição química.

No presente trabalho informações sobre a composição química do aço e das variáveis do processamento termomecânico foram coletadas em uma usina siderúrgica e utilizadas como variáveis preditoras de quatro importantes propriedades mecânicas de vergalhões de aço: Limite de escoamento (LE), Limite de Resistência à Tração (LRT), Razão LRT/LE e Alongamento Percentual (AP). Cinco técnicas preditivas foram utilizadas para prever as propriedades mecânicas: Análise de Regressão Linear (ARL), Máquina de Aprendizagem Mínima (MAM), Rede Neural Artificial (RNA), Máquina de Vetores de Suporte (MVS) e Máquina de Vetores de Suporte por Mínimos Quadrados (MVSMQ).

## 1.1 Objetivo geral

O objetivo deste trabalho é avaliar o emprego de métodos de estimação na predição de importantes propriedades mecânicas de vergalhões de aço utilizados na construção civil.

## 1.2 Objetivos específicos

1. Processamento e análise de dados de composição química e das variáveis do processamento termomecânico através do uso das ferramentas de previsão.
2. Avaliar as variáveis que mais influenciam nas quatro propriedades mecânicas finais do vergalhão: Limite de Escoamento, Limite de Resistência à Tração, Razão LRT/LE e Alongamento Percentual.

3. Avaliar o desempenho da Regressão Linear, Máquina de Aprendizagem Mínima, Rede Neural Artificial, Máquina de Vetores de Suporte e Máquina de Vetores de Suporte por Mínimos Quadrados na previsão das propriedades mecânicas do aço.

### **1.3 Produção científica**

Os resultados obtidos utilizando a ARL e a RNA foram publicados no artigo "Mathematical modelling for predicting mechanical properties in rebar manufacturing"(MURTA *et al.*, 2021). Já os resultados obtidos utilizando a MAM, a MVS e a MVSMQ irão ser submetidos no artigo "Mechanical properties prediction in rebar using kernel-based regression models".

### **1.4 Organização do restante da tese**

Os capítulos 2 e 3 apresentam alguns conceitos importantes utilizados nesse trabalho. O capítulo 2 traz informações sobre fabricação e propriedades mecânicas do aço carbono. Já o capítulo 3, aborda conceitos sobre as técnicas de estimação e os dados utilizados. Em seguida, o capítulo 4 traz os resultados e as discussões obtidas. Por fim, o capítulo 5 apresenta a conclusão da tese e propostas para trabalhos futuros.

## 2 FABRICAÇÃO E PROPRIEDADES MECÂNICAS DO AÇO CARBONO

O presente capítulo começa apresentando os conceitos básicos do aço carbono e a influência dos seus principais elementos de liga em sua propriedade mecânica. Então, fundamentos de processos de laminação e do tratamento térmico do vergalhão são abordados. Em seguida, uma breve explicação das propriedades mecânicas (Limite de Escoamento, Limite de Resistência à Tração, Razão LRT/LE e Alongamento Percentual) estudadas nesse trabalho são apresentadas.

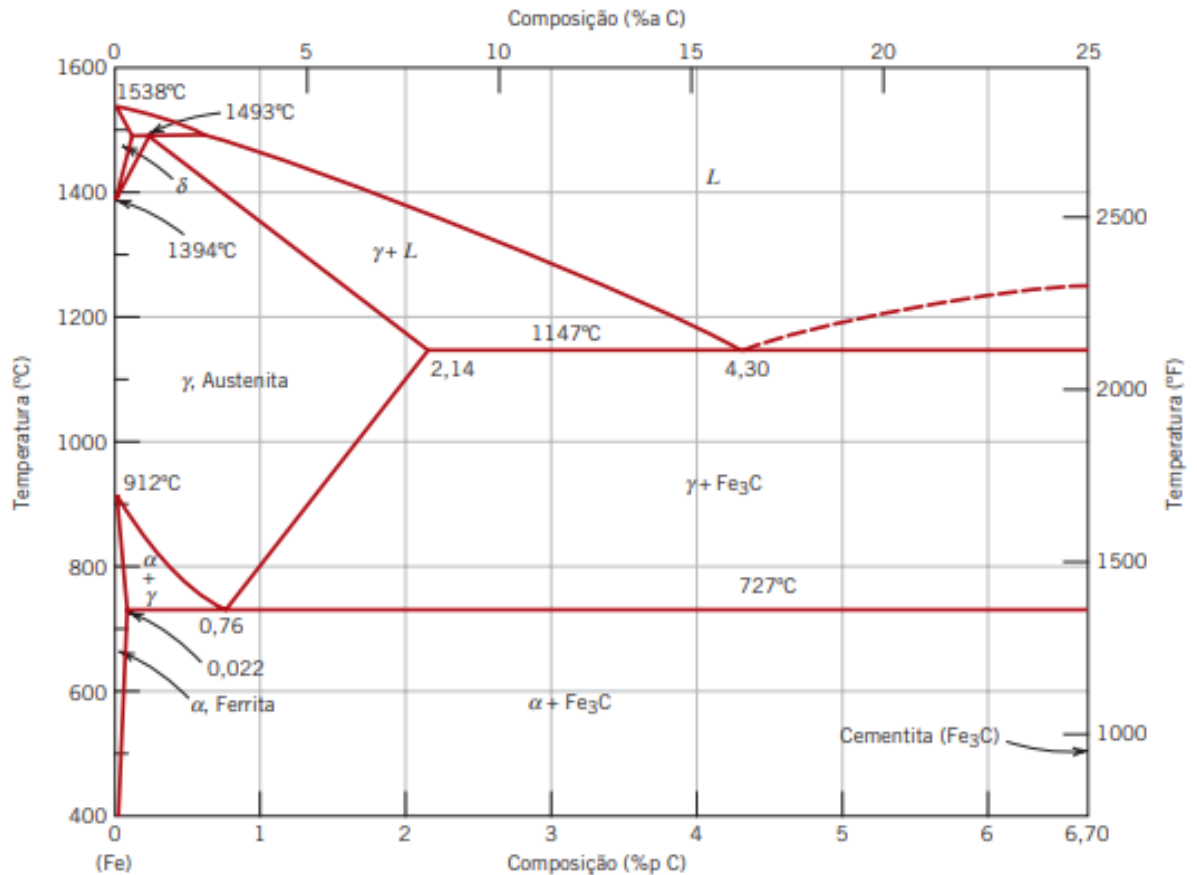
### 2.1 Aço Carbono

O aço carbono é uma liga metálica formada principalmente de ferro e carbono, contendo percentual entre 0,008% e 2,11% carbono, podendo conter outros elementos secundários como silício, manganês, magnésio, cromo e outros (LAURITO, 2010). Devido suas propriedades mecânicas, o aço é bastante utilizado no setor automobilístico, na construção civil e eletrodomésticos (BLACK; KOHSER, 2020).

A máxima solubilidade do carbono no ferro à temperatura ambiente é de 0,008% e a solubilidade máxima do carbono é de 2,11%, que ocorre a 1148°C (CHIAVERINI, 1977). Dependendo do teor de carbono no aço e da temperatura, diferentes estruturas e propriedades são obtidas em equilíbrio, como pode ser visto no diagrama do ferro-carbono apresentado na Figura 1.

A região de interesse do diagrama ferro-carbono para a produção de barras de aço é a de teor de carbono de até 0,76%, classificados como aços hipoeutetóides (ARAÚJO, 2010). Considere o resfriamento, lento (em condições de equilíbrio termodinâmico), de uma liga hipoeutetóide, que inicia no ponto “*c*” até o ponto “*f*”, apresentado na Figura 2. A transformação metalúrgica se inicia no ponto *c* (Figura 2) com a presença da fase austenita ( $\gamma$ ). Constituída por uma solução sólida intersticial de carbono no ferro com estrutura Cúbica de Face Centrada (CFC), possui propriedade não magnética, boa resistência mecânica e tenacidade (CALLISTER; RETHWISCH, 2007). A medida que a temperatura do aço diminui, pequenos grãos de ferrita ( $\alpha$ ) começam a aparecer. Os grãos se formam no contorno de grão da austenita, ver ponto *d* na Figura 2. Em temperaturas menores, maior fração de ferrita é formada (ver ponto *e*). A ferrita apresenta uma estrutura Cúbica de Corpo Centrado (CCC), com alto alongamento percentual e baixa resistência à tração.

Figura 1 – Diagrama de fases ferro-carbono

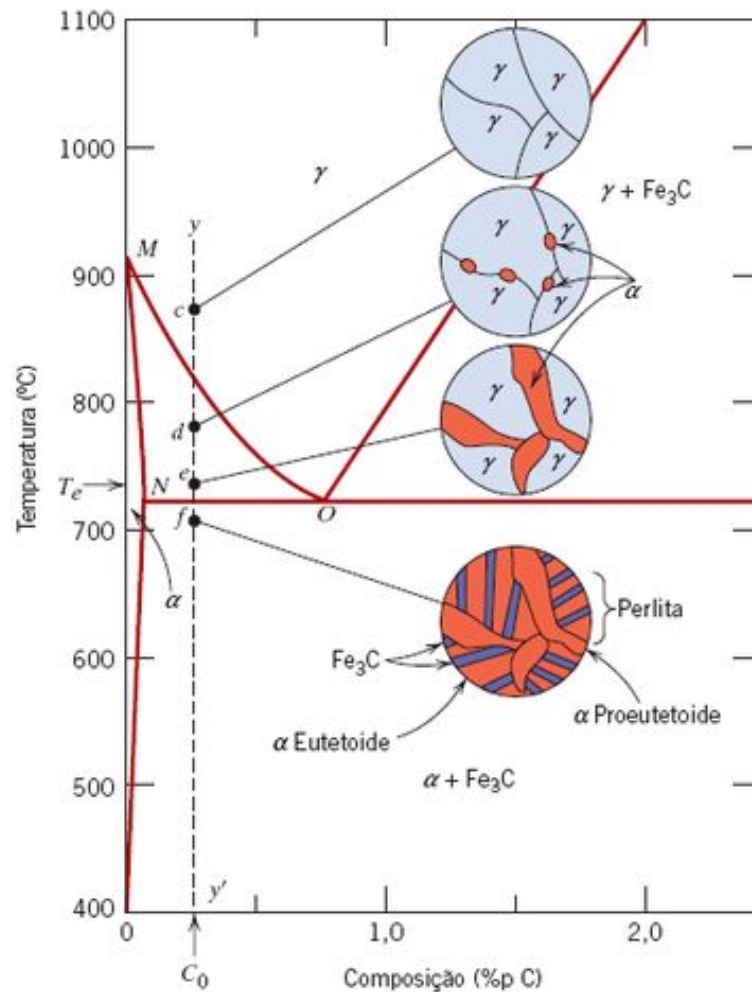


Fonte: Callister e Rethwisch (2007)

Ao resfriar abaixo de 727°C (linha NO) a austenita se transforma em um microconstituente eutetóide denominado perlita, composto por lamelas alternadas de ferrita, uma fase tenaz formada por solução de carbono intersticial na matriz do ferro cúbico de corpo centrado, e cementita, um carboneto de ferro ( $Fe_3C$ ) que apresenta estrutura ortorrômbica, duro e quebradiço, que eleva a dureza e resistência dos aços alto carbono. No ponto *f*, toda a austenita se transforma em perlita, lâminas de ferrita e cementita dispostas alternadamente, que possui as propriedades mecânicas intermediárias entre seus constituintes (CALLISTER; RETHWISCH, 2007).

Os diagramas de equilíbrio apresentados nas Figuras 1 e 2 são para condições de resfriamento lento (CALLISTER; RETHWISCH, 2007). Com o aumento da taxa de resfriamento do aço as taxas de difusão do carbono dissolvido na austenita diminuem, produzindo uma microestrutura bainítica, e podendo chegar a ocorrer transformações adifusionais e originar uma fase metaestável martensítica. Se a velocidade de resfriamento for muito elevada, irá ocorrer formação da martensita, senão, a bainita. A Figura 3 ilustra a variação da microestrutura dependendo da taxa de resfriamento para o aço 1050, onde A é austenita, F é ferrita, P é perlita,

Figura 2 – Diagrama de fases de aços hipoeutetóides

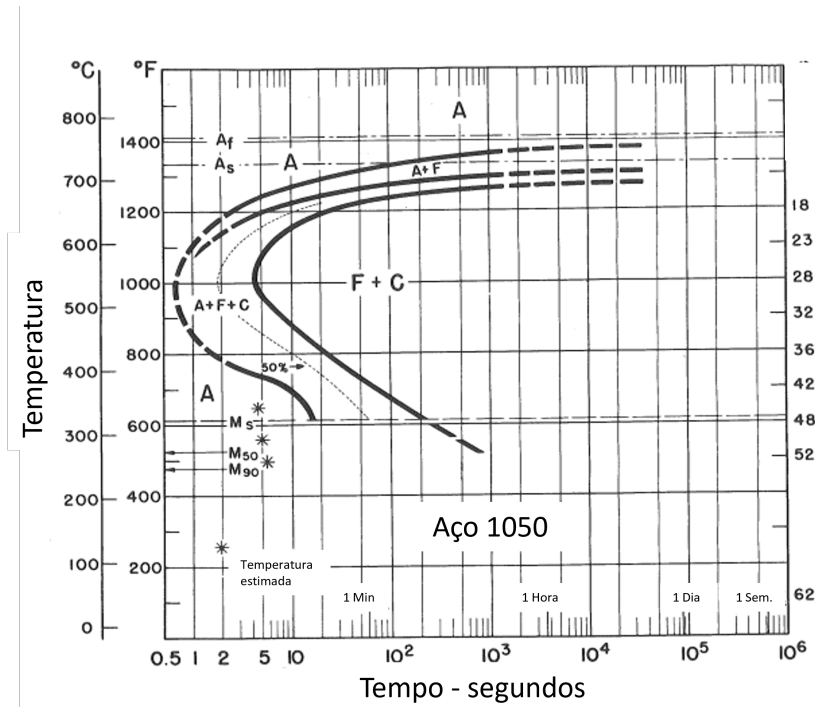


Fonte: Callister e Rethwisch (2007)

B é bainita e M é martensita. Durante o resfriamento, para ocorrer a produção de martensita é necessário que o aço atinja a temperatura de início de transformação Martensítica ( $M_s$ , ver Figura 3), assim um % de Austenita será transformado em Martensita. Se a temperatura  $M_s$  não for atingida nenhuma martensita será formada.

Assim como a perlita, a bainita também é formada por ferrita e cementita. Entretanto, são arranjadas de forma não lamelar cuja as características são dependentes da composição da liga e da temperatura de transformação. A bainita alinha elevada resistência e boa tenacidade. Já na martensita, a rápida queda de temperatura impede que os átomos de carbono se difundam para fora da estrutura cristalina dos átomos de ferro, resultando apenas em uma ligeira distorção de forma destas estruturas (CALLISTER; RETHWISCH, 2007). A martensita é extremamente rígida, mas muito frágil. O tratamento térmico para produção de martensita é chamado de têmpera.

Figura 3 – Diagrama de transformação tempo-temperatura para o aço 1050



Fonte: Adaptado de Metals (1977)

Visando obter propriedades mecânicas adequadas, após a têmpera o aço pode passar por um tratamento térmico denominado revenimento, no qual ele é reaquecido e, então, a martensita, metaestável, transforma-se em ferrita e cementita, originando a martensita revenida, mais resistente e maleável que a martensita (CALLISTER; RETHWISCH, 2007).

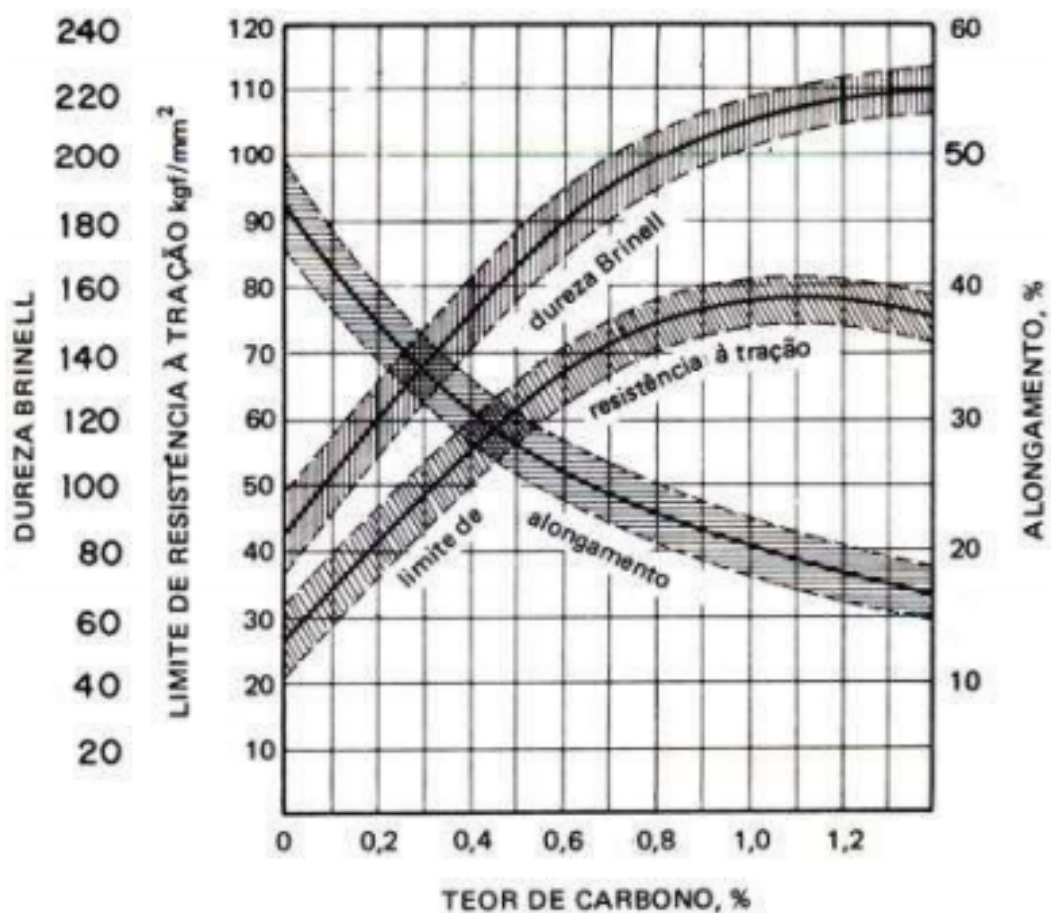
## 2.2 Elementos presentes no aço

Além do ferro e do carbono, outros elementos podem fazer parte da composição química do aço. Alguns deles são residuais, inerentes do processo de fabricação, e outros são introduzidos em sua composição para se obter propriedades mecânicas desejáveis. Entre os principais elementos presentes no aço, destacam-se: Carbono, Manganês, Nióbio, Fósforo, Silício, Molibdênio, Cobre, Vanádio, Enxofre, Cromo, Níquel e Estanho. Uma breve explicação da influência destes elementos nas propriedades mecânicas do aço estão apresentados nesta seção.

### 2.2.1 Carbono

O percentual de carbono é um dos fatores com maior influência nas propriedades mecânicas dos aços. Em geral, quanto maior é o teor de carbono maior é a resistência mecânica à tração e menor é o alongamento percentual do aço, como observado na Figura 4 (CHIAVERINI, 1977).

Figura 4 – Influência do teor de carbono sobre as propriedades mecânicas do aço carbono esfriados lentamente



Fonte: Chiaverini (1977)

O teor de carbono também influencia na transformação martensítica. O aumento do percentual de carbono causa um retardo nas transformações por difusão e diminui a temperatura de início de transformação martensítica.

### **2.2.2 Manganês**

O manganês é um elemento de liga muito importante na fabricação do aço. Além de sua função desoxidante, ele favorece a formação de carbonetos, ajudando a criar microestrutura dura por meio da têmpera (LIMA *et al.*, 2017). O manganês aumenta a resistência mecânica e a dureza do aço. De acordo com Black e Kohser (2020), o manganês combina-se com o enxofre prevenindo a fragilidade a quente. Na ausência do manganês, o enxofre combina-se com o ferro formando sulfeto de ferro, que no final da solidificação é prejudicial ao aço, podendo provocar o sucateamento de todo o material (GUIMARÃES; SOUZA, 2015).

### **2.2.3 Níquel**

O níquel favorece a estabilização da microestrutura austenítica (CARVALHO *et al.*, 2010), com isso teores corretos deste elemento proporciona o equilíbrio entre as fases. O níquel possui alta resistência à corrosão, boa condutividade térmica e elétrica a elevada temperatura e excelente propriedade magnética. É utilizado na fabricação do aço inoxidável e aços ligados.

### **2.2.4 Molibdênio**

O molibdênio pode formar carbonetos complexos na austenita e na ferrita (CHI-AVERINI, 1977), sendo responsável pelo aumento da temperatura de crescimento dos grãos austeníticos. Este aumenta a temperabilidade e reduz as temperaturas de têmpera. Adicionalmente, o molibdênio melhora a dureza a quente, a resistência mecânica e a ductilidade. É utilizado como um substituto do tungstênio, já que cerca da metade da quantidade do molibdênio em relação ao tungstênio produz resultados parecidos ao aço.

### **2.2.5 Nióbio**

De acordo com Chiaverini (1977), pequenos teores de nióbio elevam o limite de escoamento do material e o limite de resistência do aço (em menores proporções). No aço inoxidável, o nióbio forma o carboneto de nióbio e impede a formação do carboneto de cromo, composto prejudicial a alguns aços.



### 2.2.6 Fósforo

É um elemento responsável pelo aumento da resistência mecânica e diminuição da ductilidade e da tenacidade. Em aços com baixo teor de carbono, pode ser empregado para aprimorar as características de usinabilidade e resistência à corrosão atmosférica (CHIAVERINI, 1977). O fósforo tem a tendência de formar compostos estáveis com o ferro devido sua eletropositividade (PALMA, 1985), não sendo utilizado na fabricação de aços com alto teor de carbono devido fragilidade que provoca.

### 2.2.7 Silício

A principal função do silício no aço é de agente desoxidante. Quando adicionado ao aço líquido ele se combina com o oxigênio formando um composto sólido, evitando a formação de bolhas no aço através da reação química do oxigênio com o carbono ( $CO$  e  $CO_2$ ) (CHIAVERINI, 1977). A adição desse elemento aumenta a resistência mecânica e a oxidação a temperaturas elevadas, porém se dissolve na ferrita e tende a decompor cementita ( $Fe_3C$ ). É utilizado para aumentar a temperabilidade de aços que contém elementos não grafitizantes e a dureza de aços com baixo teor de liga.

### 2.2.8 Cobre

O cobre melhora consideravelmente a resistência a corrosão do aço. De acordo com Chiaverini (1977), a adição de 0,25% de cobre aumenta em duas vezes a resistência de um aço sem cobre. É um elemento responsável também pelo acréscimo na resistência mecânica e uma pequena diminuição da ductilidade. O cobre pode ocasionar fragilidade a quente na fabricação do aço, provocando trincas no lingotamento contínuo, laminação a quente e soldagem (SANTOS, 2017). Adicionando níquel a fragilidade a quente pode ser evitada.

### 2.2.9 Vanádio

O vanádio forma carbonetos estáveis que impede o crescimento do grão e contribui para a dureza a quente do aço. A microadição de vanádio ao aço causa precipitações de partículas finas de  $V_4C_3$  na matriz de ferrita, levando ao endurecimento secundário (CHIAVERINI, 1977).

### **2.2.10 Enxofre**

O enxofre é um elemento considerado indesejável para a qualidade do aço. O seu teor deve ser controlado para prevenir a formação de FeS que causa a fragilidade a quente. Este controle é feito pela adição do manganês e consequente formação de MnS em vez de FeS. Em alguns aços o enxofre pode ser adicionado para aumentar a sua usinabilidade.

### **2.2.11 Cromo**

O cromo tem grande influência na resistência à corrosão do aço, medida que o teor de cromo aumenta a suscetibilidade do aço à corrosão diminui (CHIAVERINI, 1977). Adicionalmente, medida que aumenta o teor de cromo ocorre uma redução da tendência de oxidação a temperaturas elevadas.

### **2.2.12 Estanho**

É um elemento resistente à corrosão, sendo bastante utilizado como revestimento em equipamentos de manuseio e embalagem de alimento, possui proteção não tóxica e decorativa. Entretanto, como elemento de liga, o estanho pode tornar o aço suscetível a fragilidade à quente e a fragilidade de revenido.(CHIAVERINI, 1977).

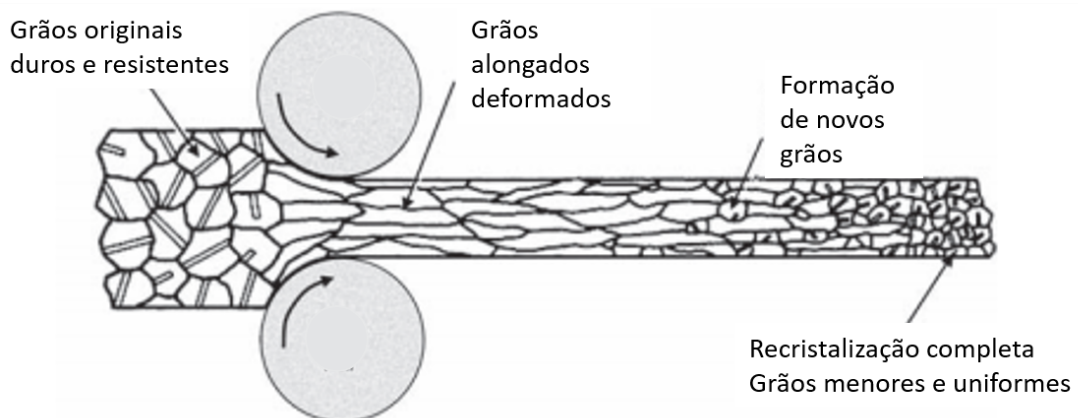
## **2.3 Processo de laminação**

O processo de conformação plástica modifica o metal em materiais com formas definidas através do uso de ferramentas ou matrizes. Podemos dividi-lo em dois grupos: os processos mecânicos, nos quais alterações na forma do produto são causadas pela aplicação de tensões externas e, às vezes, altas temperaturas, mas sem fundi-lo; e processos metalúrgicos, nos quais as modificações de forma podem estar relacionadas a tensões externas e altas temperaturas, mas com difusão de partículas de metal (como na sinterização) ou fundição. Os processos mecânicos consistem na etapa de deformação plástica do material, na qual as tensões aplicadas são geralmente inferiores à tensão de ruptura do material, e na etapa de usinagem, para a qual as tensões aplicadas são superiores à tensão de ruptura, obtendo-se assim a forma final através da remoção de metal (cavaco). Devido à sua natureza, os dois processos também são chamados de processos de formação mecânica (BOLJANOVIC, 2010).

O processo de laminação é um processo de conformação mecânica que consiste essencialmente na passagem de um corpo sólido entre dois cilindros que giram na mesma velocidade periférica, mas em direções opostas. Assim, como o corpo inicial tem uma dimensão maior que a distância entre as superfícies dos cilindros, sofre uma deformação plástica na passagem entre os cilindros, o que resulta na redução de sua seção transversal e no aumento de seu comprimento. Para obter a espessura desejada, a peça deve ser submetida a passagens sucessivas pelos cilindros, diminuindo as distâncias entre eles. A passagem da peça pelos cilindros ocorre pela ação da força de atrito que atua na superfície de contato entre as peças e os cilindros (SHIGAKI *et al.*, 2019).

O processo de laminação pode ser a frio ou a quente: é chamado de laminação a quente quando a temperatura do metal está acima da temperatura de recristalização, ao contrário, é chamado de laminação a frio. A temperatura de recristalização é a menor temperatura na qual os grãos livres de deformação aparecem. Ao aquecer o metal acima da temperatura de recristalização, assumimos que os grãos deformados, duros e resistentes, serão substituídos gradualmente por novos grãos dúcteis e sem defeitos. Isso pode reduzir a resistência a deformação plástica em cada passagem, evitando o endurecimento das passagens subsequentes, como ocorre na laminação a frio. Os grãos recristalizados são formados pelo crescimento de subgrãos selecionados na microestrutura deformada e recuperada, como está sendo ilustrado na Figura 5 (CAMPBELL, 2008). A laminação a quente é comumente aplicada em operações iniciais (operações de desbaste), onde grandes reduções de seção transversal são necessárias. O produto de laminação é utilizado em uma ampla variedade de aplicações e finalidades com diferentes geometrias; por exemplo: aço estrutural, barra e trilhos (ROBERTS, 1983).

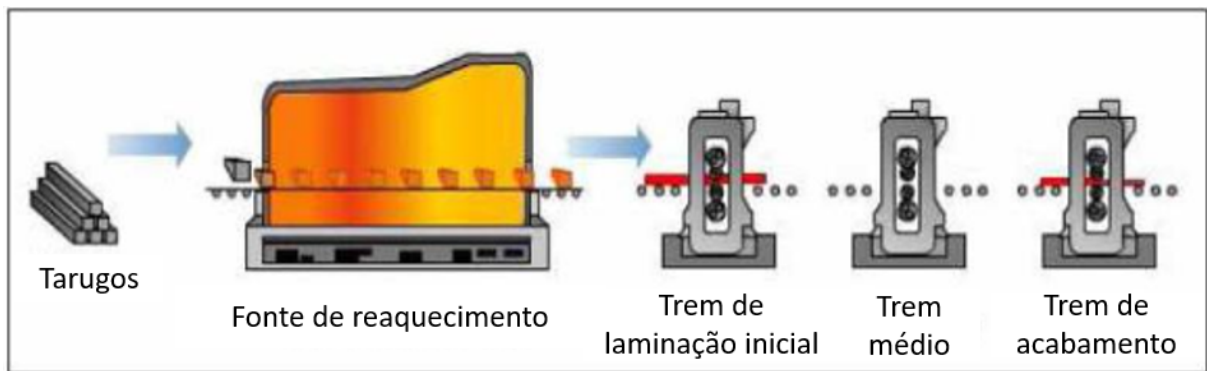
Figura 5 – Recristalização durante a laminação a quente



Fonte: Adaptado de Campbell (2008)

O laminador pode ser dividido em quatro zonas: forno de reaquecimento, trem de laminação inicial (zona de desbaste), trem médio e trem de acabamento, conforme ilustrado na Figura 6. O forno de reaquecimento visa atingir temperaturas ligeiramente superiores à temperatura de recristalização do material. O primeiro passe de rolamento ocorre no chamado trem inicial, iniciando a conformação do metal usando uma gaiola de rolamento. No trem médio, que usa até 4 gaiolas rolantes, ocorre a maior redução da área. Por fim, no trem de acabamento, que utiliza até 6 gaiolas, há pequenas reduções e visa dar as características finais do metal (ROBERTS, 1983).

Figura 6 – Representação da laminação a quente



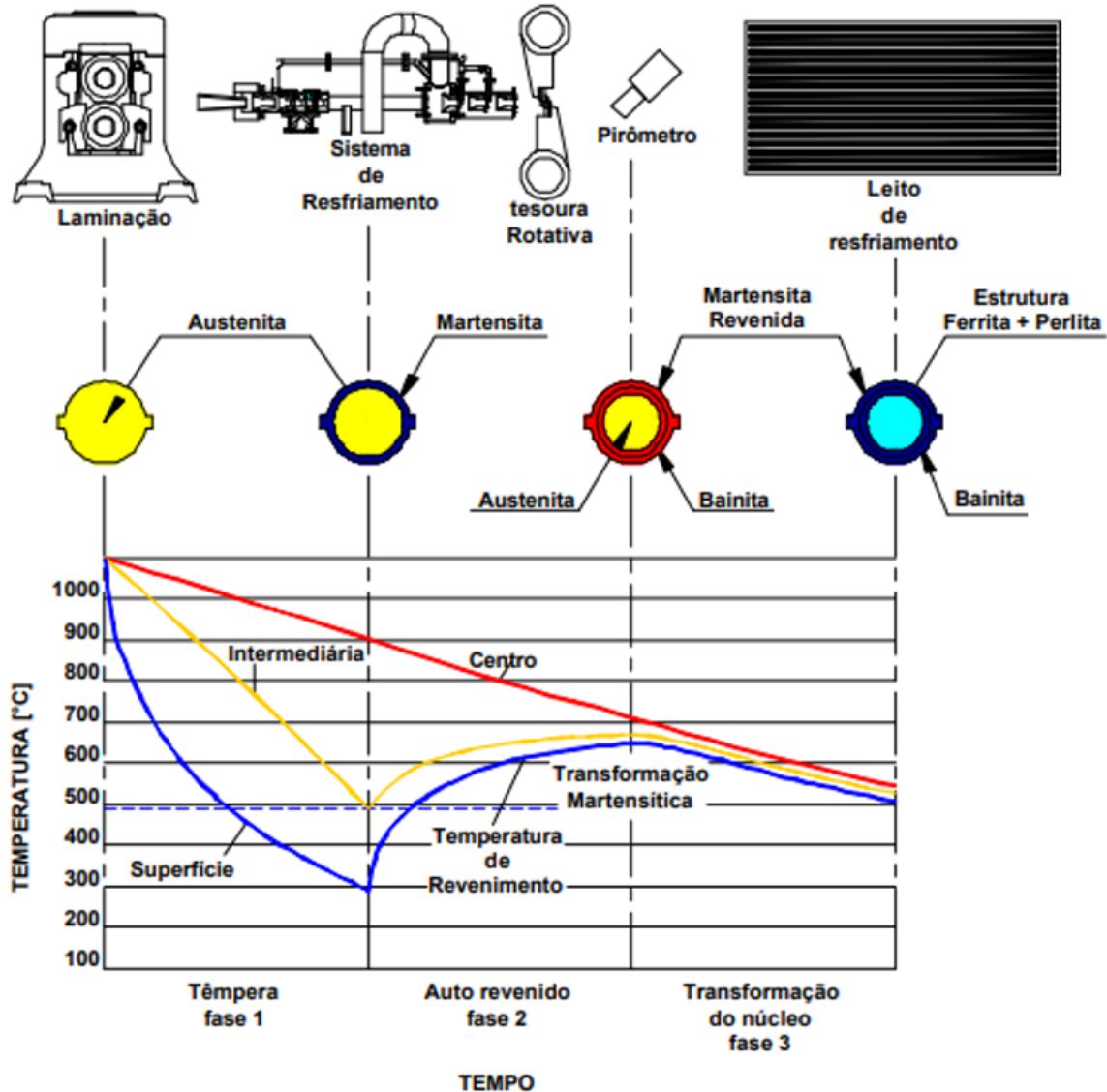
Fonte: Adaptado de Staats (2010)

## 2.4 Tratamento térmico de vergalhões de aço

Em algumas usinas e para alguns aços, após o processo de laminação pode ocorrer um tratamento térmico adicional visando melhorar as propriedades mecânicas do aço e reduzir os custos com elementos de ligas. O vergalhão ao deixar a último trem de acabamento passa através de uma série de canhões com sistema de resfriamento em água, onde é resfriado bruscamente, em seguida, sopradores de ar atuam para retirar o excesso de água do material. Esse sistema permite uma alta taxa de resfriamento e homogeneidade em toda a superfície do material (AUGUSTI, 1995). Desta forma, aços com percentual de carbono adequados transformam a sua superfície austenítica em martensita (ARAÚJO, 2010), conforme observado na Figura 7 (Fase 1).

Com a taxa de resfriamento sob controle, é necessária uma porcentagem mínima de carbono para garantir a produção de martensita apenas na superfície (J. WRIGHT M., 2016). Quando o vergalhão sai do sistema de resfriamento, cessando o resfriamento forçado com água, o núcleo ainda quente reaquece a superfície induzindo o revenimento da camada martensita (Fase

Figura 7 – Curvas de variação da temperatura da barra para diferentes posições de raio durante processo de tratamento térmico superficial



Fonte: Adaptado de Filho (2004)

2). O vergalhão vai para o leito de resfriamento e é resfriado ao ar à temperatura ambiente. Nesse processo, a bainita é formada abaixo da camada de martensita revenida e a ferrita + perlita são produzidas no núcleo do vergalhão (Fase 3). Este processo acontece pois a taxa de resfriamento do núcleo é menor, produzindo uma combinação adequada de propriedades mecânicas no final do processo (EZIO *et al.*, 2013).

A temperatura de entrada e de saída, o fluxo de água e a pressão do tratamento térmico são variáveis importantes para a obtenção das propriedades mecânicas desejadas. Para um mesmo tempo de tratamento, parâmetros do tratamento térmico e composição química,

quanto menor é a temperatura de entrada, menor é a temperatura de saída e maior é o percentual de martensita formada. Quanto maior a pressão e o fluxo da água, maior é a extração de calor e maior é o percentual de martensita formada (ver Figura 3). Quanto maior é o percentual martensítico, maior é o Limite de Escoamento (LE) e o Limite de Resistência à Tração (LRT) e menor é o Alongamento Percentual (AP) do aço.

## 2.5 Propriedades mecânicas

As propriedades mecânicas de um aço definem como o material se comporta quando sujeito a cargas externas. Dependentes de sua composição química e das variáveis do tratamento termomecânico, as propriedades mecânicas são comumente utilizadas para determinar se o produto está dentro das especificações de qualidades definidas pelo cliente. O Limite de Escoamento (LE), Limite de Resistência à Tração (LRT), Razão LRT/LE e Alongamento Percentual (AP) são propriedades que determinam se o vergalhão pode ser aplicado na construção civil.

### 2.5.1 Limite de Escoamento

Quando são submetidos a esforços mecânicos, os aços podem sofrer deformação elástica ou plástica, dependendo da magnitude da força aplicada. O comportamento elástico é observado se o material deforma durante a aplicação da carga, mas retorna às suas dimensões originais depois de cessado o carregamento. Durante a aplicação da força os átomos se afastam de suas posições iniciais e retornam as mesmas quando o carregamento cessa. Assim, define-se Limite de Escoamento como a tensão máxima que pode ser aplicada sem que o material sofra deformação plástica ou irreversível. Na fase elástica, a tensão ( $\sigma$ ) e a deformação ( $\varepsilon$ ) são proporcionais entre si e obedecem a lei de Hooke, de acordo com a seguinte equação (CALLISTER; RETHWISCH, 2007).

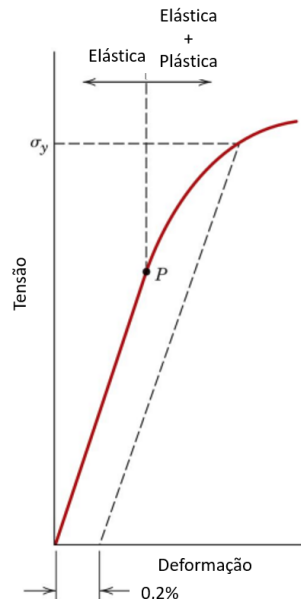
$$\sigma = E\varepsilon, \tag{2.1}$$

onde  $E$  é a constante de proporcionalidade conhecida como módulo de elasticidade ou módulo de Young. Quanto maior o módulo, menor será a deformação elástica resultante de uma determinada carga.

A deformação plástica ocorre quando o material é submetido a tensões mais elevadas e com a sua anulação o material não retorna as dimensões originais. Diferente da deformação

elástica, a deformação plástica envolve a quebra de ligações atômicas seguidas pela formação de novas ligações.

Figura 8 – Diagrama tensão-deformação mostrando a deformação elástica e plástica. O limite de escoamento está indicado pela tensão  $\sigma_y$ .



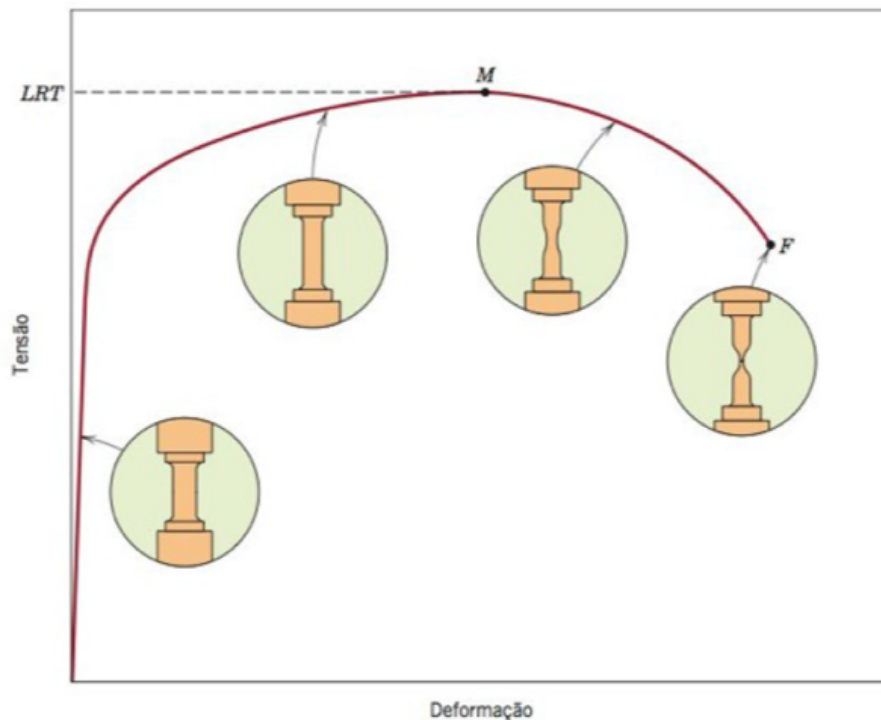
Fonte: Callister e Rethwisch (2007)

A região entre a fase elástica e plástica acontece de forma gradual, em alguns casos não podendo ser determinado com precisão. Quando a curva tensão e deformação começa a se afastar da proporcionalidade (ponto P, Figura 8) dizemos que se inicia a fase plástica (CALLISTER; RETHWISCH, 2007). Por convenção, foi estabelecido que o Limite de Escoamento (LE) é a interseção de uma linha reta paralela à porção elástica da curva tensão-deformação que após liberada causa uma pequena deformação residual, normalmente igual a 0,2%, com a curva tensão-deformação (CALLISTER; RETHWISCH, 2007).

### 2.5.2 Limite de Resistência à Tração

Após o início da fase plástica, a tensão necessária para provocar deformações progressivas no material aumenta até um valor máximo (ponto M, Figura 9) denominado Limite de Resistência à Tração (LRT). Ao alcançar esse limite (ponto M) o material apresenta uma redução bem localizada em sua seção transversal, chamada estrição, e a partir daí a tensão necessária para provocar novas deformações diminui até a sua ruptura (ponto F). Esse ponto corresponde ao Limite de Resistência à Fratura do material (CALLISTER; RETHWISCH, 2007).

Figura 9 – Diagrama tensão-deformação até o ponto de fratura, F. LRT é o Limite de Resistência à Tração e está indicado pelo ponto M.



Fonte: Callister e Rethwisch (2007)

### 2.5.3 Razão LRT/LE

Para algumas aplicações, além de conhecer o LE e o LRT, é importante saber a relação entre essas duas propriedades. A razão LRT/LE é um critério importante para muitas aplicações à prova de falhas onde um valor baixo é desejado. Os exemplos incluem aços resistentes a terremotos e casos em que uma boa resistência à fadiga é necessária (SINGH *et al.*, 1998).

De acordo com Jiao e Zhao (2001), uma das maneiras de se medir a ductilidade do material é através da razão LRT/LE e do alongamento percentual. A ductilidade é a capacidade do material se deformar sem se romper ao ser submetido a uma tração.

De acordo com as normas europeias (NORMALISATION, 2005), para o material ser considerado dúctil ele tem que respeitar as três condições listadas abaixo:

- Razão LRT/LE  $\geq 1,1$
- Alongamento Percentual  $\geq 15$
- Razão entre a deformação correspondente ao LRT e a deformação correspondente ao LE  $\geq 15$



### **2.5.4 Alongamento Percentual**

O Alongamento Percentual é uma medida da ductilidade do material e pode ser calculado a partir dos comprimentos inicial,  $c_0$ , e final,  $c_f$ , de um corpo de prova submetido ao ensaio de tração, de acordo com Callister e Rethwisch (2007)

$$\%AP = \left( \frac{c_f - c_0}{c_0} \right) \times 100. \quad (2.2)$$

## **2.6 Resumo do Capítulo**

Neste capítulo, descreveu-se conceitos básicos sobre o aço e os principais elementos químicos e parâmetros do processo que influenciam em sua propriedade mecânica. O próximo capítulo descreve os métodos de estimação e o conjunto de dados utilizados.

### 3 MÉTODOS DE ESTIMAÇÃO E CONJUNTO DE DADOS

Neste capítulo, conceitos importantes dos cinco métodos de estimação ( Análise de Regressão Linear, Máquina de Aprendizagem Mínima, Redes Neurais Artificiais, Máquinas de Vetores de Suporte e Máquinas de Vetores de Suporte por Mínimos Quadrados) são apresentados. Em seguida, o coeficiente de determinação-R<sup>2</sup> utilizado para medir a qualidade do modelo é explicado ao leitor. Por fim, o conjunto de dados utilizado nesse trabalho é apresentado.

#### 3.1 Métodos de estimação

Em muitos problemas na indústria siderúrgica há variáveis que parecem estar intrinsecamente relacionadas, sendo necessário analisar essa relação para entender melhor o problema. Os métodos de estimação apresentados nessa secção podem ser utilizados na construção de um modelo que expresse o resultado de uma variável como função de uma ou mais variáveis. Este modelo pode ser utilizado para prever uma variável em função de outras.

##### 3.1.1 Análise de Regressão Linear

A análise de regressão linear é um tipo de aproximação de função paramétrica na qual os coeficientes de regressão são procurados para que uma função linear possa ser ajustada em um espaço de estudo multidimensional. Assim, o principal objetivo desse método é estabelecer uma equação que possa ser usada para prever valores da variável dependente,  $y$ , a partir de variáveis independentes,  $x$ .

Seja a variável dependente,  $y_i \in \mathbb{R}$ , relacionada a  $x_m$  variáveis independentes. A resposta da previsão,  $y_{ip}$ , é uma variável aleatória dada por (MONTGOMERY *et al.*, 2012).

$$y_{ip} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_m x_{im} + \varepsilon_i, \quad (3.1)$$

onde  $\beta_j$  tendo  $j$  variando de 0 a  $m$  são os  $m + 1$  coeficientes da equação e  $\varepsilon_i$  é o erro aleatório relacionado ao  $i$ -ésimo dado. A variável predita pode ser reescrita de forma matricial como

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon}, \quad (3.2)$$

onde  $\mathbf{y}$  é a matriz dos dados previstos de dimensão  $N$ ;  $\mathbf{X}$  uma matriz de tamanho  $N \times D + 1$

(sendo a primeira coluna igual a 1);  $\mathbf{b}$  é o vetor dos  $D + 1$  coeficientes  $\beta$ ; e  $\varepsilon$  é a matriz dos erros associados a cada dado estimado.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad (3.3)$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1D} \\ 1 & x_{21} & x_{22} & \cdots & x_{2D} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_{D1} & x_{D2} & \cdots & x_{ND} \end{bmatrix}, \quad (3.4)$$

$$\mathbf{b} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_D \end{bmatrix}, \quad (3.5)$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}. \quad (3.6)$$

O erro residual é obtido pela diferença entre o valor da variável dependente prevista,  $y_{np}$ , e o valor observado da variável dependente,  $y_n$ . A estimativa dos coeficientes  $\beta$  é baseada no método de aproximação dos mínimos quadrados (WHITE, 1980). Por meio dele, buscamos uma maneira de minimizar o erro residual

$$S = \sum_{n=1}^N (y_{np} - y_n)^2 = \sum_{n=1}^N (e_n)^2. \quad (3.7)$$

O vetor de erros  $\mathbf{e} = \mathbf{Y} - \mathbf{X}\mathbf{b}$ , de modo a obter

$$\mathbf{S}(\mathbf{b}) = (\mathbf{Y} - \mathbf{X}\mathbf{b})^T (\mathbf{Y} - \mathbf{X}\mathbf{b}) = \mathbf{Y}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{X}\mathbf{b} - \mathbf{b}^T \mathbf{X}^T \mathbf{Y} - \mathbf{b}^T \mathbf{X}^T \mathbf{X}\mathbf{b}, \quad (3.8)$$

onde as matrizes  $(\mathbf{Y} - \mathbf{X}\mathbf{b})^T$ ,  $\mathbf{Y}^T$ ,  $\mathbf{b}^T$  e  $\mathbf{X}^T$  são as transpostas das matrizes  $(\mathbf{Y} - \mathbf{X}\mathbf{b})$ ,  $\mathbf{Y}$ ,  $\mathbf{b}$  e  $\mathbf{X}$ , respectivamente.

$$\frac{\partial \mathbf{S}(\mathbf{b})}{\partial \mathbf{b}} = (\mathbf{Y} - \mathbf{X}\mathbf{b})^T (\mathbf{Y} - \mathbf{X}\mathbf{b}) = -2\mathbf{X}^T \mathbf{Y} + 2\mathbf{X}^T \mathbf{X}\mathbf{b} = 0 \Rightarrow \mathbf{X}^T \mathbf{X}\mathbf{b} = \mathbf{X}^T \mathbf{Y}. \quad (3.9)$$

Logo,

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (3.10)$$

Se uma variável  $x_j$  tem poder explicativo sobre  $y$ , seu coeficiente  $\beta_j$  deve ser estatisticamente diferente de zero. Ou seja, este coeficiente deve ser suficientemente maior ou menor que zero. Se isso não for verdade, a variável pode ser removida do modelo sem perda de qualidade. Para verificar a significância dos coeficientes, foi realizado um teste de hipóteses para determinar se o valor de  $\beta_j$  é igual a zero. Foram elaboradas duas hipóteses para cada  $\beta_j$ :

- $H_0 : \beta_j = 0$  (Variável  $x_j$  não tem relação com a variável  $y_p$ )
- $H_1 : \beta_j \neq 0$  (Variável  $x_j$  tem relação com a variável  $y_p$ )

Foi levado em consideração que o estimador  $\beta$  é não viesado e segue uma distribuição normal com média  $\mu_{\beta_j}$  e variância  $\sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$ , onde  $\sigma^2$  é a variância do erro e  $(\mathbf{X}^T \mathbf{X})^{-1}$  é a inversa da matriz  $(\mathbf{X}^T \mathbf{X})$ . (DOUGLAS; RUNGER, 2002), dessa forma,

$$\beta \sim N(\mu_{\beta}, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}). \quad (3.11)$$

Podemos aproximar  $\sigma^2$  da variância amostral  $s^2$ , dada por

$$s^2 = \frac{\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}}{N - (D + 1)}, \quad (3.12)$$

onde  $D$  é o número de variáveis independentes.

Então a distribuição da variável aleatória  $t$  será dada por (SPIEGEL; STEPHENS, 2000):

$$t = \frac{\beta - \mu_{\beta}}{\sqrt{\sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}}} \sim t(N - D - 1). \quad (3.13)$$

Considerando a hipótese nula  $\mu_{\beta_j} = 0$ , pela  $t$  de student  $j$  é dado pela Eq. 3.14

$$t_j = \frac{\beta_j}{s\sqrt{a_{jj}}} \sim t(N-D-1), \quad (3.14)$$

onde  $a_{jj}$  é o  $j$ -ésimo elemento da diagonal da matriz  $(\mathbf{X}^T \mathbf{X})^{-1}$ .

Aplicando o valor de  $t_j$  para a  $t$  de student com  $(N-D-1)$  graus de liberdade, é possível obter coeficientes necessários para rejeitarmos a hipótese nula.

A significância de um coeficiente de regressão pode ser útil para estudar o efeito linear das variáveis independentes individualmente e descobrir se os resultados são compatíveis com os princípios metalúrgicos conhecidos. O resumo da Análise de Regressão Linear + Teste de Significância das variáveis está sendo apresentado pelo algoritmo 1.

### 3.1.2 Máquina de Aprendizagem Mínima

A MAM é um novo método de regressão baseado na distância proposto por Souza Junior *et al.* (2015). A MAM mapeia distâncias entre amostras de entrada para distâncias entre amostras de saída. Um número fixo de  $K$  amostras são aleatoriamente selecionados do conjunto de treinamento de tamanho  $N$  ( $N \gg K$ ) para servir como pontos de referência para o cálculo de distâncias. Em seguida, o algoritmo constrói um mapeamento linear da matriz de distância de entrada para a matriz de distância de saída.

Na fase de teste, o mapa de distância construído é usado para fornecer estimativas das distâncias de  $K$  pontos de referência de saída para os valores de saída desconhecidos. A fim de obter os valores de saída previstos no espaço original, métodos de otimização são usados para estimar a saída com base em um problema de multilateração, semelhante ao problema de determinar a localização de um veículo ou ponto estacionário no espaço.

Formalmente, se o vetor de variáveis dependentes,  $\mathbf{y} \in \mathbb{R}^S$ , está relacionado a um vetor  $\mathbf{x} \in \mathbb{R}^D$  de  $D$  variáveis independentes, um problema de regressão geral sobre o conjunto de  $N$  pares de entrada-saída medidos  $\mathcal{X} = \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^N \subset \mathbb{R}^D \times \mathbb{R}^S$  é frequentemente declarado como o problema de aproximar o seguinte mapeamento de múltiplas entradas e saídas:

$$\mathbf{y}_n = \mathbf{F}(\mathbf{x}_n) + \boldsymbol{\varepsilon}_n, \quad (3.15)$$

onde  $\mathbf{F} : \mathbb{R}^D \rightarrow \mathbb{R}^S$  é um mapeamento desconhecido e  $\boldsymbol{\varepsilon}_n$  é um vetor de ruído aleatório que representa incertezas nos processos de modelagem e medição.

---

**Algoritmo 1: Análise de Regressão Linear + Teste de Significância**


---

**Entrada:** Dados

**Saída:** Variável predita, coeficiente de determinação, variáveis linearmente relacionadas com a variável predita

**início**

**para**  $i=1:100$  **faça**

A separação aleatória dos dados de treino e teste;

Normalize os dados de entrada;

Construa a matriz  $\mathbf{X}$  de tamanho  $N \times D + 1$ , onde a primeira coluna é igual a 1 e as outras são as variáveis dependentes dos dados de treino;

Encontre a matriz  $\mathbf{b}$

Faça o teste de significância dos coeficientes  $\mathbf{b}$  da regressão linear;

Encontre as variáveis linearmente relacionadas com a variável predita;

Retire as variáveis que não são estatisticamente significante dos dados e construa novamente a matriz  $\mathbf{X}$  ;

Encontre a matriz  $\mathbf{b}$ ;

Construa a matriz  $\mathbf{X}$  para os dados de teste, onde a primeira coluna é igual a 1 e as outras são os dados de teste sem as variáveis que não são estatisticamente significantes;

Encontre a variável predita pela regressão ( $\mathbf{Y} = \mathbf{bX}$ );

Calcule o coeficiente de determinação entre a variável predita e o valor observado (ver seção 2.7);

**fim**

Faça o cálculo da média, mínimo, desvio dos 100 coeficientes de determinação encontrados

**fim**

---

Inicialmente, um conjunto de  $K$  pontos de referência, onde  $\mathcal{R} = \{(\mathbf{m}_k, \mathbf{t}_k)\}_{k=1}^K$ ,  $\mathcal{R} \subset \mathcal{X}$  é aleatoriamente selecionado no conjuntos de dados de treinamento  $\mathcal{X}$ . Depois, deve-se calcular duas matrizes de distância, a saber,  $\Delta_{\mathbf{x}} \in \mathbb{R}^{N \times K}$  e  $\Delta_{\mathbf{y}} \in \mathbb{R}^{N \times K}$  dados, respectivamente, pela distância Euclidiana (DANIELSSON, 1980) de  $(\mathbf{x}_n)_{n=1}^N$  e  $(\mathbf{y}_n)_{n=1}^N$  para cada ponto de referência  $(\mathbf{m}_k)_{k=1}^K$  e  $(\mathbf{t}_k)_{k=1}^K$ . Assim, a  $k$ -ésima coluna de  $\Delta_{\mathbf{x}}$  contém as distâncias  $d(\mathbf{x}_n, \mathbf{m}_k)$  entre os  $n = 1, \dots, N$  pontos de entrada  $\mathbf{x}_n$  e os  $k$ -ésimos pontos de referência  $\mathbf{m}_k$ . Da mesma forma, a  $k$ -ésima coluna de  $\Delta_{\mathbf{y}}$  contém as distâncias  $\delta(\mathbf{y}_n, \mathbf{t}_k)$  entre os  $n = 1, \dots, N$  pontos de saídas  $\mathbf{y}_n$  e os  $k$ -ésimos pontos de referência de saída  $\mathbf{t}_k$ .

Podemos relacionar  $\Delta_{\mathbf{x}}$  e  $\Delta_{\mathbf{y}}$  através de um mapeamento  $\mathbf{G}$ , de acordo com a equação:

$$\Delta_{\mathbf{y}} = \mathbf{G}(\Delta_{\mathbf{x}}) + \mathbf{E}, \quad (3.16)$$

onde  $\mathbf{E} \in \mathbb{R}^{N \times K}$  é a matriz dos erros residuais.

Assumindo que o mapeamento  $\mathbf{G}(\cdot)$  seja linear, podemos reescrever a equação 3.16 como

$$\Delta_{\mathbf{y}} = \Delta_{\mathbf{x}}\mathbf{B} + \mathbf{E}, \quad (3.17)$$

onde a matriz  $\mathbf{B}$  pode ser estimado através da minimização do erro residual (SEBER; LEE, 2012). A solução é calculada pelo conhecido Método dos Mínimos Quadrados (MMQ):

$$\hat{\mathbf{B}} = (\Delta_{\mathbf{x}}^T \Delta_{\mathbf{x}})^{-1} \Delta_{\mathbf{x}} \Delta_{\mathbf{y}}, \quad (3.18)$$

onde o sobrescrito  $T$  denota a transposição da matriz. Assim que a matriz estimada  $\hat{\mathbf{B}}$  estiver disponível, ela pode ser usada na previsão para um novo vetor de teste de entrada  $\mathbf{x} \in \mathbb{R}^D$ :

$$\hat{\delta}(\mathbf{y}, \mathcal{R}) = d(\mathbf{x}, \mathcal{R}) \hat{\mathbf{B}}, \quad (3.19)$$

onde  $d(\mathbf{x}, \mathcal{R}) = [d(\mathbf{x}, \mathbf{m}_1), \dots, d(\mathbf{x}, \mathbf{m}_K)]^T$  e  $\hat{\delta}(\mathbf{y}, \mathcal{R}) = [\hat{\delta}(\mathbf{y}, \mathbf{t}_1), \dots, \hat{\delta}(\mathbf{y}, \mathbf{t}_K)]^T$ . O vetor  $\hat{\delta}(\mathbf{y}, \mathcal{R})$  fornece uma estimativa das distâncias do ponto de saída  $\mathbf{y}$  para os pontos de referência  $\{\mathbf{t}_k\}_{k=1}^K$ . Deve-se notar, no entanto, que o valor de interesse é o vetor de saída  $\mathbf{y}$ , não o vetor de distância  $\hat{\delta}(\mathbf{y}, \mathcal{R})$ . Para estimar  $\mathbf{y}$  a partir dele, é necessário minimizar a função objetivo  $J(\mathbf{y})$  dada por

$$J(\mathbf{y}) = \sum_{k=1}^K \left( (\mathbf{y} - \mathbf{t}_k)^T (\mathbf{y} - \mathbf{t}_k) - \hat{\delta}^2(\mathbf{y}, \mathbf{t}_k) \right)^2, \quad (3.20)$$

que pode ser minimizado por qualquer método de mínimos quadrados não linear, incluindo métodos clássicos baseados em gradiente descendente. Nesta tese, usamos o algoritmo de Levenberg-Marquardt (LEVENBERG, 1944) (MARQUARDT, 1963) para este propósito. O resumo da máquina de aprendizagem mínima está apresentado no algoritmo 2.

### 3.1.3 Redes Neurais Artificiais

Inspiradas na estrutura neural de organismos inteligentes, as redes neurais artificiais adquirem conhecimento através da experiência. O primeiro neurônio artificial, criado em 1943,

---

**Algoritmo 2:** Máquina de Aprendizagem Mínima
 

---

**Entrada:** Dados

**Saída:** Variável predita e coeficiente de determinação

**início**
**para**  $i=1:100$  **faça**

A separação aleatória dos dados de treino e teste;

Normalize os dados de entrada;

 Selecione aleatoriamente  $K$  pontos de referência, onde  $\mathcal{R} = \{(\mathbf{m}_k, \mathbf{t}_k)\}_{k=1}^K$ ,

 $\mathcal{R} \subset \mathcal{X}$ ;

 Calcule as matrizes de distâncias  $\Delta_{\mathbf{x}} \in \mathbb{R}^{N \times K}$  e  $\Delta_{\mathbf{y}} \in \mathbb{R}^{N \times K}$ ;

 Calcule  $\hat{\mathbf{B}} = (\Delta_{\mathbf{x}}^T \Delta_{\mathbf{x}})^{-1} \Delta_{\mathbf{x}} \Delta_{\mathbf{y}}$ ;

 Calcule o vetor de distância  $d(\mathbf{x}, \mathcal{R}) = [d(\mathbf{x}, \mathbf{m}_1), \dots, d(\mathbf{x}, \mathbf{m}_K)]^T$ 

 Calcule  $\hat{\delta}(\mathbf{y}, \mathcal{R}) = d(\mathbf{x}, \mathcal{R}) \hat{\mathbf{B}}$ ;

 Construa o vetor  $J(\mathbf{y})$ ;

Utilize a técnica de otimização Levenberg- Marquadt para estimar o valor da variável predita;

Calcule o coeficiente de determinação entre a variável predita e o valor observado (ver seção 2.7);

**fim**

Faça o cálculo da média, mínimo, desvio dos 100 coeficientes de determinação encontrados

**fim**


---

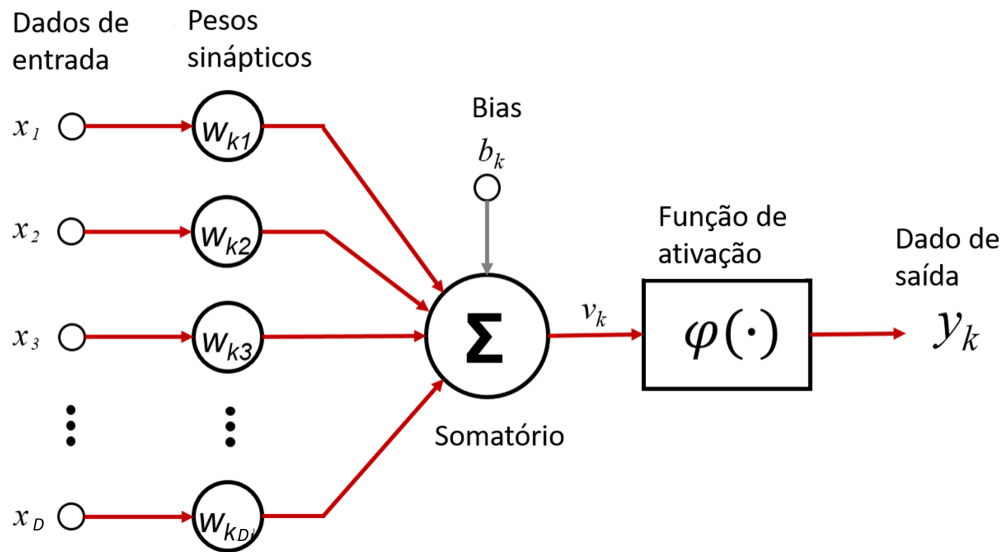
era um dispositivo binário que tinha como entrada ganhos arbitrários que podiam ser de dois tipos: excitatórias (sinal positivo) ou inibitórias (sinal negativo). Era realizado o cálculo da média ponderada das entradas, de acordo com o seu fator de ponderação, e se o resultado fosse maior ou igual a um certo valor, a saída do neurônio era pulso; caso contrário, a saída do neurônio era não pulso. Este neurônio ficou conhecido como McCulloch-Pitts (MCCULLOCH; PITTS, 1943).

Baseado no modelo de McCulloch-Pitts, outros modelos de neurônios foram criados. A representação de um desses neurônio é ilustrada na Figura 10, onde  $(x_1, x_2, \dots, x_D)$  são as características de entrada;  $(w_{k1}, w_{k2}, \dots, w_{kD})$  são os ganhos associados a cada vetor de entrada (pesos das conexões);  $b_k$  é o bias, entrada de valor inicial aleatório associado a cada neurônio;  $v_k$  é a saída do somatório;  $\varphi(\cdot)$  é a função de ativação;  $y_k$  é a saída do neurônio.

Desta forma, a saída do somatório,  $v_k$ , pode ser escrito pela Equação 3.21 e a saída



Figura 10 – Modelo matemático de um neurônio artificial



Fonte: Próprio autor.

do neurônio,  $y_k$ , pela Equação 3.22.

$$v_k = \sum_{i=1}^D x_i w_{ki} + b_k, \quad (3.21)$$

$$y_k = \varphi(v_k) = f\left(\sum_{i=1}^D x_i w_{ki} + b_k\right). \quad (3.22)$$

A função de ativação  $\varphi(\cdot)$  avalia o potencial de ativação  $v_k$  e define a saída do neurônio. Segundo Haykin (2009), podemos identificar três tipos básicos de funções de ativação:

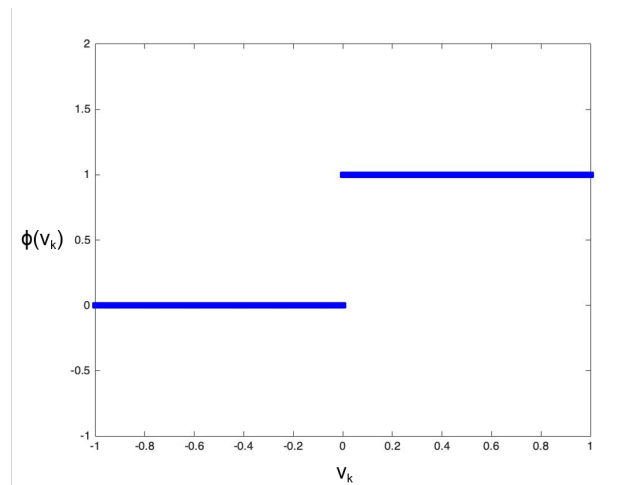
- **Função de Limiar:** É a função do modelo de McCulloch, descrito anteriormente. Nesta função a saída do neurônio assume o valor 1 se o campo induzido do neurônio é positivo e 0 caso contrário, como podemos observar na equação 3.23.

$$y_k = \begin{cases} 1, & \text{se } v_k \geq 0, \\ 0, & \text{se } v_k < 0. \end{cases} \quad (3.23)$$

A função está sendo representada graficamente na Figura 11

- **Função Linear por Partes:** Nesta função a saída do neurônio pode assumir valores entre 0 e 1. A saída  $y_k$  é 1 para valores de  $v_k$  maiores ou iguais a 1. Para valores de  $v_k$  menores

Figura 11 – Função de Limiar

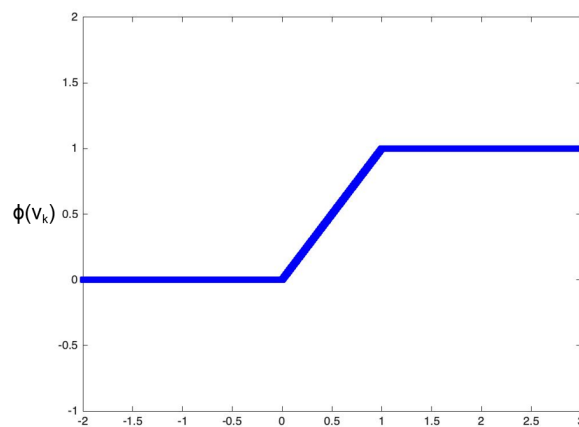


Fonte: Próprio autor.

ou iguais a 0 é 0 e entre 0 e 1 é igual a  $v_k$ , como pode ser observado na Equação 3.24 e na Figura 12.

$$y_k = \begin{cases} 1, & \text{se } v_k \geq 1, \\ v_k, & \text{se } 1 > v_k > 0, \\ 0, & \text{se } v_k \leq 0. \end{cases} \quad (3.24)$$

Figura 12 – Função Linear por Partes



Fonte: Próprio autor.

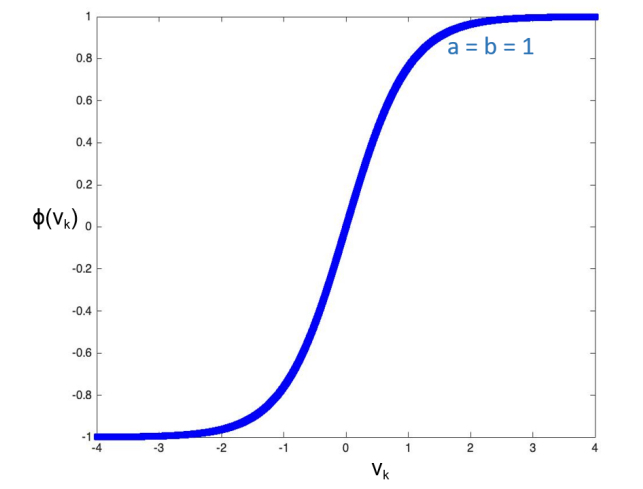
- **Função Tangente Hiperbólica:** É uma função monotônica crescente que possui propriedade assintótica (tende a uma constante quando  $v_k$  tende  $\pm\infty$ ) e suavidade (possui derivada

de todas ordens). Esta função é representada pela seguinte equação

$$y_k = a \left( \frac{1 - e^{-2bv_k}}{1 + e^{-2bv_k}} \right), \quad (3.25)$$

onde  $a$  e  $b$  são hiperparâmetros do modelo. O gráfico correspondente é mostrado na Figura 13.

Figura 13 – Tangente hiperbólica



Fonte: Próprio autor.

De acordo com Haykin (2009), os valores encontrados na literatura que apresentam melhores resultados para as constantes  $a$  e  $b$  são

$$a = 1 \text{ e } b = 1, \quad (3.26)$$

ou

$$a = 1.7159 \text{ e } b = 2/3. \quad (3.27)$$

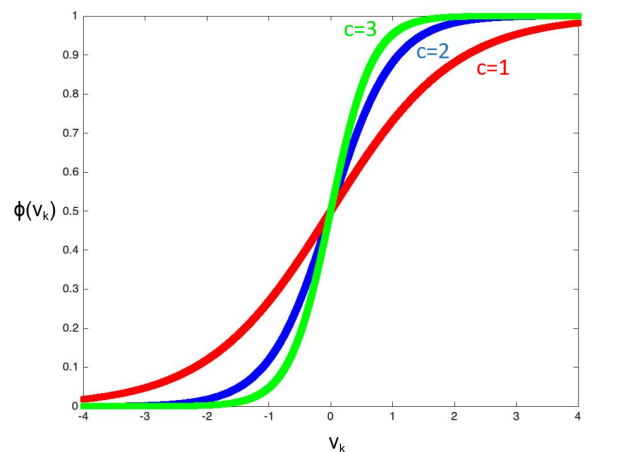
Para valores de  $a = 1.7159$  e  $b = 2/3$ ,  $y_k = 1$  se  $v_k = 1$  e  $y_k = -1$  se  $v_k = -1$ , a inclinação da função de ativação fica próxima da unidade na origem e a derivada segunda de  $y_k$  atinge seu valor máximo em  $v_k = 1$ .

- **Função Sigmóide logística:** Assim como a tangente hiperbólica, a função sigmóide logística é monotônica crescente que possui propriedade assintótica e suavidade. Devido as suas propriedades a função sigmóide logística e tangente hiperbólica são uma das mais utilizadas nas redes neurais. A função sigmóide logística é dada pela expressão

$$y_k = \frac{1}{1 + e^{-cu_k}}, \quad (3.28)$$

onde  $c > 0$  é a inclinação da curva. A Figura 14 ilustra o gráfico da função sigmóide logística para diferentes valores de  $c$ . Para valores elevados de  $c$  a função se aproxima da limiar.

Figura 14 – Função sigmóide logística para diferentes valores de  $c$ .



Fonte: Próprio autor.

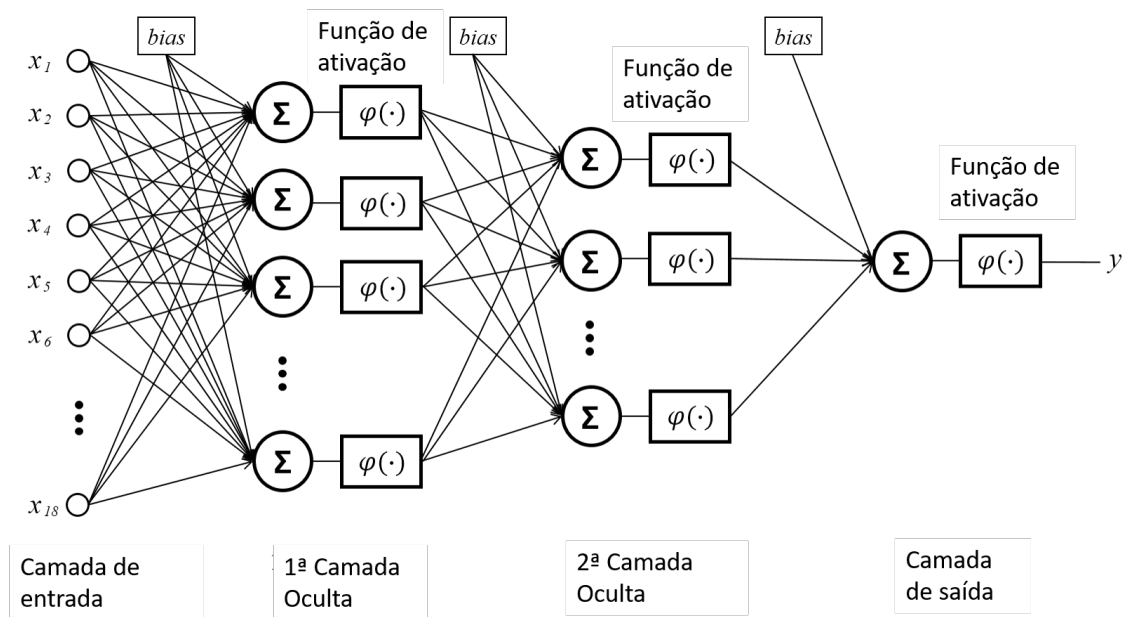
Valores encontrados na literatura que apresentam melhores resultados para as constantes a constante  $c$  são Haykin (2009):

$$c = 1 \text{ ou } 2,2. \quad (3.29)$$

Para resolver problemas mais complexos, os quais não podem ser resolvidos por um único neurônio, foi criada a rede neural de multicamadas. Esta rede é composta por neurônios organizados em três ou mais camadas. A primeira camada é a de entrada, onde os dados são apresentados ao algoritmo. Em seguida temos as camadas ocultas, que tornam a rede capaz de extrair estatísticas de ordem elevada e a última camada é a de saída, que é a resposta da rede para o usuário.

Os sinais de saída da primeira camada são utilizados como entrada para a segunda camada, que são processados pela segunda camada e utilizados como entrada para a terceira camada, e assim por diante para o resto da rede. Desta forma, temos uma rede totalmente conectada, onde cada nó em uma camada se conecta, com um certo peso, em um nó da camada seguinte. Esse tipo de rede neural é chamada de perceptron de múltiplas camadas (Figura 15).

Figura 15 – Modelo matemático da rede perceptron de múltiplas camadas



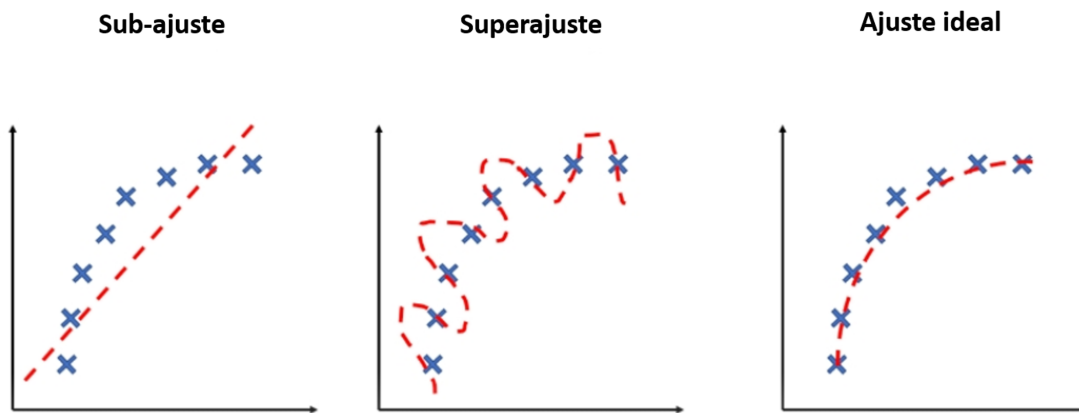
Fonte: Próprio autor.

O número de neurônios ocultos deve ser adequadamente determinado, pois a escolha inadequada pode provocar subajuste ou sobreajuste. No subajuste, a rede tem poucos parâmetros ajustáveis e não consegue capturar as características específicas dos dados, generalizando muito o modelo. Já no sobreajuste, a rede tem muitos parâmetros ajustáveis e captura características muito específicas dos dados, generalizando pouco o modelo (HAYKIN, 2009). A Figura 16 ilustra o subajuste, sobreajuste e ajuste ideal de um modelo. Denominaremos de  $q_1$ ,  $q_2$  e  $m$ , respectivamente, aos números de neurônio da primeira e segunda camadas ocultas e da camada de saída.

A rede neural pode aprender a reconhecer padrões e ser utilizada na classificação e previsão de dados, ela pode ser dividida em dois tipos: não supervisionada ou supervisionada. No primeiro caso, a rede aprende através das características dos dados extraídos do modelo. Já no segundo caso, é apresentado os dados de saída desejado ao modelo para o mesmo aprender a reconhecer padrões. Neste trabalho, utilizamos uma rede neural supervisionada para a previsão de dados.

Em uma rede neural supervisionada, o algoritmo é dividido em duas fases, a fase de treino e a de teste. Na fase de treino, uma parte dos dados é utilizada para o aprendizado do algoritmo. Desta forma, são apresentados os dados e as suas respectivas classificações ao programa, que tem como objetivo ajustar os ganhos e o limiar, até obter a maior taxa média de acerto na classificação desses dados. Já na fase de teste, é verificado se o algoritmo aprendeu a

Figura 16 – Representação gráfica do subajuste, sobreajuste e ajuste ideal de um modelo



Fonte: Traduzido de Ramasubramanian e Moolayil (2019)

classificar os dados corretamente. A entrada é um conjunto dos dados que não foram utilizados anteriormente, e os ganhos e o bias da rede são aqueles obtidos na fase de treinamento. Nesta fase, os dados são processados e a previsão da variável independente é obtida.

O ganho inicial e o bias são escolhidos aleatoriamente e, depois, o ganho é ajustado na fase de treinamento do algoritmo. De acordo com Hebb (2005), os ganhos devem ser ajustados para cada camada da rede, começando da camada de saída e seguindo em direção a camada de entrada, da seguinte forma:

$$w_{ki}^{novo} = w_{ki}^{atual} + \eta \delta_k y_i + \alpha (w_{ki}^{atual} - w_{ki}^{atual-1}), \quad (3.30)$$

onde o parâmetro  $\eta$  é definido como taxa de aprendizado e  $\alpha$  como constante de momento, parâmetros informados pelo usuário, e  $\delta_k$  é o gradiente local, dado por:

$$\delta_k = (y_k^o - y_k) \phi_k', \quad (3.31)$$

onde  $\phi_k'$  é a derivada da função de ativação e  $y_k^o$  é a saída observada.

A cada vez que todo o conjunto de treinamento é apresentado à rede neural dizemos que ocorreu uma "época". O ajuste dos pesos sinápticos é repetido até que o critério de parada seja alcançado. Para parada antecipada de treinamento baseada na validação cruzada, um subconjunto dos dados utilizados na fase de treinamento da rede (dados de validação) é separado para determinar o número ideal de épocas de treinamento. A apresentação do conjunto de treinamento é repetida até que a curva de aprendizagem obtida para o subconjunto de validação apresente o menor erro médio quadrático (ponto mínimo na curva de aprendizagem), conforme ilustrado

na Figura 17. O uso desta técnica de parada diminui a possibilidade de ocorrer sobreajuste dos dados.

Figura 17 – Regra da parada antecipada baseada na validação cruzada



Fonte: Haykin (2009)

Após obtermos os valores do bias e ajustarmos os valores dos ganhos, começa a fase de teste. Na fase de teste, as características de entrada são apresentadas ao algoritmo e este retorna a previsão dos dados de saída. O resumo da rede neural artificial está sendo apresentado pelo algoritmo 3.

### 3.1.4 Máquinas de Vetores de Suporte

A Máquina de Vetores de Suporte (MVS) proposta por Drucker *et al.* (1997) vem se tornando uma poderosa ferramenta quando aplicada a problemas de classificação de padrões não lineares e de regressão. A ideia básica por trás do algoritmo de regressão por máquinas de vetores de suporte é transformar os dados de entrada  $\mathcal{X}$  em um espaço  $\mathcal{H}$  de dimensão muito alta (possivelmente infinita) e, em seguida, realizar uma regressão linear no espaço transformado (SUYKENS *et al.*, 2002). Segundo o teorema de Cover (HAYKIN, 2009), se aplicarmos uma transformação não-linear para um espaço de mais alta dimensão temos alta probabilidade dos dados se tornarem linearmente separáveis.

Exemplificando, sejam dois conjuntos de dados não linearmente separáveis, como

---

**Algoritmo 3:** Redes Neurais Artificiais
 

---

**Entrada:** Dados, taxa de aprendizagem, função de ativação das camadas ocultas, parâmetros da função de ativação ( $a$ ,  $b$  e  $c$ ), número de épocas máximo ( $E_m$ ), número de neurônios da primeira e da segunda camada oculta

**Saída:** Variável predita e coeficiente de determinação

**início**

**para**  $i=1:100$  **faça**

A separação aleatória dos dados de treino, validação e teste;

A escolha aleatória de todos os ganhos e dos bias iniciais que serão utilizados na rede;

Normalização dos dados de entrada;

**para**  $ep=1:E_m$  **faça**

**para** *cada dado de treino* **faça**

O processamento dos dados pela rede;

Se a saída for incorreta, ajuste o peso através do algoritmo de propagação backpropagation;

**fim**

**para** *cada dado de validação* **faça**

O processamento dos dados com os ganhos obtidos na fase de treino;

Calcule o erro quadrático médio obtido entre os dados de validação e os dados observados;

**fim**

Armazene o valor do erro quadrático médio para cada época e os ganhos correspondentes;

**fim**

Verifique quantas épocas proporcionou o mínimo erro quadrático médio na previsão dos dados de validação, este será o número de épocas ótimo;

Os ganhos correspondente ao número de épocas ótimo será usado para testar o algoritmo.

**para** *cada dado de teste* **faça**

O processamento com os ganhos obtidos na fase de treino ;

Encontre a variável predita pela rede;

**fim**

Calcule o coeficiente de determinação entre a variável predita e o valor observado;

**fim**

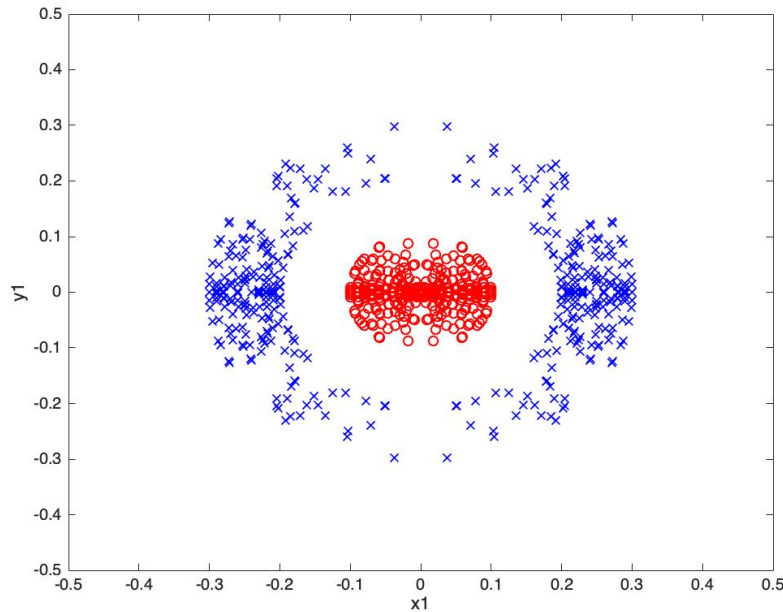
Faça o cálculo da média, mínimo, desvio dos 100 coeficientes de determinação encontrados;

**fim**



mostra a Figura 18. Podemos mapear os dados do espaço  $\mathbb{R}^2$  para um espaço característico  $\mathbb{R}^3$  através de uma transformação  $\phi(\cdot)$ , como mostra a Figura 19. Nota-se, que nesse mapeamento é possível separar os dados linearmente.

Figura 18 – Exemplo de um problema não linearmente separável



Fonte: Próprio autor.

Sem perda de generalidade, vamos considerar os mesmos dados definidos para a Máquina de Aprendizagem Mínima, exceto que a dimensão das amostras de saída é de  $S = 1$ :

$$\mathcal{X} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\} \subset \mathbb{R}^D \times \mathbb{R}. \quad (3.32)$$

A função  $\phi(\mathbf{x})$  é usada para transformar o espaço de entrada  $\mathbb{R}^D$  em um espaço de características de alta dimensão  $\mathcal{H}$ . Neste caso, o conjunto de dados em  $\mathcal{H}$  é dado por

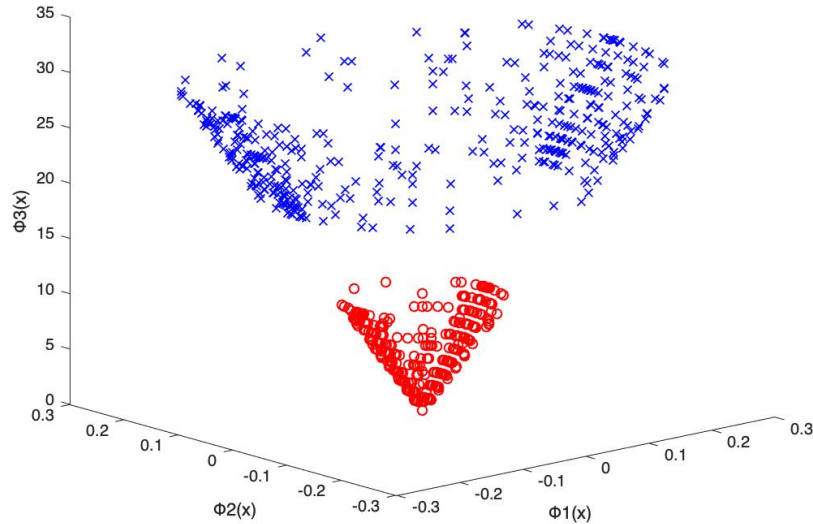
$$(\phi(\mathbf{x}_1), y_1), (\phi(\mathbf{x}_2), y_2), \dots, (\phi(\mathbf{x}_N), y_N) \in \dim(\mathcal{H}) \times \mathbb{R}. \quad (3.33)$$

Nesse espaço transformado, a regressão linear pode ser construída como

$$\hat{y}(\mathbf{x}_n) = f(\mathbf{x}_n) = \mathbf{w}^T \phi(\mathbf{x}_n) + b, \quad (3.34)$$

onde  $f(\mathbf{x}_n) \in \mathbb{R}$ ,  $\mathbf{x}_n \in \mathbb{R}^D$  (para  $n = 1, \dots, N$ ),  $\mathbf{w}$  é o vetor de peso,  $\dim(\mathbf{w}) = \dim(\mathcal{H})$ ,  $b \in \mathbb{R}$  e  $\phi(\cdot) : \mathcal{X} \rightarrow \mathcal{H}$ . O desenvolvimento do modelo MVS começa com a minimização da seguinte função objetivo, também conhecida como fórmula primária

$$C(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 + c \sum_{n=1}^N (\xi_n + \xi_n^*), \quad (3.35)$$

Figura 19 – Exemplo do mapeamento no  $\mathbb{R}^3$ 

Fonte: Próprio autor.

onde

$$f(\mathbf{x}_n) - \mathbf{w}^T \phi(\mathbf{x}_n) - b \leq \varepsilon + \xi_n, \quad (3.36)$$

$$-f(\mathbf{x}_n) + \mathbf{w}^T \phi(\mathbf{x}_n) + b \leq \varepsilon + \xi_n^*, \quad (3.37)$$

$$\xi_n, \xi_n^* \geq 0, \quad (3.38)$$

para  $n = 1, \dots, N$ . A constante de regularização  $c > 0$  regula a penalidade a ser imposta às amostras que estão fora da margem ( $\varepsilon$ ) e é usada para evitar sobreajuste. As variáveis de folga  $\{\xi_n, \xi_n^*\}$  geram algum relaxamento nas restrições. A função objetivo na Equação 3.35 e as restrições correspondentes podem ser reescritas como a seguinte função lagrangiana:

$$\begin{aligned} L(\mathbf{w}, b, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*) = & \frac{1}{2} \|\mathbf{w}\|^2 + c \sum_{n=1}^N (\xi_n + \xi_n^*) - \sum_{n=1}^N \alpha_n (\mathbf{w}^T \phi(\mathbf{x}_n) + b + \varepsilon + \xi_n - y_n) \\ & - \sum_{n=1}^N \alpha_n^* (-\mathbf{w}^T \phi(\mathbf{x}_n) - b + \varepsilon + \xi_n^* + y_n) - \sum_{n=1}^N (\eta_n \xi_n + \eta_n^* \xi_n^*), \end{aligned} \quad (3.39)$$

onde  $\alpha_n, \alpha_n^*, \eta_n, \eta_n^*$  são múltiplos positivos de Lagrange.

De acordo com Karush-Kuhn-Tucker, o ponto ótimo da otimização convexa é obtido fazendo-se as derivadas parciais do lagrangiano  $L()$  em relação a  $\mathbf{w}$ ,  $b$ ,  $\xi_n$ , e  $\xi_n^*$  e, em seguida,

igualando-as todas a zero (HAYKIN, 2009). Dessa forma, temos

$$\frac{\partial L(\mathbf{w}, b, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*)}{\partial \mathbf{w}} = \mathbf{w} - \sum_{n=1}^N (\alpha_n - \alpha_n^*) \phi(\mathbf{x}_n) = \mathbf{0} \Rightarrow \mathbf{w} = \sum_{n=1}^N (\alpha_n - \alpha_n^*) \phi(\mathbf{x}_n), \quad (3.40)$$

$$\frac{\partial L(\mathbf{w}, b, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*)}{\partial b} = \sum_{n=1}^N (\alpha_n - \alpha_n^*) = 0, \quad (3.41)$$

$$\frac{\partial L(\mathbf{w}, b, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*)}{\partial \xi_n} = 0 \Rightarrow c - \alpha_n - \eta_n = 0, \quad (3.42)$$

$$\frac{\partial L(\mathbf{w}, b, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*)}{\partial \xi_n^*} = 0 \Rightarrow c - \alpha_n^* - \eta_n^* = 0. \quad (3.43)$$

Podemos reescrever a Equação 3.35 substituindo os resultados obtidos nas Equações 3.40 até 3.43, então temos uma função objetivo alternativa que deve ser maximizada

$$\begin{aligned} \max C(\alpha, \alpha^*) = & -\frac{1}{2} \sum_{n,k=1}^N (\alpha_n + \alpha_n^*)(\alpha_k + \alpha_k^*) \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) - \varepsilon \sum_{n=1}^N (\alpha_n + \alpha_n^*) \\ & - \varepsilon \sum_{n=1}^N (\alpha_n + \alpha_n^*) + \sum_{y=1}^N y_n (\alpha_n - \alpha_n^*). \end{aligned} \quad (3.44)$$

A Equação 3.44 é conhecida como a formulação dual do problema de otimização da MVS. Neste caso, os parâmetros a serem estimados são os coeficientes de Lagrange  $\alpha$  e  $\alpha^*$ . Os coeficientes  $M$ ,  $M \ll N$ , com valores diferentes de zero são aqueles associados aos vetores de suporte. Apenas  $M$  das  $N$  amostras de treinamento são usadas para construir o modelo, o regressor baseado em MVS é considerado um modelo esparsos.

Após os coeficientes de Lagrange  $\alpha$  e  $\alpha^*$  serem calculados, pode-se representar  $f(\mathbf{x})$  como

$$\hat{y}(\mathbf{x}) = f(\mathbf{x}) = \sum_{n=1}^M (\alpha_n - \alpha_n^*) \phi(\mathbf{x})^T \phi(\mathbf{x}_n) + b. \quad (3.45)$$

No entanto, o mapeamento não linear  $\phi(\mathbf{x}_n)$  é desconhecido e a Equação 3.45 não pode ser usada para predição em aplicações práticas de regressão. A solução depende do conhecido *kernel trick* em que o cálculo do produto escalar  $\phi(\mathbf{x})^T \phi(\mathbf{x}_n)$  é substituído pelo cálculo da função kernel  $k(\mathbf{x}_i, \mathbf{x}_j)$ ,  $\forall i, j = 1, \dots, M$ . Qualquer função contínua e simétrica  $k(\mathbf{x}_i, \mathbf{x}_j)$  que satisfaça o teorema de Mercer (MERCER, 1909) pode ser usado para representar o produto

Tabela 1 – Exemplo de funções de kernel.

Kernel	Expressões	Hiperparâmetros
Linear	$k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$	-
Polinomial	$k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + a)^b$	$a, b$
Gaussiana	$k(\mathbf{x}_i, \mathbf{x}_j) = \exp \left\{ -\frac{\ \mathbf{x}_i - \mathbf{x}_j\ ^2}{2\sigma^2} \right\}$	$\sigma$
Sigmóide	$k(\mathbf{x}_i, \mathbf{x}_j) = \tanh(k_1 \mathbf{x}_i^T \mathbf{x}_j + k_2)$	$k_1, k_2$

escalar dos vetores  $\mathbf{x}_i$  e  $\mathbf{x}_j$  no espaço de característica  $\mathcal{H}$ . A função kernel linear, polinomial, gaussiana e sigmóide estão entre as mais utilizadas (HAYKIN, 2009). As descrições destas funções estão sendo apresentadas na Tabela 1.

Devido ao truque do kernel, a Equação 3.45 pode ser reescrita como

$$\hat{y}(\mathbf{x}) = f(\mathbf{x}) = \sum_{n=1}^M (\alpha_n - \alpha_n^*) k(\mathbf{x}, \mathbf{x}_i) + b. \quad (3.46)$$

Neste trabalho o algoritmo MVS foi implementado no MATLAB (R2015b) e a função *fitrsvm* foi utilizada para treinar o algoritmo, conforme apresentado no algoritmo 4.

---

#### Algoritmo 4: Máquina de Vetores de Suporte

---

**Entrada:** Dados, tipo de função kernel

**Saída:** Variável predita e coeficiente de determinação

**início**

**para**  $i=1:100$  **faça**

    A separação aleatória dos dados de treino e teste;

    Normalize os dados de entrada;

    Calcule os vetores de suporte e os hiperparâmetros da função kernel utilizando os dados de treino e o tipo de função kernel como entrada da função *fitrsvm*;

    Calcule a variável predita dos dados de teste com os vetores de suporte e os hiperparâmetros da função kernel encontrados;

    Calcule o coeficiente de determinação entre a variável predita e o valor observado;

**fim**

    Faça o cálculo da média, mínimo, desvio dos 100 coeficientes de determinação encontrados

**fim**

---

### 3.1.5 Máquinas de Vetores de Suporte por Mínimos Quadrados

A Máquina de Vetores de Suporte por Mínimos Quadrados (MVSMQ) é uma técnica motivada pelo MVS do Vapnik descrito na seção anterior. MVSMQ foi introduzida por Suykens e Vandewalle (1999) como uma ferramenta de aprendizado de máquina com menor custo computacional em relação ao MVS, sem perda de acurácia (SANTOS; BARRETO, 2017). A principal diferença entre os dois modelos é que MVSMQ não é esparso, uma vez que todas as amostras de dados são utilizadas como vetor de suporte. No entanto, técnicas de esparsificação podem ser utilizadas a posteriori para reduzir o número de vetores de suporte (SANTOS; BARRETO, 2017).

Como a MVS, o modelo MVSMQ também implementa uma projeção das amostras de entrada  $\mathbf{x} \in \mathbb{R}^D$  para um espaço de mais alta dimensão  $\mathcal{H}$  e em seguida realizar a regressão linear no espaço transformado. A operação inicia na Eq. (3.34). O problema do MVSMQ consiste em minimizar a função custo, que diferente da técnica MVS é dada por

$$C(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 + \gamma \frac{1}{2} \sum_{n=1}^N e_n^2, \quad (3.47)$$

onde  $\mathbf{w}$  é o vetor de pesos,  $\gamma \gg 0$  é o parâmetro de regularização e  $e_n = y_n - f(\mathbf{x}_n)$  é o erro associado ao  $n$ -ésimo dado. Se compararmos a função objetivo MVS na Eq. (3.35) com a função do modelo MVSMQ na Eq. (3.47), uma grande diferença entre eles é observada na norma das variáveis relacionadas ao erro. Enquanto na Eq. (3.35) a  $L_1$ -norma é usada, na Eq. (3.47)  $L_2$ -norma (ou seja, norma quadrática) é preferida.

A minimização da função objetivo na Eq. (3.47) é realizada com a ajuda do método do multiplicador de Lagrange, a partir do qual a seguinte função Lagrangiana é obtida:

$$L(\mathbf{w}, b, e_n, \alpha_n) = \frac{1}{2} \|\mathbf{w}\|^2 + \gamma \frac{1}{2} \sum_{n=1}^N e_n^2 - \sum_{n=1}^N \alpha_n (\mathbf{w}^T \phi(\mathbf{x}_n) + b + e_n - y_n), \quad (3.48)$$

onde  $y_n$  é a saída associada ao vetor de entrada atual  $\mathbf{x}_n$ ;  $e_n$  é o valor do erro correspondente; e  $\alpha_n$  é o multiplicador de Lagrange associado ao  $n$ -ésimo dado  $\mathbf{x}_n$ .

De acordo com a condição de Karush-Kuhn-Tucker, o valor mínimo do problema de otimização convexa declarado em Eq. (3.48) é obtido fazendo-se as derivadas parciais da função de Lagrange  $\frac{\partial L}{\partial \mathbf{w}}$ ,  $\frac{\partial L}{\partial b}$ ,  $\frac{\partial L}{\partial e_n}$  e  $\frac{\partial L}{\partial \alpha_n}$  e, em seguida, igualando-as todas a zero (HAYKIN, 2009): (HAYKIN, 2009):

$$\frac{\partial L(\mathbf{w}, b, e_n, \alpha_n)}{\partial \mathbf{w}} = \mathbf{w} - \sum_{n=1}^N \alpha_n \phi(\mathbf{x}_n) = \mathbf{0} \Rightarrow \mathbf{w} = \sum_{n=1}^N \alpha_n \phi(\mathbf{x}_n), \quad (3.49)$$

onde usamos um vetor de zeros. Continuando com os cálculos, obtemos

$$\frac{\partial L(\mathbf{w}, b, e_n, \alpha_n)}{\partial b} = \sum_{n=1}^N \alpha_n = 0, \quad (3.50)$$

$$\frac{\partial L(\mathbf{w}, b, e_n, \alpha_n)}{\partial e_n} = \sum_{i=1}^N (\gamma e_n - \alpha_n) = 0 \Rightarrow \alpha_n = \gamma e_n, \quad (3.51)$$

e

$$\frac{\partial L(\mathbf{w}, b, e_n, \alpha_n)}{\partial \alpha_n} = \mathbf{w}^T \phi(\mathbf{x}_n) + b + e_n - y_n = 0 \Rightarrow y_n = \mathbf{w}^T \phi(\mathbf{x}_n) + b + e_n. \quad (3.52)$$

Pode-se reescrever a Eq. (3.52) substituindo os valores de  $\mathbf{w}$  e  $e_n$  com aqueles obtidos das Eqs. (3.49) e (3.51), respectivamente. Desta forma temos

$$y_n = \sum_{n=1}^N \alpha_n \phi(\mathbf{x}_n)^T \phi(\mathbf{x}_n) + b + \frac{\alpha_n}{\gamma}, \quad (3.53)$$

$$= \sum_{n=1}^N \alpha_n k(\mathbf{x}_n, \mathbf{x}_n) + b + \frac{\alpha_n}{\gamma}, \quad (3.54)$$

onde usamos o truque do kernel para obter  $k(\mathbf{x}_n, \mathbf{x}_n) = \phi(\mathbf{x}_n)^T \phi(\mathbf{x}_n)$ . Logo, as Eq. (3.50) e (3.53) podem ser reescritas em forma matricial, respectivamente, como

$$\mathbf{1}_N^T \alpha = 0, \quad (3.55)$$

$$\mathbf{y} = (\mathbf{K} + \gamma^{-1} \mathbf{I}_N) \alpha + b \mathbf{1}_N, \quad (3.56)$$

onde  $\mathbf{I}_N \in \mathbb{R}^{N \times N}$  é a  $N \times N$  matriz identidade,  $\mathbf{1}_N \in \mathbb{R}^N$  é um  $N$ -dimensional vetor de 1's, e  $\mathbf{K} \in \mathbb{R}^{N \times N}$  é a  $N \times N$  matriz de Kernel.

As Equações 3.55 e 3.56 podem ser colocadas juntas em forma matricial

$$\begin{bmatrix} \mathbf{1}_N^T & 0 \\ \mathbf{K} + \gamma^{-1} \mathbf{I}_N & \mathbf{1}_N \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}, \quad (3.57)$$

onde o vetor de coeficiente de Lagrange  $\alpha \in \mathbb{R}^N$  e o bias  $b$  são as soluções do sistema linear. Uma vez que esses parâmetros são calculados, a saída prevista do modelo de regressão MVSMQ para um novo vetor de entrada  $\mathbf{x}$  é dada por

$$\hat{y}(\mathbf{x}) = \hat{f}(\mathbf{x}) = \sum_{n=1}^N \alpha_n k(\mathbf{x}, \mathbf{x}_n) + b. \quad (3.58)$$

Conforme mencionado no início desta seção, os multiplicadores de Lagrange associados às amostras de entrada de treinamento  $\mathbf{x}_n$  são em geral não nulos, levando a um modelo MVSMQ não esparso. Isso significa que todos os vetores de entrada de treinamento são considerados vetores de suporte. Para conjuntos de dados muito grandes, esse recurso pode impor dificuldades para a construção de modelos e o uso de estratégias de redução de dados pode ser necessário (consulte Rocha Neto e Barreto (2013) e suas referências).

Assim como algoritmo MVS, o algoritmo MVSMQ foi implementado no MATLAB (R2015b) com o uso das funções *trainlssvm* e *simlssvm* (Toolbox LSSVM), conforme apresentado no algoritmo 5.

---

**Algoritmo 5:** Máquina de Vetores de Suporte por Mínimos Quadrados

---

**Entrada:** Dados, tipo de função kernel

**Saída:** Variável predita e coeficiente de determinação

**início**

**para**  $i=1:100$  **faça**

    A separação aleatória dos dados de treino e teste;

    Normalize os dados de entrada;

    Calcule os hiperparâmetros da função kernel utilizando os dados de treino e o tipo de função Kernel como entrada da função *trainlssvm*;

    Calcule a variável predita dos dados de teste utilizando os vetores de suporte e os hiperparâmetros da função kernel encontrados como entrada da função ;

    Calcule o coeficiente de determinação entre a variável predita e o valor observado;

**fim**

  Faça o cálculo da média, mínimo, desvio dos 100 coeficientes de determinação encontrados

**fim**

---

### 3.2 Coeficiente de determinação - $R^2$

Os dados foram processados pelos métodos de estimação citados acima e o coeficiente de determinação foi utilizado para avaliar a qualidade dos modelos.

O Coeficiente de determinação ( $R^2$ ) foi medido entre a variável predita e a variável observada, dada por

$$R2 = \frac{SQ_{EXP}}{SQ_{TOTAL}}, \quad (3.59)$$

onde  $SQ_{EXP}$  é a soma dos quadrados explicados e  $SQ_{TOTAL}$  corresponde a soma dos quadrados totais.

A soma dos quadrados explicados,  $SQ_{EXP}$ , corresponde a soma do quadrado da diferença entre o valor predito para cada observação e a média das observações, dada por

$$SQ_{EXP} = \sum_{n=1}^N (\hat{y}_n - \bar{y})^2, \quad (3.60)$$

onde  $\hat{y}_n$  é o valor predito e  $\bar{y}$  é a média das observações.

A soma dos quadrados totais,  $SQ_{TOTAL}$ , corresponde a soma dos quadrados das diferenças entre o valor observado e a média dos valores, dada por

$$SQ_{TOTAL} = \sum_{n=1}^N (y_n - \bar{y})^2, \quad (3.61)$$

onde  $y_n$  é o valor observado.

Desta forma,  $R2$  pode ser escrito como

$$R2 = \frac{\sum_{n=1}^N (\hat{y}_n - \bar{y})^2}{\sum_{n=1}^N (y_n - \bar{y})^2}. \quad (3.62)$$

O coeficiente de determinação pode assumir valores entre 0 e 1, indicando, em percentagem, a proporção da variável dependente do modelo que é explicada pelas variáveis independentes. Quanto mais próximo de 1 maior parte da variável predita está sendo explicada pelo modelo (SPIEGEL; STEPHENS, 2000). Na indústria, utilizando dados reais e com ruídos inerentes ao processo, o  $R2$  acima de 0,5 já é considerado satisfatório.

### 3.3 Conjunto de dados

Os métodos de estimação: Regressão Linear Múltipla, Máquina de Aprendizagem Mínima, Rede Neural Artificial, Máquina de Vetores de Suporte e Máquina de Vetores de Suporte por Mínimos Quadrados foram utilizadas para estimar o Limite de Escoamento, Limite de Resistência à Tração, Razão LRT/LE e o Percentual de Alongamento a partir de informações sobre a composição química do aço e as variáveis termomecânicas de processamento coletadas em uma usina siderúrgica. Ao todo, este estudo resultou em 20 análises.



A Tabela 2 fornece informações sobre os parâmetros estudados referentes a temperatura de laminação inicial, temperatura de entrada e de saída do tratamento térmico, ao tamanho da área de seção transversal do material e à pressão e vazão da água utilizada no tratamento térmico. O controle das temperaturas de laminação inicial permite uma melhor compreensão do grau de refinamento do grão austenítico formado no aço quando submetido a temperaturas acima da temperatura de recristalização do material, o que é um importante indicador das propriedades mecânicas finais do aço. Uma pequena área de seção transversal requer um fluxo mais baixo de pressão da água do tratamento térmico para a criação de uma camada externa martensítica, o que causa alterações nas propriedades mecânicas do vergalhão. Os diâmetros dos vergalhões estudados foram 8 mm, 10 mm e 12 mm, 1/2 "(12,7 mm), 3/8"(9,525 mm) e 5/8 "(15,875 mm).

Tabela 2 – Parâmetros da laminação do tratamento térmico do vergalhão do aço.

Variáveis de Entrada	Mínimo	Máximo	Média	Desvio Padrão
Área da secção transversal [ $mm^2$ ]	50	199	93,2	46,5
Temperatura do trem de laminação inicial [ $^{\circ}C$ ]	971,8	1081,5	1032,4	18,8
Temperatura de entrada do tratamento térmico [ $^{\circ}C$ ]	867,0	961,8	921,1	18,4
Temperatura de saída do tratamento térmico [ $^{\circ}C$ ]	337,6	584,1	436,8	47,6
Pressão de água do tratamento térmico [Bar]	0	15,5	5,8	5,0
Fluxo da água do tratamento térmico [ $m^3/h$ ]	0	68,2	29,2	23,7

Em relação à composição química, foram considerados elementos que possuem influências nas propriedades mecânicas finais do produto, conforme discutido na seção 3.2. A análise metalúrgica dos parâmetros foi aplicada a fim de otimizar o desempenho dos métodos utilizados. Na Tabela 3 são mostrados os valores mínimo, máximo, médio e desvio padrão da composição química do conjunto de dados estudado nesta pesquisa. Os teores de Nióbio e Vanádio apresentados na Tabela 3, são pequenos para apresentar os efeitos descritos nas seções 2.2.5 e 2.2.9, porém a remoção desses parâmetros piorou a previsão das propriedades mecânicas, sendo necessário utilizar como entrada do modelo todos os 12 dados de composição química.

Os algoritmos estimaram, a partir dos 18 dados de entrada, as 4 propriedades mecânicas mostradas na Tabela 4. A tabela fornece informações sobre as faixas, médias e desvios-padrões dos parâmetros estudados. Os dados estudados possuem um alto desvio padrão, o que indica que aços com diferentes propriedades e aplicações foram analisados neste trabalho.

Para todas as propriedades mecânicas e para cada algoritmo estudado, 80 % dos 1300 pares de entrada/saída disponíveis (1040 pares) foram escolhidos aleatoriamente para treinar o modelo, enquanto 20 % dos pares restantes (260 pares) foram utilizados para testar o modelo. O

Tabela 3 – Variáveis de entrada relacionadas à composição química [%].

Variáveis de Entrada	Mínimo	Máximo	Média	Desvio Padrão
Carbono	0,156	0,474	0,3216	0,039
Nióbio	0,000	0,008	0,0026	0,001
Fosforo	0,002	0,013	0,0064	0,002
Silício	0,099	0,208	0,1567	0,021
Molibdênio	0,006	0,029	0,0111	0,002
Cobre	0,187	0,445	0,2814	0,040
Manganês	0,456	1,149	0,9001	0,229
Vanádio	0,002	0,019	0,0081	0,006
Enxofre	0,020	0,050	0,0370	0,006
Cromo	0,056	0,271	0,1314	0,036
Níquel	0,049	0,143	0,0745	0,011
Estanho	0,001	0,040	0,0193	0,003

Tabela 4 – Propriedades mecânicas do vergalhão previstas pelos modelos.

Variáveis Previstas	Mínimo	Máximo	Média	Desvio Padrão
Limite de escoamento [MPa]	421	661	521,7	41,6
Limite de resistência à tração [MPa]	596	914	682,8	35,1
Razão do LRT/LE	1,12	1,55	1,314	0,072
Alongamento Percentual [%]	8	20	15,1	1,897

coeficiente de determinação foi utilizado para avaliar a qualidade dos dados estimados para cada modelo.

### 3.4 Resumo do Capítulo

Neste capítulo discutiu-se sobre as técnicas de aprendizado de máquina e o conjunto de dados que foram utilizadas neste trabalho. No próximo capítulo, os resultados obtidos estão sendo apresentados e discutidos.

## 4 RESULTADOS E DISCUSSÕES DAS SIMULAÇÕES NUMÉRICAS

Nesse capítulo serão apresentados e discutidos os resultados obtidos na previsão de quatro propriedades mecânicas do vergalhão, são elas: LE, LRT, Razão do LRT/LE e AP.

Para executar os experimentos numéricos relatados a seguir, usamos o Matlab R2015b rodando em um MacBook Pro com macOS Mojave 10.14.6, Intel Core i5, processador de 2,4 Ghz e 4 GB de RAM. Para o modelo baseado em MVS, usamos o *Statistics and Machine Learning Toolbox*, em particular o comando `fitrsvm`. Para o modelo baseado em MVSMQ, usamos a *LS-SVMlab toolbox*, que está disponível publicamente para download em <https://www.esat.kuleuven.be/sista/lssvmlab/>. Para o modelo de ARL, MAM e RNA, foi desenvolvido o código do zero.

### 4.1 Análise de Regressão Linear

O LE foi a primeira das propriedades mecânicas a ser analisada pela regressão linear múltipla. O teste de significância mostrou que os coeficientes relacionados aos teores de cobre, níquel, fósforo e estanho não são estatisticamente significativos e que não estão linearmente correlacionados com o LE. Por outro lado, o teste revelou que os coeficientes associados aos teores de carbono, manganês e vanádio, a temperatura de entrada no tratamento térmico e a área da seção transversal estão fortemente correlacionados ao LE. Esse resultado é consistente com os relatados por Singh et al (SINGH *et al.*, 1998).

Os efeitos do carbono e manganês no ferro são bem conhecidos e documentados. Carbono e manganês formam soluções sólidas de ferro, formando, respectivamente, solução sólida intersticial e substitucional em ferro. Além disso, o manganês favorece a formação de carboneto, ajudando a criar uma microestrutura por meio de têmpera.

Amin e Pickering (1982) relataram que os carbonitreto de nióbio e vanádio não apenas evitam o engrossamento dos grãos de austenita durante o aquecimento, mas também ajudam a refinar a estrutura dos grãos de austenita durante a laminação a quente, fixando os limites dos grãos e retardando a recristalização. Os aços analisados nesse trabalho tiveram adição de vanádio menor do que 200 ppm, e embora esses teores não apresentem melhoras relevantes nas propriedades mecânicas, o coeficiente associado ao teor de vanádio está fortemente correlacionado ao LE.

Após remover as entradas relacionadas às quatro entradas estatisticamente insignifi-

cantes (cobre, níquel, fósforo, estanho), o valor estimado do LE foi recalculado para 1000 opções distintas de conjuntos de treinamento e teste. O coeficiente de determinação entre o LE previsto e seu valor observado nos dados de teste atingiu um máximo de 0,6291. A média calculada nas 1000 simulações foi de 0,5072 com um desvio padrão de 0,0515. Os resultados são mostrados na Tabela 5.

Tabela 5 – Análise de regressão linear para as propriedades mecânicas do vergalhão após 1000 execuções de treino/teste.

Variáveis Preditas	R2 Máximo	R2 médio	R2 desvio
<b>Limite de escoamento (LE)</b>	0,6291	0,5072	0,0515
<b>Limite de resistência à tração (LRT)</b>	0,3660	0,2398	0,0440
<b>Razão LRT/LE</b>	0,8255	0,7618	0,0289
<b>Alongamento Percentual (AP)</b>	0,6066	0,4950	0,0424

O LRT foi a segunda propriedade mecânica estimada pela regressão linear múltipla. O teste de significância mostrou que a temperatura de saída do tratamento térmico, os teores de molibdênio, nióbio, enxofre e silício, o fluxo de água usado no tratamento térmico, a temperatura do trem de laminação inicial, o teor de carbono, a temperatura de entrada do tratamento térmico, a área da seção transversal, o conteúdo de cromo, manganês e vanádio estão, nesta ordem crescente, fortemente correlacionados com o limite resistência à tração.

Os efeitos do carbono, manganês, nióbio e vanádio foram explicados anteriormente. A área de seção transversal está relacionada a pressão da água do tratamento térmico para a criação de uma camada externa martensítica, o que também causa alterações nas propriedades mecânicas do vergalhão. Quanto menor esta área, menor é a pressão necessária para formar uma superfície martensítica. Esta propriedade exerce forte influência no LRT.

Por outro lado, o teste de significância mostrou novamente que o conteúdo de cobre, níquel, fósforo e estanho, bem como a pressão da água usada no tratamento térmico, não estão significativamente correlacionada com essa propriedade. Após remover as cinco entradas mencionadas acima, a análise de regressão foi repetida para 1000 conjuntos de dados escolhidos aleatoriamente.

O coeficiente máximo de determinação entre os valores previstos e observados do LRT foi de 0,3660. A média calculada sobre 1000 conjuntos de dados foi de 0,2498 com um desvio padrão de 0,0440. Esses resultados mostram que a regressão linear não é capaz de prever satisfatoriamente o LRT. Os resultados também são mostrados na Tabela 5.

Em seguida, a regressão linear múltipla foi utilizada para prever a razão do LRT/LE. O teste de significância realizado nos coeficientes obtidos pela regressão mostrou que os teores de fósforo, cobre, enxofre, nióbio, cromo, silício, níquel e estanho, além da área de seção transversal, não são estatisticamente significantes para estimar a Razão LRT/LE. Todos os outros coeficientes estão fortemente correlacionados, principalmente aqueles relacionados aos teores de carbono e manganês, à pressão da água utilizada no tratamento térmico e à temperatura de saída do tratamento térmico.

Após remover as entradas estatisticamente insignificantes, a análise de regressão foi repetida para 1000 conjuntos de dados escolhidos aleatoriamente. O coeficiente médio de determinação e o desvio padrão calculado em 1000 conjuntos aleatórios de testes foram 0,7618 e 0,0289, respectivamente. O coeficiente máximo de determinação entre os valores previstos e medidos da relação LRT/LE alcançada foi de 0,8255.

Por fim, a análise de regressão linear múltipla foi usada para prever o AP do aço. O teste de significância foi realizado e mostrou que os coeficientes associados à área da seção transversal, aos teores de fósforo, cobre, enxofre, nióbio, cromo, silício e níquel não estão linearmente correlacionados com o AP. Os teores de carbono, manganês, vanádio e molibdênio e todos as temperaturas estão fortemente correlacionados.

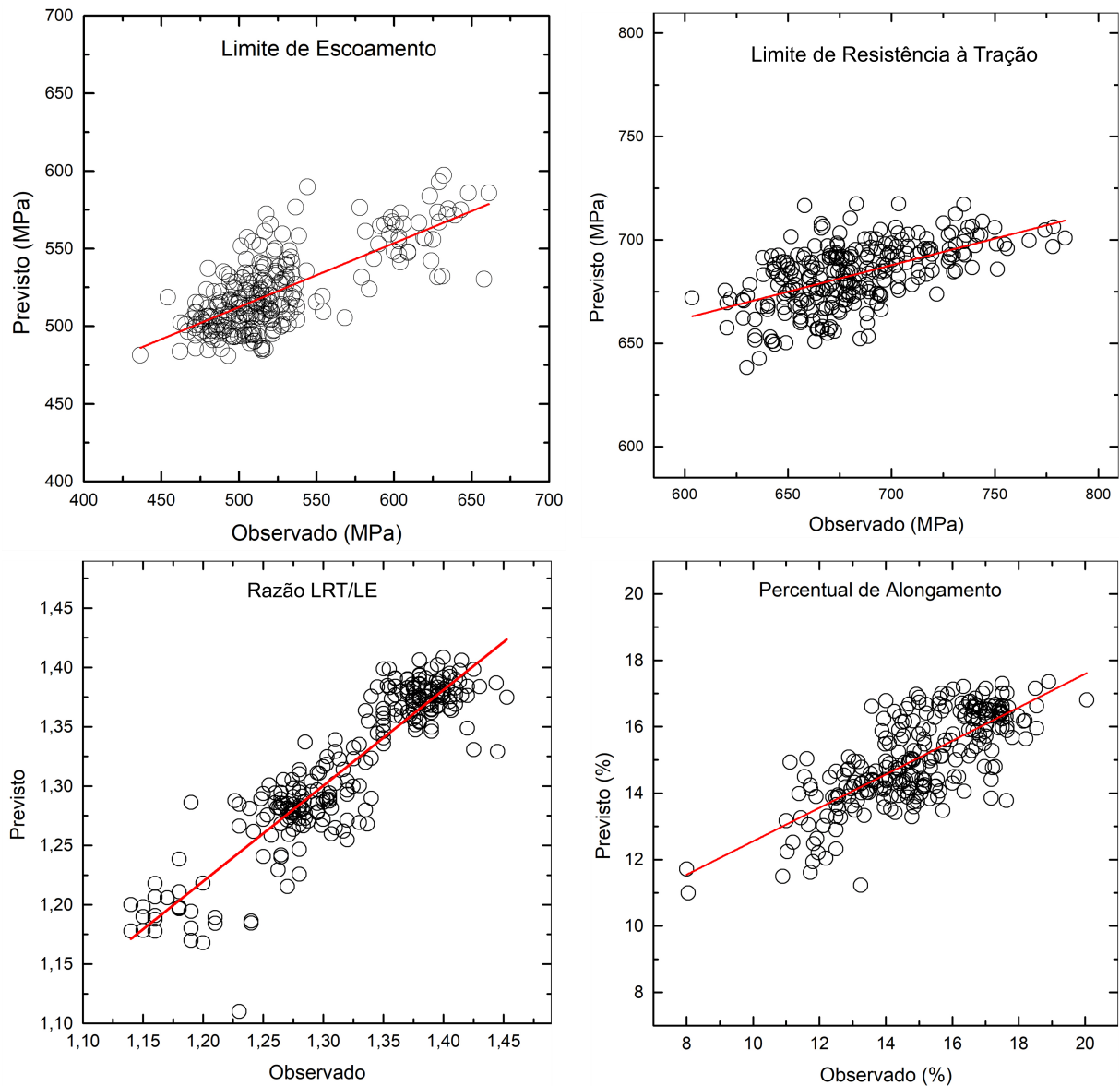
A análise de regressão foi repetida apenas com as entradas significativas para 1000 conjuntos de dados escolhidos aleatoriamente. O coeficiente máximo de determinação entre os valores previstos e medidos de alongamento, a média e o desvio padrão calculados nos 1000 conjuntos de dados foram 0,6066, 0,4950 e 0,0424, respectivamente. A tabela 5 resume os resultados obtidos pela análise de regressão linear múltipla.

A Figura 20 fornece uma representação gráfica do coeficiente de determinação para regressão linear múltipla. Quanto maior o coeficiente, mais próximos os dados estão da linha de regressão ajustada e maior a porcentagem dos dados que é explicada pelo modelo.

## **4.2 Máquina de Aprendizagem Mínima**

A MAM também foi utilizada para prever as propriedades mecânicas do vergalhão de aço. Foram realizados 100 experimentos para cada uma das propriedades com  $K$  pontos de referência variando de 20 até 1040 com incremento de 20. Para cada experimento, foram separados aleatoriamente 80 % (1040) dos dados para treinar e 20 % (260) dos dados para testar o algoritmo. O valor máximo e a média do  $R^2$  foram avaliados em 100 escolhas aleatórias dos

Figura 20 – Previsão das propriedades mecânicas pela ARL.



dados de teste e estão sendo apresentados nas Figuras 21, 22, 23 e 24.

Para o LE, o melhor resultado em termos de valores médios e máximos de  $R^2$  foi obtido para  $K = 980$  pontos de referência: (média, máximo) = (0,8022, 0,8838). Para a variável de saída LRT, os melhores resultados foram obtidos para  $K = 960$  pontos de referência: (média, máximo) = (0,5416, 0,7849). A previsão da razão LRT/LE produziu os maiores valores de  $R^2$  entre as quatro variáveis de saída de interesse. Para esta saída, um modelo de regressão baseado em MAM com  $K = 940$  alcançado (média, máximo) = (0,8426, 0,9068). Finalmente, para a variável de saída AP, um modelo de regressão baseado em MAM com  $K = 1000$  alcançado (média, máximo) = (0,6679, 0,7929).

A tabela 15 apresenta o  $R^2$  obtidos para cada uma das propriedades depois das 100

Figura 21 – Previsão da limite de escoamento pela MAM

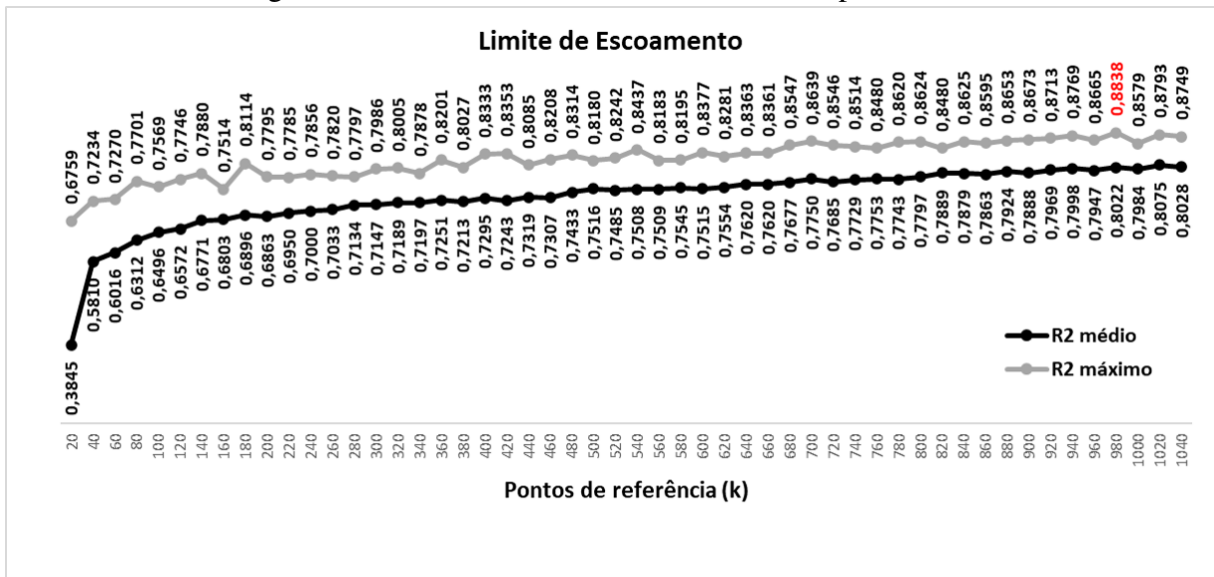
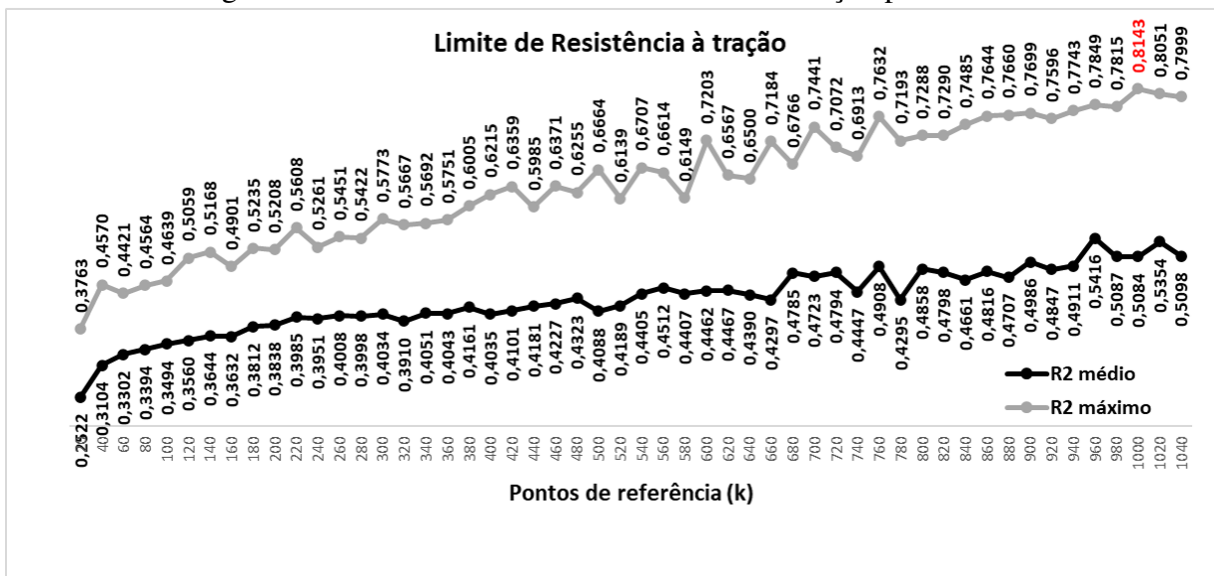


Figura 22 – Previsão da limite de resistência à tração pela MAM



execuções de treino/teste. O  $R^2$  máximo e médio obtido com MAM foi consideravelmente melhor do que o obtido com Análise de Regressão Linear para todas as quatro propriedades estudadas. Este resultado já era esperado devido a regressão conseguir extrair apenas características lineares enquanto a máquina de aprendizagem mínima consegue extrair características lineares e não lineares do modelo.

Quanto maior o número de pontos de referência, maior é o esforço computacional e o conjunto de dados de armazenamento usados para prever a propriedade mecânica. Nesta pesquisa, os melhores resultados foram obtidos utilizando cerca de 1000 pontos de referência, o que não foi um problema para este modelo. No entanto, dependendo do tamanho do conjunto de dados estudados, o uso de muitos pontos de referência pode se tornar uma desvantagem da

Figura 23 – Previsão da Razão LRT/LE pela MAM

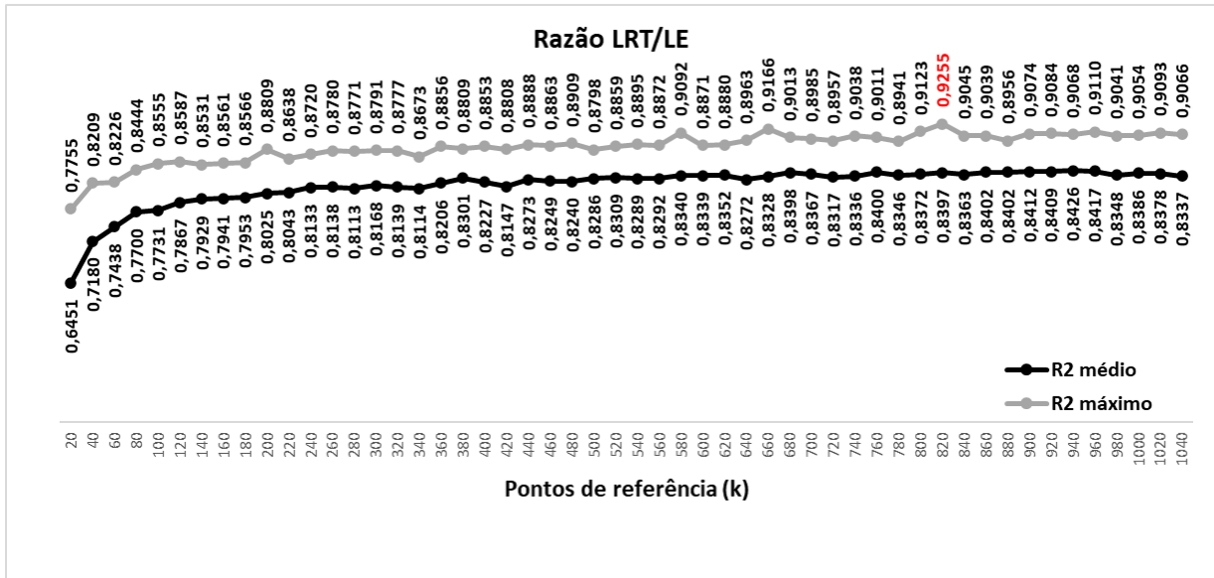
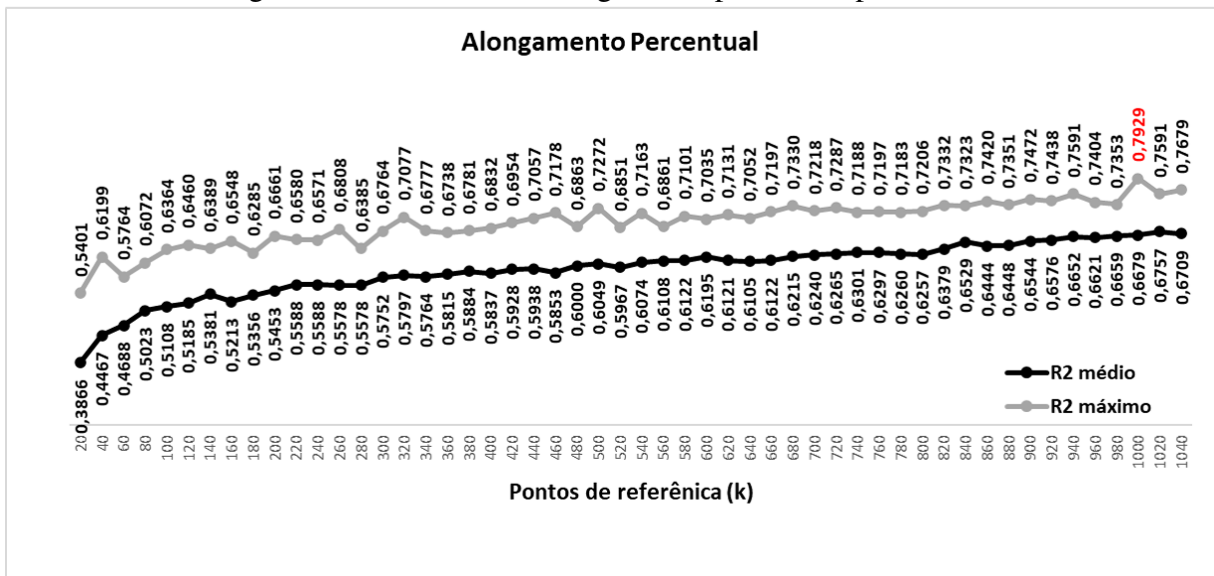


Figura 24 – Previsão do alongamento percentual pelo MAM



MAM. A Figura 25 fornece uma representação gráfica do coeficiente de determinação entre os valores alvo e cada uma das quatro propriedades mecânicas estimadas pelo MAM. Os valores previstos usando a MAM estão mais próximos da linha de regressão ajustada do que aqueles previstos pela Análise de Regressão Linear.

### 4.3 Rede neural de múltiplas camadas

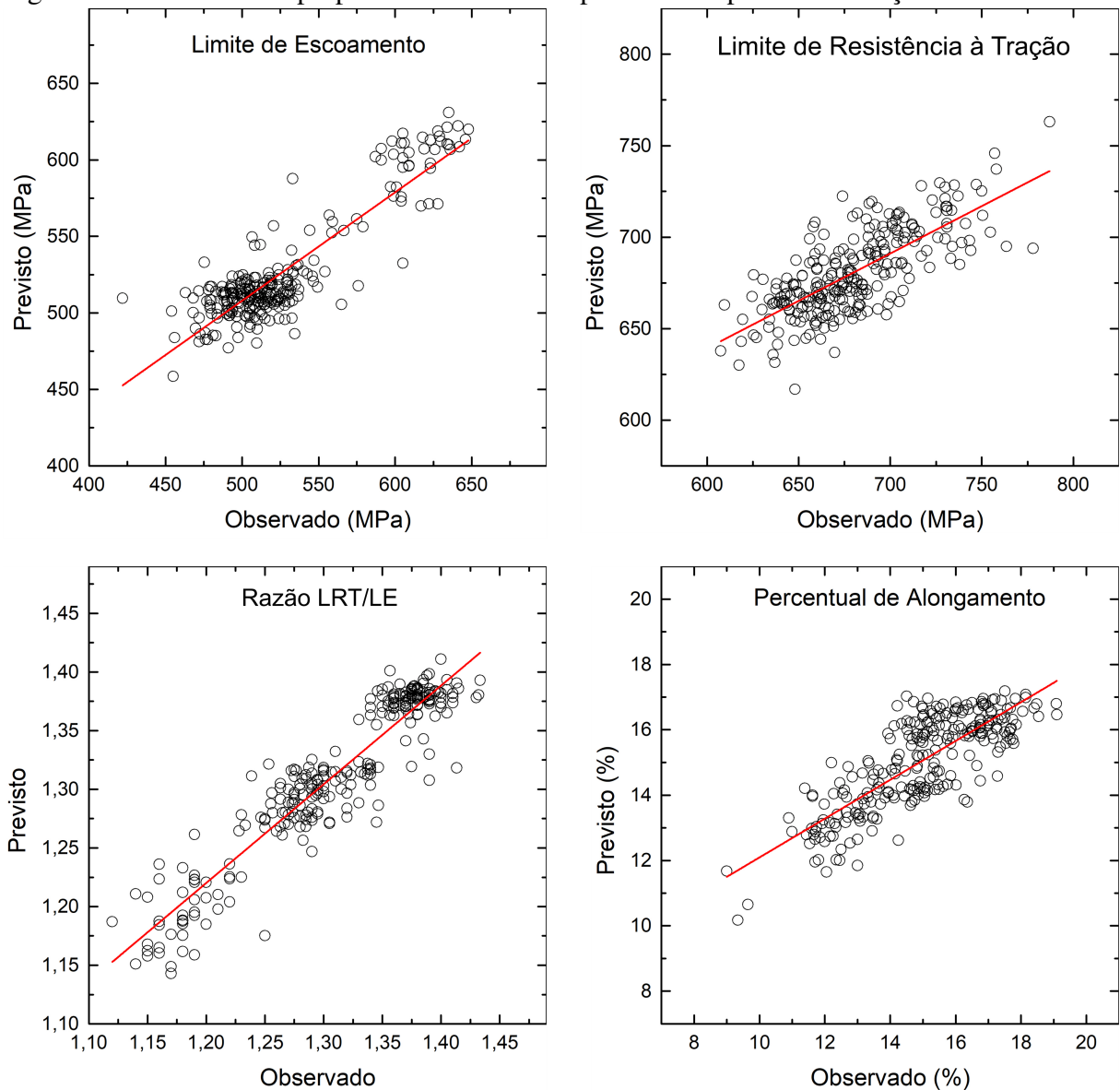
A RNA também foi utilizada para prever as propriedades mecânicas do vergalhão de aço. Grande parte do tempo deste trabalho foi dedicado à determinação da arquitetura da rede neural capaz de produzir a melhor estimativa para cada propriedade mecânica estudada.



Tabela 6 – Estatísticas descritivas do desempenho da MAM após 100 execuções de treinamento/teste para dois valores diferentes do hiperparâmetro  $K$  (número de pontos de referência). Os melhores resultados são mostrados em negrito.

Variáveis Preditas	$K$	$R2$ médio	$R2$ máx	$R2$ mín	$R2$ desvio
<b>Limite de escoamento (LE)</b>	980	<b>0,8022</b>	<b>0,8838</b>	0,6595	0,0368
	1020	0,8075	0,8793	0,6862	0,0365
<b>Limite de resistência à tração (LRT)</b>	960	<b>0,5416</b>	<b>0,7849</b>	-0,0103	0,2165
	1000	0,5084	0,8143	-0,0062	0,2398
<b>Razão LRT/LE</b>	820	0,8397	0,9255	0,7174	0,0359
	940	<b>0,8426</b>	<b>0,9068</b>	0,7575	0,0317
<b>Alongamento Percentual (AP)</b>	1000	<b>0,6679</b>	<b>0,7929</b>	0,4400	0,0595
	1020	0,6757	0,7591	0,4560	0,0441

Figura 25 – Previsão das propriedades mecânicas pela MAM após 100 execuções de treino/teste.



Nesta etapa, foram avaliadas todas as possíveis arquiteturas de rede com duas camadas ocultas, variando o número de neurônios da primeira e da segunda camada oculta de 3 a 15. Isso resultou em 169 diferentes arquiteturas do perceptron multicamadas.

Para cada uma das 169 arquiteturas possíveis, foram avaliados os efeitos dos seis parâmetros de rede no coeficiente de determinação: funções de ativação das camadas ocultas {Tangente Hiperbólica (Tanh); Sigmoid Logística (Sgm)}, constantes das funções de ativação  $\{a; b; c\}$  (Equações 3.26, 3.27 e 3.29) e constante de momento  $\alpha \{0,90; 0,95\}$ .

A função de ativação do neurônio de saída e a taxa de aprendizado  $\eta$  foram definidas como linear e 0,01, respectivamente, para todas as configurações de redes testadas. A combinação dos seis parâmetros de rede mencionados acima resultou em 16 configurações.

A combinação de 169 arquiteturas diferentes e 16 configurações gerou 2704 experimentos para cada propriedade mecânica. Portanto, os experimentos foram numerados de 1 a 2704 para melhor organização. Os números na primeira coluna das tabelas 7 a 14 são para identificação desses experimentos. Os valores q1-q2-m na segunda coluna correspondem, respectivamente, aos números de neurônio da primeira e segunda camadas ocultas e camada de saída.

Assim como na regressão linear múltipla, para todas as propriedades mecânicas e para cada experimento realizado, 80 % (1040) dos 1300 pares de entrada/saída disponíveis foram selecionados aleatoriamente para treinar a rede e os 20 % (260) pares de entrada/saída restante testaram o algoritmo. O máximo, a média e o desvio padrão do  $R^2$  foram avaliados em 100 escolhas aleatórias dos dados de teste.

Os resultados foram divididos em dois grupos: melhores  $R^2$  máximos e melhores  $R^2$  médios. As tabelas 7, 8, 9 exibem as arquiteturas e parâmetros da rede neural que atingiram o  $R^2$  máximo para todas as quatro propriedades mecânicas. No entanto, é importante para o modelo não apenas alcançar um  $R^2$  alto, mas também o maior coeficiente de determinação médio possível, com o menor desvio padrão possível. Isso implica em uma melhor generalização da rede neural. Por esse motivo, as Tabelas 11, 12, 13 e 14 mostram as configurações que produziram o melhor  $R^2$  médio.

O coeficiente de determinação entre a Razão LRT/LE prevista e seu valor observado para o conjunto de dados de teste atingiu um máximo de 0,9170. A média calculada em mais de 100 simulações foi de 0,8413, com um pequeno desvio padrão de 0,0306. Os resultados são mostrados na Tabela 8.

Tabela 7 – Sumário dos resultados obtidos pela rede neural artificial para o LE classificado pelo  $R^2$  máximo após 100 execuções de treino/teste.

Experimento	Número de neurônios	Função de ativação das camadas Ocultas e de Saída	Parâmetros $a, b$ e $c$	$\alpha$	$R^2$ máx	$R^2$ médio	$R^2$ desvio
2321	14-15-1	Tanh / Tanh	1, 1 e 0	0,95	0,8599	0,7472	0,0540
1613	10-12-1	Sgm / Tanh	1, 1 e 2,2	0,95	0,8560	0,7312	0,0557
1789	11-10-1	Sgm / Tanh	1, 1 e 2,2	0,95	0,8556	0,7242	0,0644
2261	13-14-1	Sgm / Tanh	1, 1 e 1	0,95	0,8530	0,6947	0,0803
1107	8-7-1	Tanh / Tanh	1,7159, 2/3 e 0	0,95	0,8524	0,7195	0,0624

Tabela 8 – Sumário dos resultados obtidos pela rede neural artificial para o LRT classificado pelo  $R^2$  máximo após 100 execuções de treino/teste.

Experimento	Número de neurônios	Função de ativação das camadas Ocultas e de Saída	Parâmetros $a, b$ e $c$	$\alpha$	$R^2$ máx	$R^2$ v médio	$R^2$ desvio
2641	15-12-1	Tanh / Tanh	1, 1 e 0	0,95	0,6868	0,4218	0,0815
2273	13-15-1	Tanh / Tanh	1, 1 e 0	0,95	0,6817	0,4454	0,0971
2593	15-9-1	Tanh / Tanh	1, 1 e 0	0,95	0,6768	0,4249	0,0949
2369	14-8-1	Tanh / Tanh	1, 1 e 0	0,95	0,6646	0,4206	0,0870
2305	14-4-1	Tanh / Tanh	1, 1 e 0	0,95	0,6568	0,4233	0,0876

Tabela 9 – Sumário dos resultados obtidos pela rede neural artificial para a razão do LRT/LE classificado pelo  $R^2$  máximo após 100 execuções de treino/teste.

Experimento	Número de neurônios	Função de ativação das camadas Ocultas e de Saída	Parâmetros $a, b$ e $c$	$\alpha$	$R^2$ máx	$R^2$ médio	$R^2$ desvio
993	7-13-1	Tanh / Tanh	1, 1 e 0	0,95	0,9170	0,8413	0,0306
17	11-9-1	Tanh / Tanh	1, 1 e 0	0,90	0,9123	0,8280	0,0315
785	6-13-1	Tanh / Tanh	1, 1 e 0	0,95	0,91183	0,83584	0,0369
847	7-3-1	Sgm / Tanh	1,7159, 2/3 e 2,2	0,95	0,9117	0,8339	0,0303
2369	14-8-1	Tanh / Tanh	1, 1 e 0	0,95	0,9109	0,8428	0,0289

Tabela 10 – Sumário dos resultados obtidos pela rede neural artificial para o AP classificado pelo  $R^2$  máximo após 100 execuções de treino/teste.

Experimento	Número de neurônios	Função de ativação das camadas Ocultas e de Saída	Parâmetros $a, b$ e $c$	$\alpha$	$R^2$ máx	$R^2$ médio	$R^2$ desvio
897	7-7-1	Tanh / Tanh	1, 1 e 0	0,95	0,7436	0,5847	0,0545
1507	10-6-1	Tanh / Tanh	1,7159, 2/3 e 0	0,95	0,7406	0,5692	0,0619
1535	10-7-1	Sgm / Tanh	1,7159, 2/3 e 2,2	0,95	0,7403	0,5640	0,0717
2577	15-8-1	Tanh / Tanh	1, 1 e 0	0,95	0,7388	0,5978	0,0636
849	7-4-1	Tanh / Tanh	1, 1 e 0	0,95	0,7328	0,5841	0,0694

O coeficiente de determinação obtido para o LE foi o segundo melhor para todas as análises, seguido do AP. O pior desempenho obtido foi o do LRT.

Singh *et al.* (1998) mostrou em um estudo com redes neurais a influência de elementos de liga como carbono, manganês, vanádio e nióbio na predição de algumas propriedades do aço comercial. Duas conclusões principais deste estudo são coerentes com o presente trabalho: o LRT é mais influenciado pela adição de elementos de liga e resíduos do que o LE; e o LRT apresenta maiores erros de previsão quando comparado ao LE a partir da variação desses elementos. Assim, pode-se dizer que pequenas variações na composição química do material em estudo

podem levar a maiores variações no LRT quando comparado ao LE, tornando o LRT mais difícil de prever do que o LE.

Vale ressaltar que o aumento da taxa de deformação aumenta a tensão necessária para qualquer deformação plástica adicional. O efeito da taxa de deformação na tensão necessária para produzir alguma deformação, com tensão e temperatura fixas, segue uma lei de potência cujo o expoente é a sensibilidade à taxa de deformação. Portanto, o LE será grande se a sensibilidade à taxa de deformação ou a taxa de deformação relativa for alta. Esse expoente pode ser obtido, por exemplo, executando dois testes de tração contínuos a diferentes taxas de deformação e comparando os níveis de tensão na mesma deformação fixa (HOSFORD, 2010). Em outras palavras, o LE depende da taxa de deformação e da sensibilidade à taxa de deformação. Perto da temperatura ambiente, o valor da sensibilidade à taxa de deformação para o aço é baixo e, conseqüentemente, o efeito é pequeno, porém, para trabalhos futuros, pode-se verificar se a adição da sensibilidade à taxa de deformação como parâmetro do modelo melhora ainda mais os resultados.

Para trabalhos futuros, também pode ser adicionado alguma informação sobre a microestrutura do material, já que possui importante influência nas propriedades mecânicas finais.

Tabela 11 – Sumário dos resultados obtidos pela rede neural artificial para o limite de escoamento classificado pelo  $R2$  médio após 100 execuções de treino/teste.

Experimento	Número de neurônios	Função de ativação das camadas Ocultas e de Saída	Parâmetros $a, b$ e $c$	$\alpha$	$R2$ máx	$R2$ médio	$R2$ desvio
2657	15-13-1	Tanh/Tanh	1,1,0	0,95	0,8286	0,7509	0,0435
2689	15-15-1	Tanh/Tanh	1,1,0	0,95	0,8380	0,7503	0,0546
1617	10-13-1	Tanh/Tanh	1,1,0	0,95	0,8333	0,7486	0,0416
2321	14-5-1	Tanh/Tanh	1,1,0	0,95	0,8600	0,7472	0,0540
2561	15-7-1	Tanh/Tanh	1,1,0	0,95	0,8404	0,7455	0,0581

Tabela 12 – Sumário dos resultados obtidos pela rede neural artificial para o Limite de Resistência à tração classificado pelo  $R2$  médio após 100 execuções de treino/teste.

Experimento	Número de neurônios	Função de ativação das camadas Ocultas e de Saída	Parâmetros $a, b$ e $c$	$\alpha$	$R2$ máx	$R2$ médio	$R2$ desvio
2273	13-15-1	Tanh/Tanh	1,1 e 0	0,95	0,6817	0,4455	0,0972
2257	13-14-1	Tanh/Tanh	1,1 e 0	0,95	0,6103	0,4408	0,0823
2689	15-15-1	Tanh/Tanh	1,1 e 0	0,95	0,6347	0,4370	0,0911
2385	14-9-1	Tanh/Tanh	1,1 e 0	0,95	0,6073	0,4327	0,0862
1729	11-7-1	Tanh/Tanh	1,1 e 0	0,95	0,6108	0,4320	0,0783

Os melhores coeficientes de determinação máximos e as melhores médias foram

Tabela 13 – Sumário dos resultados obtidos pela rede neural artificial para a razão do LRT/LE classificado pelo  $R2$  médio após 100 execuções de treino/teste,

Experimento	Número de neurônios	Função de ativação das camadas Ocultas e de Saída	Parâmetros $a, b$ e $c$	$\alpha$	$R2$ máx	$R2$ médio	Desvio $R2$
1473	10-4-1	Tanh/Tanh	1,1 e 0	0,95	0,9006	0,8489	0,0260
1041	8-3-1	Tanh/Tanh	1,1 e 0	0,95	0,8998	0,8488	0,0272
2321	14-5-1	Tanh/Tanh	1,1 e 0	0,95	0,9071	0,8485	0,0272
2001	12-11-1	Tanh/Tanh	1,1 e 0	0,95	0,9002	0,8477	0,0287
2673	15-14-1	Tanh/Tanh	1,1 e 0	0,95	0,9006	0,8470	0,0340

Tabela 14 – Sumário dos resultados obtidos pela rede neural artificial para o alongamento percentual classificado pelo  $R2$  médio após 100 execuções de treino/teste.

Experimento	Número de neurônios	Função de ativação das camadas Ocultas e de Saída	Parâmetros $a, b$ e $c$	$\alpha$	$R2$ máx	$R2$ médio	$R2$ desvio
1985	12-10-1	Tanh/Tanh	1,1 e 0	0,95	0,6978	0,5991	0,0546
1169	8-11-1	Tanh/Tanh	1,1 e 0	0,95	0,7034	0,5981	0,0561
2577	15-8-1	Tanh/Tanh	1,1 e 0	0,95	0,7388	0,5978	0,0636
2385	14-9-1	Tanh/Tanh	1,1 e 0	0,95	0,7032	0,5968	0,0570
1841	11-4-1	Tanh/Tanh	1,1 e 0	0,95	0,7168	0,5962	0,0530

obtidos com a função de ativação do tipo tangente hiperbólica, tanto nas camadas ocultas quanto na camada de saída para todas as quatro propriedades mecânicas estudadas. Esse melhor desempenho provavelmente se deve ao uso de uma função de ativação assimétrica comumente proposta por técnicas heurísticas de otimização de rede neural. Nesse sentido, pode-se supor que 1,7159 e 2/3 são os valores mais apropriados para as constantes  $a$  e  $b$  para a função de ativação do tipo tangente hiperbólica (HAYKIN, 2009).

Embora um desempenho melhor fosse esperado para essa configuração, isso não foi confirmado pelos sinais de teste. Entretanto, tanto o coeficiente máximo de determinação quanto as melhores médias foram obtidas para  $\varepsilon$  igual a 0 e parâmetros  $a, b$  e  $c$  iguais a 1, 1 e 0, respectivamente, para todas as propriedades mecânicas estudadas. Os melhores resultados foram obtidos com constante de momento,  $\alpha$ , igual a 0,95 para todas as propriedades mecânicas estudadas, mas com um número diferente de neurônios em camadas ocultas. As tabelas 11, 12, 13 e 14 mostram os resultados.

O coeficiente de determinação obtido para o LE foi o segundo melhor para todas as análises, seguido pelo AP. O pior desempenho obtido por esse método foi para o LRT.

O modelo previu LRT e LE com erro menor do que  $\pm 5\%$  ( $\pm 10\%$ ) em mais de 84% (98,8%) dos dados e o AP e a Razão LRT / LE com um erro menor do que  $\pm 5\%$  ( $\pm 10\%$ ) em mais de 99% (100%) dos dados.

Quando comparamos a regressão linear múltipla (Tabela 5) com a rede neural (Tabelas 7 até 14), fica claro que os resultados obtidos com a rede neural são melhores para

Tabela 15 – Estatísticas descritivas do desempenho da MVS após 100 execuções de treino/teste.

Variáveis Preditas	R2 médio	R2 máximo	R2 mínimo	R2 desvio	# Vetores de Suporte (média)
<b>Limite de Escoamento (LE)</b>	0,7296	0,8086	0,6450	0,0338	572
<b>Limite de Resistência à tração (LRT)</b>	0,5180	0,6242	0,3487	0,0588	566
<b>Razão LRT/LE</b>	0,8469	0,9002	0,7616	0,0271	483
<b>Alongamento Percentual(AP)</b>	0,6456	0,7332	0,5523	0,0383	560

todas as propriedades estudadas. A melhora foi significativa até para a propriedade mecânica que apresentou mais dificuldade em ser estimada, o LRT. Esses resultados eram esperados, pois a regressão linear pode apenas modelar uma relação linear entre as variáveis explicativas (independentes) e a variável resposta (dependente) (MONTGOMERY *et al.*, 2012), enquanto um perceptron multicamada com uma camada oculta pode mapear quaisquer funções não lineares. É possível observar que os menores desvios padrões do  $R2$  foram obtidos para a Razão LRT/LE, tanto pela rede neural quanto pela regressão linear.

A rede neural de duas camadas atingiu valores de  $R2$  (máximo, médio) = (0,91117, 0,8489) para a razão LRT/LE, valores ligeiramente inferiores aos relatado na Tabela 15 para o modelo MAM. O modelo MAM também teve um desempenho melhor do que o modelo RNA para as outras três variáveis de saída, valores de  $R2$  (máximo, médio) = (0,8599, 0,7509) para a variável LE (máximo, média) = (0,7436, 0,5991) para a variável AP. A principal diferença de desempenho entre os modelos MAM e RNA foi observada para a variável LRT, com o modelo RNA alcançou valores de  $R2$ (máximo, média) = (0,6868, 0,4455), que é um resultado inferior em comparação com aquele alcançado pelo modelo MAM.

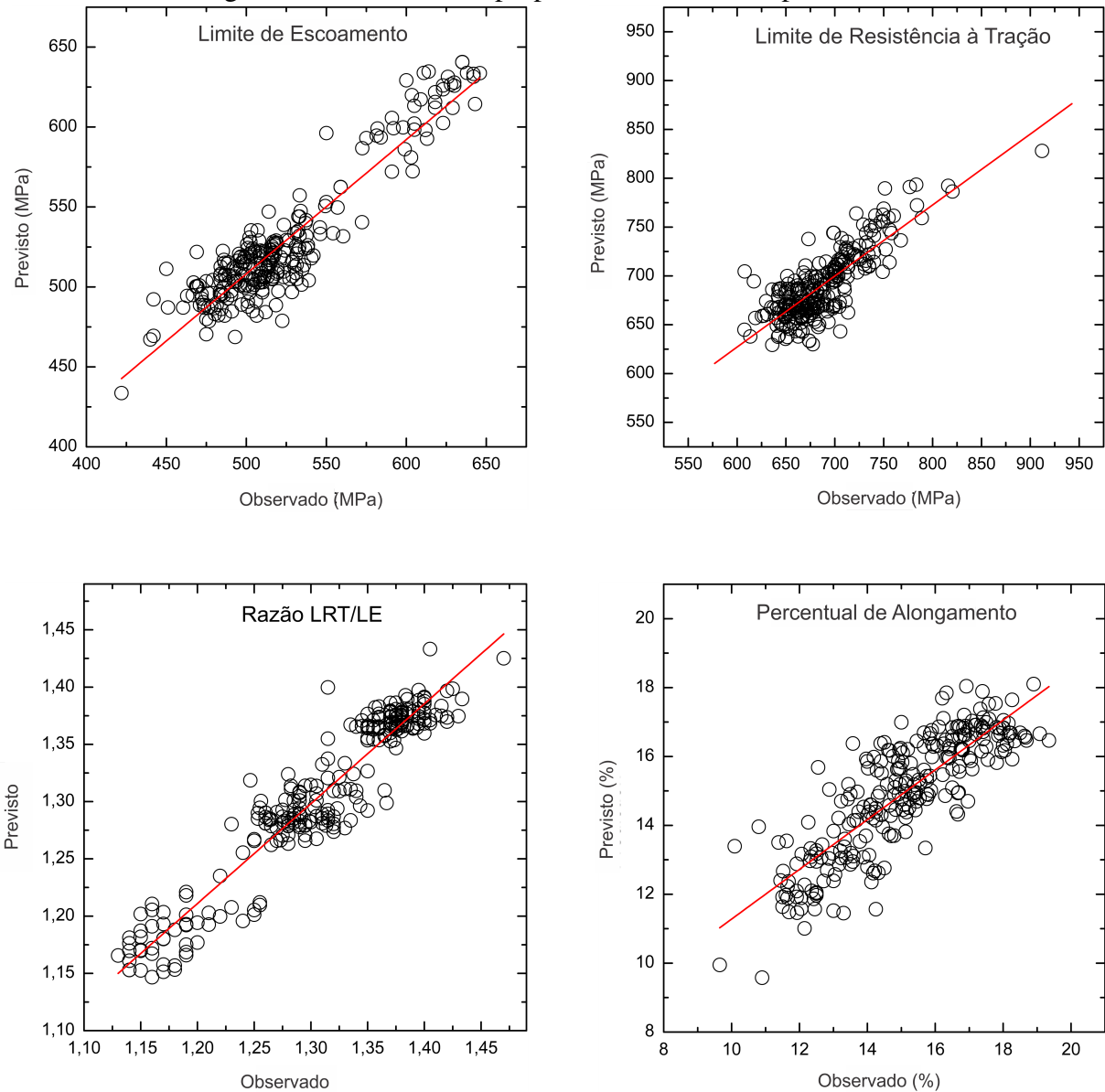
A Figura 26 mostra uma representação gráfica do coeficiente de determinação entre os valores-alvo e cada uma das quatro propriedades mecânicas estimadas pela rede neural.

#### 4.4 Máquinas de Vetores de Suporte

As quatro propriedades mecânicas do vergalhão também foram estimadas utilizando a MVS. Inicialmente, foi separado aleatoriamente 80% dos dados para treinar e 20% para testar o algoritmo. Para cada um desses testes, diferentes  $\sigma$ 's (parâmetro da função gaussiana) foram escolhidos. Os melhores resultados foram obtidos com a função gaussiana. Para cada propriedade mecânica foi treinado uma MVS. Os resultados obtidos para cada umas das propriedades estão sendo apresentados na tabela abaixo.

Para o LE, os valores de  $R2$  alcançados foram (máximo, médio) = (0,8086, 0,7296). Para a variável LRT, o modelo apresentou valores de  $R2$  (máximo, média) = (0,6242, 0,5180).

Figura 26 – Previsão das propriedades mecânicas pela RNA.



Para a variável AP, os valores de  $R^2$  alcançados foram (máximo, médio) = (0,6456, 0,7332). Já para a razão LRT/LE, o modelo apresentou  $R^2$  (máximo, médio) = (0,9002, 0,8469).

Para essas LR, LRT e o AP, o modelo baseado em MAM teve um desempenho melhor do que o modelo baseado em MVS em termos de valores médios de  $R^2$ . Os resultados encontrados para a razão LRT/LE foram estatisticamente equivalentes. Apesar de seu desempenho inferior quando comparado ao modelo baseado em MAM, um aspecto positivo do modelo baseado em MVS é que o número médio de vetores de suporte ficou em torno de 550, metade do número de pontos de referência ( $K$ ) exigidos pela MAM. Assim como os pontos de referência no MAM, o número do vetor de suporte é proporcional aos esforços computacionais e ao tamanho do conjunto de dados de armazenamento para prever as propriedades mecânicas.

Ao compararmos os resultados obtidos com a RNA e a MVS, os valores de  $R^2$  máximo encontrados para a razão entre LRT/LE e AP foram bem parecidos. Já para o Limite de Resistência à Tração e de escoamento, o coeficiente de determinação máximo foi melhor para a rede neural do que utilizando o algoritmo MVS. O valor do  $R^2$  médio encontrado entre as propriedades mecânicas estimadas e o observado apresentou melhoras para o LRT e o AP utilizando o método MVS.

A representação gráfica do coeficiente de determinação entre os valores alvo e os valores preditos para as quatro propriedades mecânicas estimadas pelo MVS é mostrada na Figura 27.

Figura 27 – Previsão das propriedades mecânicas pela MVS.

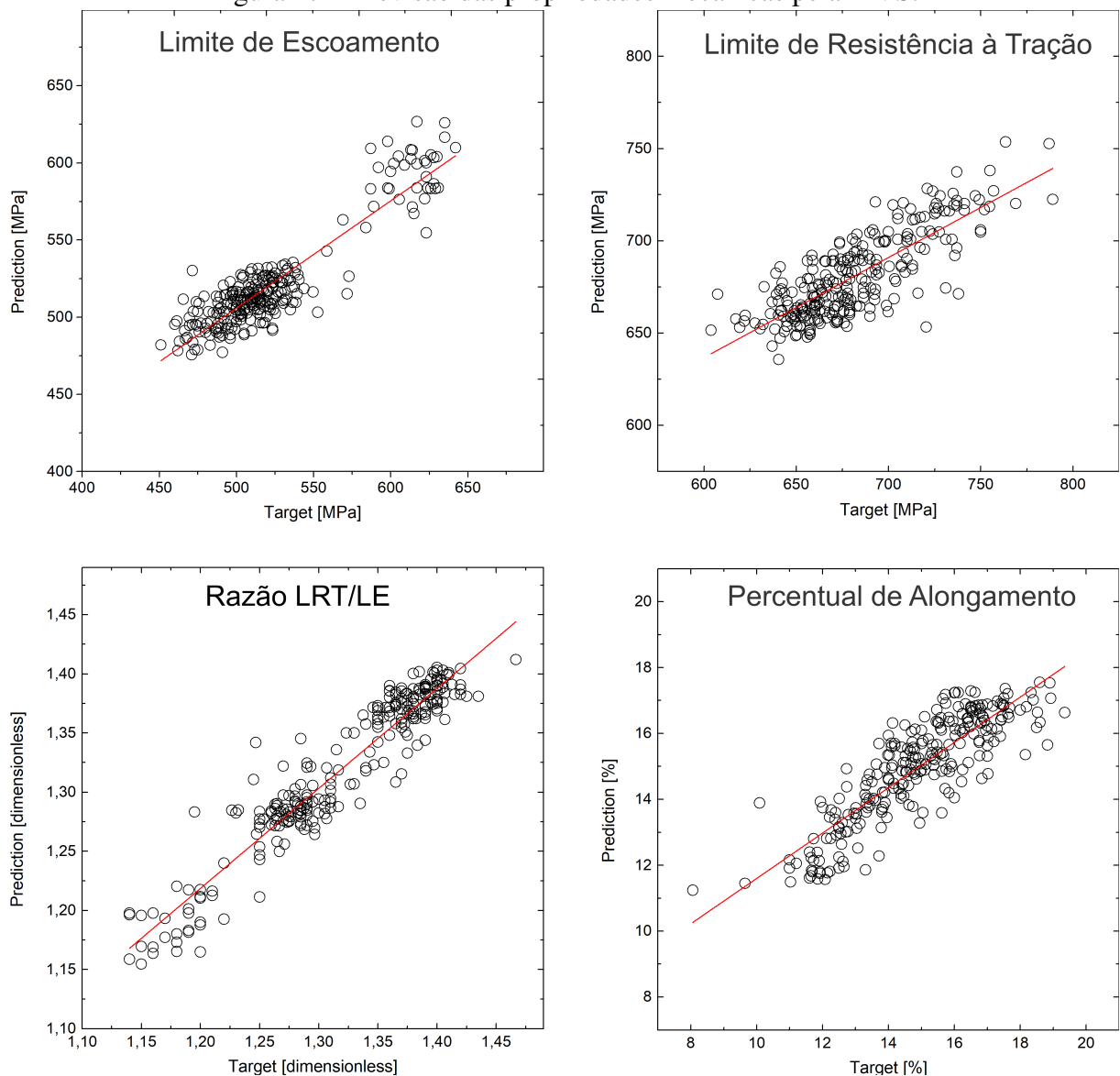




Tabela 16 – Estatísticas descritivas do desempenho da MVSMQ após 100 execuções de treinamento/teste.

Variáveis Preditas	$R^2$ médio	$R^2$ máximo	$R^2$ mínimo	$R^2$ desvio	# Vetores de Suporte (média)
<b>Limite de Escoamento (LE)</b>	0,7989	0,8665	0,6973	0,0342	899
	0,8048	0,8603	0,7343	0,0246	904
<b>Limite de Resistência à tração (LRT)</b>	0,7389	0,8436	0,5236	0,0630	906
<b>Razão LRT/LE</b>	0,8603	0,9156	0,7986	0,0240	889
	0,8737	0,9107	0,8127	0,0212	894
<b>Alongamento Percentual (AP)</b>	0,6883	0,7737	0,6009	0,0315	895
	0,7048	0,7663	0,6196	0,0309	901

#### 4.5 Máquinas de Vetores de Suporte por Mínimos Quadrados

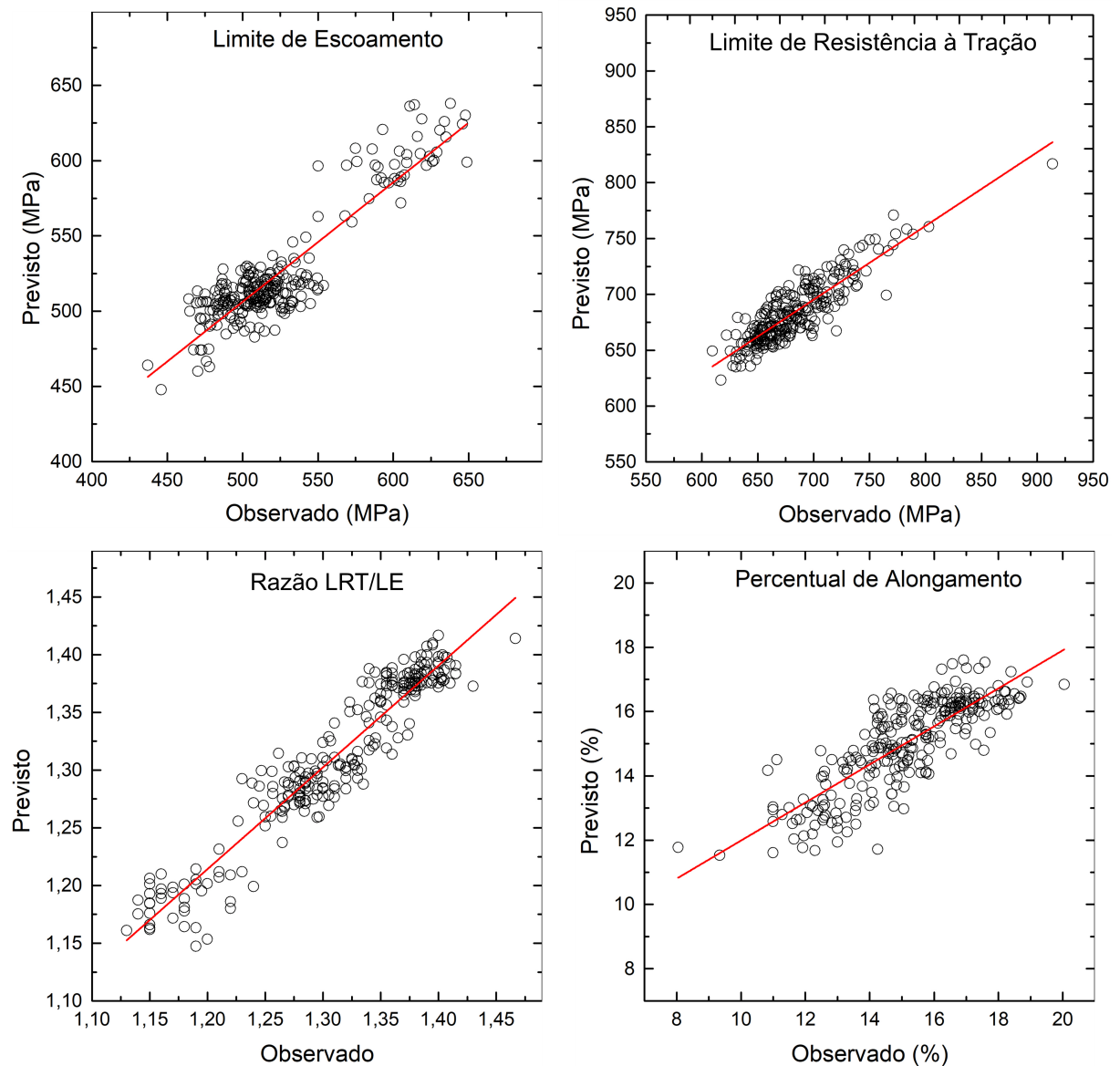
O último conjunto de experimentos de simulação envolve a avaliação do modelo baseado na MVSMQ. Inicialmente, uma série de testes preliminares foi realizada para escolher a melhor função kernel (entre as apresentadas na Tabela 1) e os melhores resultados encontrados foram obtidos com a função gaussiana. Então, 20 % do conjunto de dados foi separada para estimar os hiperparâmetros  $\gamma$  (parâmetro de regularização) e  $\sigma$  (largura da função kernel gaussiana), 60 % do conjunto de dados para treinar o modelo e 20 % do conjunto de dados para testá-lo. O procedimento de treinamento/teste é repetido por 100 execuções, cada execução com seleção aleatória de amostras para os conjuntos de dados de validação, treinamento e teste. Em resumo, para cada execução foram estimados valores diferentes para os  $\sigma$ 's e  $\gamma$ 's, então foi rodado o modelo mais 100 vezes embaralhando os dados de treinamento e teste para validar os hiperparâmetros escolhidos. Quatro diferentes modelos de regressão baseados na MVSMQ foram treinados, um para cada variável de saída. Após o treinamento, um determinado modelo baseado em MVSMQ é submetido a um procedimento de esparsificação para reduzir o número de vetores de suporte. Os melhores resultados obtidos para cada variável de saída prevista são relatados em Table 16.

Para a variável LE, os melhores valores de  $R^2$  alcançados foram (máximo, médio) = (0,8665, 0,7989) para o modelo com 899 vetores de suporte. De fato, os resultados para os modelos com 899 e 904 vetores de suporte são estatisticamente equivalentes. Neste caso, damos preferência ao modelo com menos vetores de suporte. Decisões semelhantes são tomadas para as outras três variáveis de saída. Por exemplo, para a variável AP, os melhores valores alcançados foram (máximo, médio) = (0,7736, 0,6883) para o modelo com 895 vetores de suporte. Para a razão LRT/LE, os melhores valores alcançados de  $R^2$  foram (máximo, médio) = (0,9156, 0,8603) para o modelo com 889 vetores de suporte. Vale ressaltar que essa variável atingiu os maiores

valores de  $R^2$  para todos os modelos avaliados, sugerindo ser a variável de saída mais fácil de prever. Finalmente, para a variável LRT, os valores de  $R^2$  foram (máximo, média) = (0,8436, 0,7389) para um modelo com 906 vetores de suporte.

A representação gráfica do coeficiente de determinação entre os valores-alvo e cada uma das quatro propriedades mecânicas estimadas pela MVSMQ estão sendo apresentadas na Figura 28.

Figura 28 – Previsão das propriedades mecânicas pela MVSMQ.



Ao compararmos os resultados obtidos com a MVS e MVSMQ, o  $R^2$  médio,  $R^2$  máximo e  $R^2$  mínimo obtido para o MVSMQ foi maior para todas as 4 propriedades mecânicas estudadas. Na regressão baseada em MVSMQ, todos os modelos exigiram cerca de 900 vetores de suporte para prever adequadamente as variáveis de saída de interesse. Isso explica basicamente

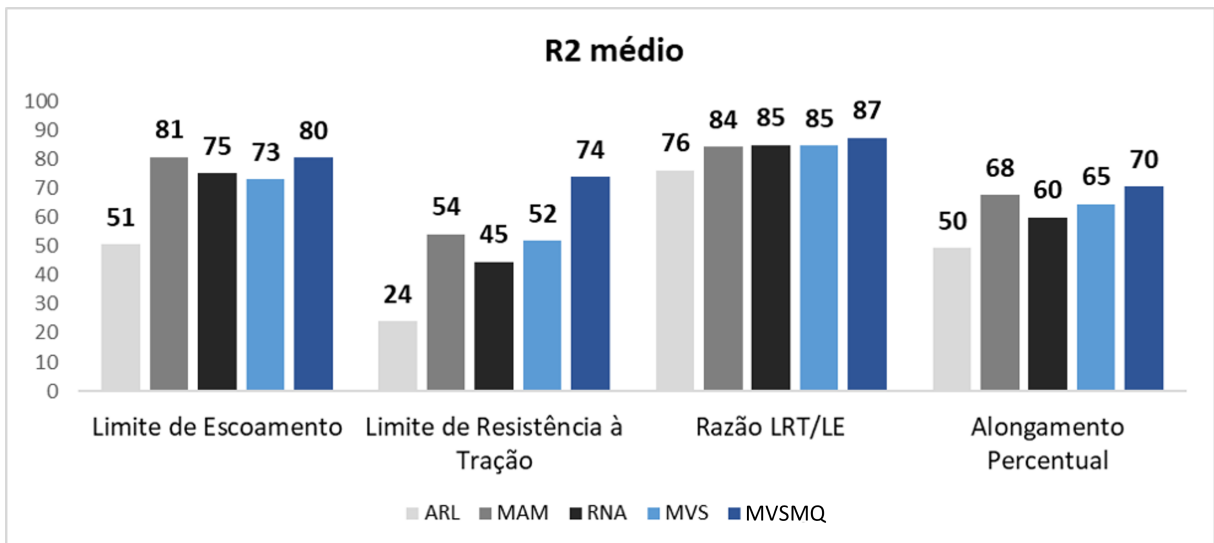
os desempenhos superiores dos modelos de regressão baseados em MVSMQ sobre aqueles alcançados pelo modelo baseado em MVS.

Ao compararmos os resultados obtidos com a RNA e o MVSMQ, os valores de  $R^2$  máximo encontrado para as propriedades LE, razão entre LRT/LE e AP foram bem parecidos. Já para o LRT, propriedade mecânica que apresentou mais dificuldade em ser estimada, o coeficiente de determinação máximo foi de 0,6868 usando a rede neural de multicamadas e de 0,7906 utilizando o algoritmo MVSMQ.

#### 4.6 Considerações Finais dos Resultados

Nesta seção, pretendemos fornecer uma avaliação geral dos resultados apresentados nas seções anteriores. Para tanto, as Figuras 29 e 30 mostram, respectivamente, os melhores resultados em termos de valores máximos e médios de  $R^2$  alcançados pelos algoritmos: RLM, MAM, RNA, MVS, MVSMQ.

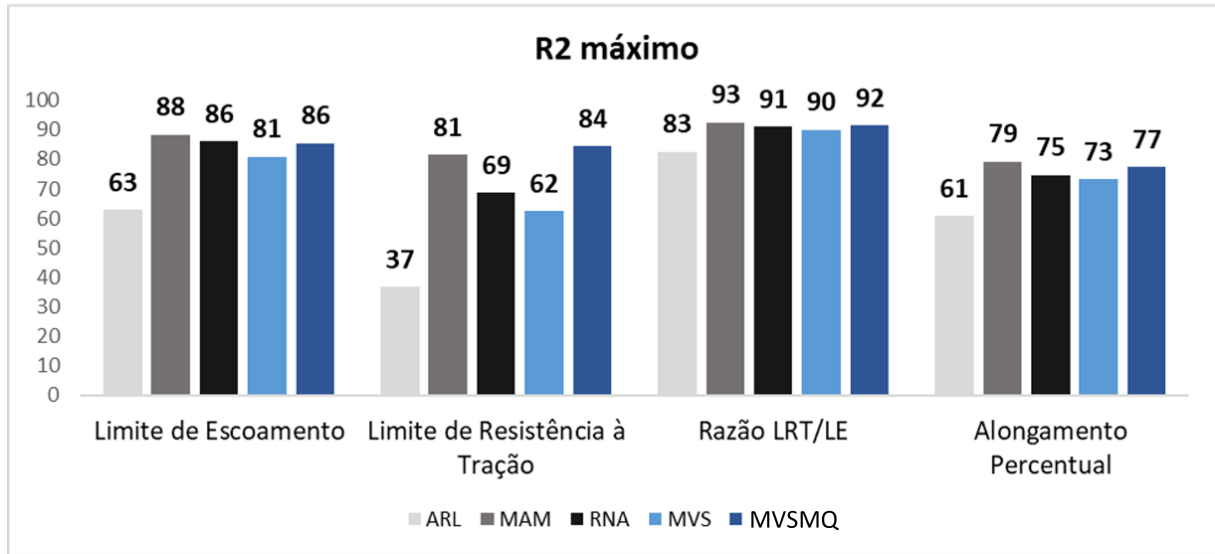
Figura 29 – Comparação geral do desempenho dos modelos avaliados em termos dos valores médios de  $R^2$ .



Ao analisar essas figuras, pode-se inferir que em termos de valores médios de  $R^2$  o melhor desempenho geral foi alcançado pelo modelo de regressão MVSMQ. O segundo melhor desempenho foi alcançado pelo modelo de regressão baseado em MAM. Em termos de valores máximos de  $R^2$ , o modelo de regressão baseado em MAM tem desempenho similar com os modelos baseados em MVSMQ.

Em termos de velocidade de treinamento (ou seja, construção de modelo), o modelo de regressão baseado em MAM traz alguma vantagem sobre o modelo baseado em MVSMQ,

Figura 30 – Comparação geral do desempenho dos modelos avaliados em termos dos valores máximos de  $R^2$ .



já que o primeiro requer o ajuste de um único hiperparâmetro ( $K$ , o número de pontos de referência), enquanto o último requer o ajuste de dois hiperparâmetros ( $\gamma$  e  $\sigma$ ). Além disso, o modelo baseado em MVSMQ requer uma escolha cuidadosa da função do kernel para funcionar corretamente, um grau adicional de liberdade com o qual o praticante tem que lidar. O modelo baseado em MAM é essencialmente um modelo linear no espaço de matrizes de distância, não exigindo nenhuma função kernel explícita para funcionar<sup>1</sup>.

A fim de entender a aplicabilidade dos modelos, foi calculado a porcentagem dos dados de teste que apresentaram um erro médio menor do que 5% e 10%, os resultados obtidos estão sendo apresentados na Tabela 17. O erro médio é calculado pelo módulo do valor observado menos o valor previsto dividido pelo valor observado.

A MAM apresentou a maior porcentagem de dados com erro menor do que 5% e 10% para todas as propriedades mecânicas e ARL a menor porcentagem de dados. O AP foi a propriedade que obteve maior erro médio, porém a MAM, RNA, MVS e MVSMQ obtiveram erro menor do que 10% para mais do que 80% dos dados.

Do exposto, os experimentos relatados indicam claramente dois modelos de regressão baseados em kernel competitivos para serem usados na previsão das propriedades mecânicas de vergalhões de aço: os modelos de regressão baseados em MVSMQ e MAM.

<sup>1</sup> O uso de cálculos de matrizes de distâncias quadradas pelo modelo MAM o qualifica como um método kernel. Veja (SCHÖLKOP, 2001) para detalhes.

Tabela 17 – Porcentagem dos dados de teste que apresentaram um erro médio menor do que 5% e 10%.

Modelo	Erro médio	LE	LRT	Razão LRT/LE	AP
<b>ARL</b>	$\leq 5\%$	60,7%	76,5%	96,9%	39,2%
	$\leq 10\%$	91,9%	98,8%	100%	74,6%
<b>MAM</b>	$\leq 5\%$	92,3%	96,6%	99,2%	62,7%
	$\leq 10\%$	100%	100%	100%	89,6%
<b>RNA</b>	$\leq 5\%$	85,3%	90,7%	98,4%	55,7%
	$\leq 10\%$	99,2%	99,6%	100%	86,6%
<b>MVS</b>	$\leq 5\%$	85,3%	90,7%	98,4%	55,7%
	$\leq 10\%$	99,2%	99,6%	100%	86,4%
<b>MVSMQ</b>	$\leq 5\%$	82,6%	86,1%	99,6%	56,9%
	$\leq 10\%$	98,8%	100%	100%	83,4%

#### 4.7 Resumo do Capítulo

Neste capítulo foi apresentado e discutido os principais resultados obtidos com os algoritmos de aprendizado de máquina. No próximo capítulo está sendo apresentado a conclusão do trabalho e propostas para trabalhos futuros.

## 5 CONCLUSÃO

A composição química (teores de C, Mn, Nb, V, P, S, Si, Cr, Mo, Ni, Cu e Sn) e os parâmetros do tratamento térmico (temperaturas do trem de laminação inicial, área da secção transversal, temperaturas de entrada e saída, pressão e a vazão de água do tratamento térmico) estão relacionados às quatro principais propriedades mecânicas finais do aço (LE, LRT, Razão do LRT/LE e AP).

O teste de significância aplicado à regressão linear foi realizado para analisar se cada propriedade mecânica considerada neste estudo está linearmente relacionada às variáveis independentes. O teste demonstrou que o carbono, vanádio e manganês estão fortemente relacionados às quatro propriedades mecânicas estudadas. Este resultado é consistente com os encontrados na literatura.

A ARL foi a técnica que pior estimou todas as 4 propriedades estudadas devido a capacidade do método de determinar apenas a relação linear entre a variável de entrada e saída, enquanto os outros métodos estudados podem detectar correlações lineares e não lineares entre as variáveis.

O coeficiente de determinação ( $R^2$ ) entre os valores estimados pela MAM e os respectivos valores observados indica que esta técnica conseguiu muito boa estimação para LE, Razão LRT/LE e AP e pode ser utilizada na previsão das propriedades mecânicas de vergalhões de aço.

A MAM apresentou a maior porcentagem de dados com erro menor do que 5% e 10% para todas as propriedades mecânicas e ARL a menor porcentagem de dados.

A MVSMQ foi capaz de prever melhor todas as quatro propriedades estudadas do que a MVS. Estes resultados eram esperados considerando que o método MVSMQ utiliza 1,6 vezes mais vetores de suporte do que o utilizado pelo algoritmo MVS.

Os experimentos numéricos revelaram informações interessantes, como o fato de que dois dos modelos baseados em kernel avaliados, a saber, aqueles baseados nos algoritmos MVSMQ e MAM, superaram consistentemente o modelo de regressão baseado em rede neural artificial. Esta é uma conquista relevante no sentido de que os modelos baseados em MVSMQ e MAM são consideravelmente mais simples de treinar e avaliar do que o modelo de regressão baseado em RNA.

Em termos de velocidade de treinamento, o modelo de regressão baseado em MAM traz alguma vantagem sobre o modelo baseado em MVSMQ, já que o primeiro requer o ajuste

de um único hiperparâmetro ( $K$ ), enquanto o último requer o ajuste de dois hiperparâmetros ( $\gamma$  e  $\sigma$ ). Além disso, o modelo baseado em MVSMQ requer uma escolha cuidadosa da função do kernel para funcionar corretamente e o modelo baseado em MAM não exige nenhuma função kernel explícita para funcionar.

Os experimentos relatados indicam que os dois modelos de regressão baseados em kernel podem ser utilizados como ferramenta de previsão em uma indústria siderúrgica: MVSMQ e MAM.

Em trabalhos futuros, análises de sensibilidade das correlações não-lineares entre as variáveis de entrada e de saída podem ser realizadas para melhorar a previsibilidade do modelo. Novas estratégias em modelos de aprendizado profundo kernel (WILSON *et al.*, 2016) podem ser avaliadas com o objetivo de aumentar ainda mais os valores de  $R^2$  para as saídas de interesse. Pode também ser realizado um estudo mais concentrado separando os aços que obedecem alguma norma, dimensional e aplicação específica.

## REFERÊNCIAS

- AMIN, R.; PICKERING, F. Thermomechanical processing of microalloyed austenite, eds. **A. J. Derdo, GA Ratz and PJ Wray, TMS-AIME, Warrendale, PA, USA**, p. 377–402, 1982.
- ARAÚJO, I. S. D. **Produção brasileira de aço e a competitividade do Brasil no período de 2016 a 2018**. Monografia (Graduação) – Pontifícia Universidade Católica de Goiás. Goiânia, 6 2021.
- ARAÚJO, V. T. **Caracterização e quantificação microestrutural de barras de aço produzidas via processo thermex**. Monografia (Graduação) – Universidade Federal do Ceará. Fortaleza, 12 2010.
- ARSHADI, H.; KHEYRODDIN, A.; NADERPOUR, H. High-strength reinforcement effects on the seismic behaviour of beam–column joints. **Proceedings of the Institution of Civil Engineers-Structures and Buildings**, Thomas Telford Ltd, p. 1–12, 2020.
- ASSOCIATION, W. S. **World Steel in Figures**. Disponível dia 09/12/2021 às 21h em: "<https://www.worldsteel.org/en/dam/jcr:976723ed-74b3-47b4-92f6-81b6a452b86e/World%2520Steel%2520in%2520Figures%25202021.pdf>", 2021.
- AUGUSTI, A. In-line thermal treatment applies to straight and coiled rebars. **Steel Times International**, Quartz Business Media Ltd, v. 19, n. 2, p. 44, 1995.
- BELAYADI, A.; BOURAHLA, B. Neural network model for 7000 (al-z) alloys: Classification and prediction of mechanical properties. **Physica B: Condensed Matter**, Elsevier, v. 554, p. 114–120, 2019.
- BLACK, J. T.; KOHSER, R. A. **DeGarmo's materials and processes in manufacturing**. [S.l.]: John Wiley & Sons, 2020.
- BOËR, C. R.; REBELO, N. M.; RYDSTAD, H. A.; SCHRÖDER, G. **Process modelling of metal forming and thermomechanical treatment**. [S.l.]: Springer Science & Business Media, 2012.
- BOLJANOVIC, V. **Metal shaping processes: casting and molding, particulate processing, deformation processes, and metal removal**. [S.l.]: Industrial Press Inc., 2010.
- CALLISTER, W. D.; RETHWISCH, D. G. **Materials science and engineering: an introduction**. [S.l.]: John Wiley & Sons New York, 2007. v. 7.
- CAMPBELL, F. C. **Elements of metallurgy and engineering alloys**. [S.l.]: ASM International, 2008.
- CARVALHO, A.; FARIA, G.; MOREIRA, P. Efeito do tempo e da temperatura de tratamento isotérmico sobre o tamanho de grão, dureza e susceptibilidade à corrosão intergranular de um aço inoxidável ferrítico. **ABM-Associação Brasileira de Metalurgia, Materiais e Mineração**, v. 78, 2010.
- CHIAVERINI, V. **Aços e ferros fundidos**. [S.l.]: ABM-Associação Brasileira de Metalurgia, Materiais e Mineração, 1977.
- DANIELSSON, P.-E. Euclidean distance mapping. **Computer Graphics and image processing**, v. 14, n. 3, p. 227–248, 1980.



DOUGLAS, C.; RUNGER, G. C. **Applied statistics and probability for engineers**. [S.l.]: John Wiley & Sons, Inc, 2002.

DRUCKER, H.; BURGESS, C. J.; KAUFMAN, L.; SMOLA, A.; VAPNIK, V. Support vector regression machines. **Advances in neural information processing systems**, Morgan Kaufmann Publishers, v. 9, p. 155–161, 1997.

EZIO, C.; MATTEO, D.; DANIELE, F.; NICOLETTA, T.; CARLO, A. Mechanical behaviour of quenched and self-tempered reinforcing steel in tension under high strain rate. **Materials and Design**, Elsevier, 2013.

FILHO, C. H. M. A. **Análise da Influência da Taxa de Resfriamento no Gradiente Microestrutural de Barras Laminadas**. Tese (Doutorado) — Pontifícia Universidade Católica do Paraná. Curitiba, 2004.

GHAISARI, J.; JANNESARI, H.; VATANI, M. Artificial neural network predictors for mechanical properties of cold rolling products. **Advances in Engineering Software**, Elsevier, v. 45, n. 1, p. 91–99, 2012.

GIGOVIĆ-GEKIĆ, A.; ORUC, M.; AVDUSINOVIĆ, H.; SUNULAHPASIĆ, R. Regression analysis of the influence of a chemical composition on the mechanical properties of the steel nitronic 60. **Materials and Technologies**, v. 48, n. 3, p. 433–437, 2014.

GUIMARÃES, D. O. M.; SOUZA, C. F. de. Estudo da fragilização a quente nos aços produzidos pela companhia siderúrgica nacional. **ABM-Associação Brasileira de Metalurgia, Materiais e Mineração**, 2015.

HAYKIN, S. S. **Neural networks and learning machines**. [S.l.]: Pearson Upper Saddle River, 2009. v. 3.

HEBB, D. O. **The organization of behavior: A neuropsychological theory**. [S.l.]: Psychology Press, 2005.

HOSFORD, W. F. **Mechanical behavior of materials**. [S.l.]: Cambridge university press, 2010.

J. WRIGHT M., W. P. M. Reinforcing bar: Hardening mechanisms and performance in use. **AISTech**, 2016.

JIAO, H.; ZHAO, X.-L. Material ductility of very high strength (vhs) circular steel tubes in tension. **Thin-walled structures**, Elsevier, v. 39, n. 11, p. 887–906, 2001.

JUUTILAINEN, I.; RÖNING, J.; MYLLYKOSKI, L. Modelling the strength of steel plates using regression analysis and neural networks. In: **Proceedings of International Conference on Computational Intelligence for Modelling, Control and Automation**. [S.l.: s.n.], 2003. p. 681–691.

LAURITO, D. F. **Estudo da propagação da trinca por fadiga em um aço microligado com diferentes condições microestruturais**. Tese (Doutorado) — Universidade de São Paulo, 2010.

LEVENBERG, K. A method for the solution of certain non-linear problems in least squares. **Quarterly of Applied Mathematics**, v. 2, n. 2, p. 164–168, 1944.

LIMA, M. N. d. S.; MAIA, P. P. N.; ABREU, H. F. G. d. Microestrutura e dureza do aço baixo carbono com alto teor de manganês (27% p) laminado. **ABM-Associação Brasileira de Metalurgia, Materiais e Mineração**, 2017.

MARQUARDT, D. An algorithm for least-squares estimation of nonlinear parameters. **SIAM Journal on Applied Mathematics**, v. 11, n. 2, p. 431–441, 1963.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Springer, v. 5, n. 4, p. 115–133, 1943.

MERCER, J. XYI Functions of positive and negative type, and their connection the theory of integral equations. **Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character**, The Royal Society London, v. 209, n. 441-458, p. 415–446, 1909.

METALS, A. S. for. **Atlas of isothermal transformation and cooling transformation diagrams**. [S.l.]: American Society for Metals Materials Park, OH, 1977.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. [S.l.]: John Wiley & Sons, 2012. v. 821.

MURTA, R. H. F.; BRAGA, F. D.; MAIA, P. P. N.; DIÓGENES, O. B. F.; de MOURA, E. P. Mathematical modelling for predicting mechanical properties in rebar manufacturing. **Ironmaking & Steelmaking**, v. 48, n. 2, p. 161–169, 2021.

NORMALISATION, C. E. de. Eurocode 3: Design of steel structures—part 1-1: General rules and rules for buildings. **Brussels, Belgium. CEN**, 2005.

PALMA, E. S. **Desenvolvimento de um aço sinterizado ligado ao manganês e fósforo**. Dissertação (Mestrado) — Universidade Federal de Santa Catarina, 1985.

RAMASUBRAMANIAN, K.; MOOLAYIL, J. **Applied Supervised Learning with R: Use Machine Learning Libraries of R to Build Models that Solve Business Problems and Predict Future Trends**. [S.l.]: Packt Publishing, Limited, 2019.

ROBERTS, W. L. **Hot rolling of steel**. [S.l.]: CRC Press, 1983.

Rocha Neto, A. R.; BARRETO, G. A. Opposite maps: Vector quantization algorithms for building reduced-set SVM and LSSVM classifiers. **Neural Processing Letters**, v. 37, p. 3–19, 2013.

SAMPAIO, R. S.; JONES, J.; VIEIRA, J. Hot metal strategies for the eaf industry. **Iron & steel technology**, n. 2, p. 31–37, 2009.

SANTOS, J. D. A.; BARRETO, G. A. A regularized estimation framework for online sparse lssvr models. **Neurocomputing**, Elsevier, v. 238, p. 114–125, 2017.

SANTOS, L. B. D. **ESTUDO COMPARATIVO DO COMPORTAMENTO DO COBRE NOS ENSAIOS DE TRAÇÃO E COMPRESSÃO à QUENTE NOS AÇOS**. Monografia (Graduação) – Universidade Federal do Espírito Santos. Vitória, 2017.

SCHÖLKOP, B. The kernel trick for distances. In: LEEN, T.; DIETTERICH, T.; TRESP, V. (Ed.). **Advances in Neural Information Processing Systems**. MIT Press, 2001. v. 13, p. 1–7. Disponível em: <<https://proceedings.neurips.cc/paper/2000/file/4e87337f366f72daa424dae11df0538c-Paper.pdf>>.

- SEBER, G. A. F.; LEE, A. J. **Linear regression analysis**. [S.l.]: John Wiley & Sons, 2012. v. 329.
- SHIGAKI, Y.; SILVA, J. M.; ROMIE, R. N. A.; SANTOS, S. C. Modelamento, análise e validação do processo de laminação de encruamento pelo método dos elementos finitos. **Tecnologia em Metalurgia, Materiais e Mineração**, ABM-Associação Brasileira de Metalurgia, Materiais e Mineração, v. 16, n. Especial, p. 0–0, 2019.
- SINGH, S.; BHADESHIA, H.; MACKAY, D.; CAREY, H.; MARTIN, I. Neural network analysis of steel plate processing. **Ironmaking and Steelmaking**, Citeseer, v. 25, n. 5, p. 355–365, 1998.
- Souza Junior, A. H.; CORONA, F.; BARRETO, G. A.; MICHE, Y.; LENDASSE, A. Minimal learning machine: A novel supervised distance-based approach for regression and classification. **Neurocomputing**, v. 164, p. 34–44, 2015.
- SPIEGEL, M. R.; STEPHENS, L. J. **Estatística: Coleção Schaum**. [S.l.]: Bookman, 2000.
- STAATS, M. E. **Estudo para otimização do projeto de navalha para corte a quente de vergalhão de 25 milímetros**. Monografia (Graduação) – Universidade Federal do Rio Grande do Sul. Porto Alegre, 2010.
- SUYKENS, J. A.; VANDEWALLE, J. Least squares support vector machine classifiers. **Neural processing letters**, Springer, v. 9, n. 3, p. 293–300, 1999.
- SUYKENS, J. A. K.; Van Gestel, T.; De Brabanter, J.; De Moor, B.; VANDEWALLE, J. **Least squares support vector machines**. [S.l.]: World Scientific, 2002.
- WHITE, H. Using least squares to approximate unknown regression functions. **International Economic Review**, JSTOR, p. 149–170, 1980.
- WILSON, A. G.; HU, Z.; SALAKHUTDINOV, R.; XING, E. P. Deep kernel learning. In: GRETTON, A.; ROBERT, C. C. (Ed.). **Proceedings of the 19th International Conference on Artificial Intelligence and Statistics**. Cadiz, Spain: PMLR, 2016. (Proceedings of Machine Learning Research, v. 51), p. 370–378. Disponível em: <<https://proceedings.mlr.press/v51/wilson16.html>>.
- WORRELL, E. Energy efficiency improvement and cost saving opportunities for the us iron and steel industry. **Ernest Orlando Lawrence Berkeley National Laboratory**, 2011.
- YANG, Y. Y.; MAHFOUF, M.; LINKENS, D. A.; ZHANG, Q. Tensile strength prediction for hot rolled steels by bayesian neural network model. **IFAC Proceedings Volumes**, Elsevier, v. 42, n. 23, p. 255–260, 2009.
- YILDIRIM, I. Z.; PREZZI, M. Experimental evaluation of eaf ladle steel slag as a geo-fill material: Mineralogical, physical & mechanical properties. **Construction and Building Materials**, Elsevier, v. 154, p. 23–33, 2017.