



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

DIÊGO FARIAS DE OLIVEIRA

MÁQUINA DE APRENDIZAGEM MÍNIMA PARA DADOS INTERVALARES

FORTALEZA

2019

DIÉGO FARIAS DE OLIVEIRA

MÁQUINA DE APRENDIZAGEM MÍNIMA PARA DADOS INTERVALARES

Dissertação apresentada ao Curso de do Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Ciência da Computação. Área de Concentração: Teoria da Computação

Orientador: Prof. Dr. João Paulo Pordeus Gomes

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

O46m Oliveira, Diêgo Farias de.

Máquina de Aprendizagem Mínima para dados intervalares / Diêgo Farias de Oliveira. – 2019.
48 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Ciência da Computação, Fortaleza, 2019.

Orientação: Prof. Dr. João Paulo Pordeu Gomes.

1. Máquina de Aprendizagem Mínima. 2. Dados Intervalares. 3. Modelos de Regressão. 4. Modelos de Classificação. I. Título.

CDD 005

DIÊGO FARIAS DE OLIVEIRA

MÁQUINA DE APRENDIZAGEM MÍNIMA PARA DADOS INTERVALARES

Dissertação apresentada ao Curso de do Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Ciência da Computação. Área de Concentração: Teoria da Computação

Aprovada em:

BANCA EXAMINADORA

Prof. Dr. João Paulo Pordeus Gomes (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. João Fernando Lima Alcântara
Universidade Federal do Ceará (UFC)

Prof. Dr. Leonardo Ramos Rodrigues
Instituto de Aeronáutica e Espaço (IAE)

Aos meus pais e minha avó.

“Tudo o que temos de decidir é o que fazer com o tempo que nos é dado.”

(Gandalf)

RESUMO

Resolver problemas de aprendizagem de máquina com conjuntos de dados intervalares é uma tarefa desafiadora que pode surgir em muitas aplicações do mundo real, por exemplo, na predição da frequência cardíaca. Motivado por esse fato, muitos pesquisadores propuseram métodos de regressão não-linear e classificadores para lidar com dados intervalares nos últimos anos. Neste trabalho, são propostas quatro variantes da Máquina de Aprendizagem Mínima (MLM) para dados intervalares, sendo duas com o foco em tarefas de classificação e regressão, e outras duas exclusivamente para regressão. A escolha do MLM é explicada pelo seu notável desempenho em muitas aplicações e pela necessidade da definição de um único hiperparâmetro. Para validar os métodos propostos, nos problemas de regressão, eles foram comparados com cinco métodos não-lineares: três variantes do *Extreme Learning Machine* (iELM) e duas extensões da regressão de kernel. Já na classificação, três modelos de regressão logística foram utilizados. Os experimentos em conjuntos de dados sintéticos (com diferentes configurações) e dados reais são apresentados e mostram que as variantes do MLM para dados intervalares obtêm resultados similares ou melhores aos demais. A partir deste estudo, nota-se que os métodos propostos apresentaram resultados competitivos e podem ser considerados boas opções.

Palavras-chave: máquina de aprendizagem mínima; dados; intervalares; modelos; regressão; classificação.

ABSTRACT

Solving machine learning problems with interval-valued data is a challenging task that can arise in many real world applications, for example, in heart rate prediction. Motivated by this fact, many researchers have proposed nonlinear regression methods and classifiers to deal with interval-valued data in recent years. In this work, four variants of the Minimal Learning Machine (MLM) are proposed for interval-valued data, two focusing on regression problems and two for classification and regression. The choice of MLM is explained by its outstanding performance in many applications and by the need to define a single hyperparameter. To validate the proposed methods, in regression problems, they were compared with five non-linear regressors: three Extreme Learning Machine (iELM) variants and two extensions of kernel regression. In the classification, three logistic regression models were used. Experiments on synthetic datasets (with different configurations) and real data are presented and they show that the variants of the MLM for interval-valued data obtain similar or better results than the others. From this study, it is noted that the proposed methods presented competitive results and can be considered good options.

Keywords: regression; models; minimal learning machine; interval; data.

LISTA DE FIGURAS

Figura 1 – Exemplificação do problema de multilateração	17
Figura 2 – Ilustração do conjunto sintético $DS1_1$	31
Figura 3 – Ilustração do conjunto sintético $DS2_2$	32
Figura 4 – Ilustração do conjunto sintético linear 1 com $\gamma_1, \gamma_2 = \mathcal{U} \sim (1, 10)$	37
Figura 5 – Ilustração do conjunto sintético linear 2 com $\gamma_1, \gamma_2 = \mathcal{U} \sim (1, 10)$	38
Figura 6 – Ilustração do conjunto sintético não linearmente separável 1 (DSLN1) com $\gamma_1, \gamma_2 = \mathcal{U} \sim (1, 10)$	39
Figura 7 – Ilustração do conjunto sintético não linearmente separável 2 (DSNL2) com $\gamma_1, \gamma_2 = \mathcal{U} \sim (1, 10)$	40

LISTA DE TABELAS

Tabela 1 – Parâmetros utilizados para geração dos conjuntos de dados sintéticos	32
Tabela 2 – Parâmetros utilizados para geração do segundo esquema dos conjuntos de dados sintéticos	33
Tabela 3 – Configurações conjunto de dados linearmente separável 1 (DSL1)	37
Tabela 4 – Configurações conjunto de dados linearmente separável 2 (DSL2)	38
Tabela 5 – RMSEB obtidos nos conjuntos de dados sintéticos com relação fraca	42
Tabela 6 – RMSEB obtidos na primeira configuração dos dados sintéticos com relação direta	42
Tabela 7 – RMSEB obtidos na segunda configuração dos dados sintéticos com relação direta	43
Tabela 8 – RMSEB obtidos nos conjuntos reais	43
Tabela 9 – Acurácia dos métodos nos Dados Sintéticos	44
Tabela 10 – Acurácia obtida nos dados reais	44

LISTA DE SÍMBOLOS

x^C	Centro de um intervalo de entrada
x^R	Raio de um intervalo de entrada
x^L	Limite inferior de um intervalo de entrada
x^U	Limite superior de um intervalo de entrada
\mathbf{x}^C	Vetor de centros de intervalos de entrada
\mathbf{x}^R	Vetor de raios de intervalos de entrada
\mathbf{x}^L	Vetor de limites inferiores de intervalos de entrada
\mathbf{x}^U	Vetor de limites superiores de intervalos de entrada
\mathbf{m}_k	Ponto de referência k
K	Número de ponto de referências
$\mathbf{\Delta}$	Matriz de distâncias no conjunto de saída
\mathbf{D}	Matriz de distâncias no conjunto de entrada

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Justificativa e Objetivos	13
1.2	Objetivos	14
2	MÁQUINA DE APRENDIZAGEM MÍNIMA	15
2.1	Formulação Básica	15
2.2	Algoritmo de Aprendizagem	16
2.3	Estimação da Saída	16
2.3.1	<i>MLM para classificação</i>	17
3	MLM PARA DADOS INTERVALARES	19
3.1	MLM com Limite Superior e Inferior Independentes	19
3.2	MLM com Centro e Raio Independentes	21
3.3	MLM com Limite Superior e Inferior Dependentes	23
3.4	MLM com centro e raio Dependentes	24
3.5	Classificação	26
4	CONFIGURAÇÕES DOS EXPERIMENTOS NA TAREFA DE REGRESSÃO	27
4.1	Métodos Regressão	27
4.1.1	<i>Extreme Learning Machine</i>	27
4.1.1.1	<i>Modelo 1</i>	27
4.1.1.2	<i>Modelo 2</i>	28
4.1.1.3	<i>Modelo 3</i>	28
4.1.2	<i>Interval Kernel Regression</i>	29
4.1.2.1	<i>Modelo 1</i>	29
4.1.2.2	<i>Modelo 2</i>	30
4.2	Conjuntos de dados	30
4.2.1	<i>Conjuntos sintéticos esquema 1</i>	30
4.2.1.1	<i>Conjunto de dados sintético 1 (DS1)</i>	30
4.2.1.2	<i>Conjuntos de dados sintéticos 2 (DS2)</i>	31
4.2.2	<i>Conjuntos Sintéticos Esquema 2</i>	33
4.2.3	<i>Conjuntos Reais</i>	33

4.3	<i>Root Mean Square Error Bounds</i>	34
5	CONFIGURAÇÕES DOS EXPERIMENTOS NA TAREFA DE CLASSIFICAÇÃO	35
5.1	Métodos Classificação	35
5.1.1	<i>Regressão Logística para Dados Intervalares</i>	35
5.1.1.1	<i>Modelo 1 (LOGIT-C) e 2 (LOGIT-LU)</i>	35
5.1.1.2	<i>Modelo 3 (LOGIT_L_U)</i>	36
5.2	Conjunto de Dados	36
5.2.1	<i>Conjuntos de Dados Sintéticos</i>	36
5.2.1.1	<i>Conjunto Linearmente Separáveis (DSL)</i>	37
5.2.1.2	<i>Conjuntos Não-Linearmente Separáveis (DSNL)</i>	38
5.2.2	<i>Conjuntos de dados reais</i>	41
6	RESULTADOS	42
6.1	Resultados Experimentos de Regressão	42
6.2	Resultados Experimentos de Classificação	43
7	CONCLUSÕES E TRABALHOS FUTUROS	45
	REFERÊNCIAS	46

1 INTRODUÇÃO

Algoritmos de Aprendizagem de Máquina, geralmente, são desenvolvidos para trabalhar com conjunto de dados clássicos, onde cada amostra é representada como um único ponto em \mathbb{R}^d , sendo d a dimensão. Porém em muitas situações esses dados podem ser intervalos e utilizar a forma clássica pode não representar bem o problema, causando perda de informação ou dados que não são consistentes. Agora, uma amostra é um vetor de intervalos (BILLARD, 2006). Esse tipo de dado surge em diversas situações, exemplos são:

- alguns dados são inerentemente intervalos, como a pressão sanguínea;
- os problemas podem ser modelados como intervalos, por exemplo, dados que são coletados a partir de um questionário, atributos como idade ou renda geralmente são representados por intervalos;
- dados temporais podem ser melhores representados por intervalos, tendo como exemplo uma situação em que a massa corporal da pessoa é importante no início de um período de tempo e no fim;
- dados que são coletados a partir de equipamentos com imprecisão, em que os valores medidos estão em uma faixa de valores.

Objetivando superar o problema de manipular esses dados e conseguir utilizar métodos de aprendizagem de máquina clássicos, os intervalos são alterados e muitas vezes são considerados apenas seus pontos médios (BILLARD, 2006), assim um intervalo $[0, 100]$ é substituído por 50, porém se existir um intervalo $[40, 60]$ eles serão representados pelo mesmo ponto médio, mesmo sendo completamente diferentes. Outra abordagem é simplesmente ignorar os atributos intervalares, o que não é a melhor opção.

Para tentar contornar essa adversidade, o desenvolvimento e adaptação de algoritmos clássicos para intervalos vem atraindo grande interesse nos últimos anos. No âmbito de regressão, o trabalho de Roberta *et al.* (2013) propõe uma regressão não paramétrica para intervalos; em Zhao e Wang (2014), três variantes da máquina de aprendizagem extrema (*extreme learning machines*) são adaptadas para intervalos; Billard e Diday (2000), Neto e Carvalho (2010) e Li *et al.* (2017) propuseram diferentes modelos para regressão linear, a primeira cria uma formulação baseada nos centros dos intervalos, a segunda baseia-se no centro e raio dos intervalos, além de adicionar uma restrição para manter a consistência dos intervalos de saída, a terceira cria um modelo de regressão linear restrito baseado na teoria de conjuntos aleatórios; a rede perceptron multicamadas foi adaptada em Roque *et al.* (2007); um modelo de rede neural probabilística foi

apresentado em Kowalski e Kulczycki (2017); recentemente Yang *et al.* (2019) desenvolveram um modelo de redes neurais regularizada para intervalos. Já problemas de classificação são menos estudados, alguns trabalhos como o artigo de Queiroz *et al.* (2018), cria uma variante da análise do discriminante para intervalos; em (SOUZA *et al.*, 2011), foram apresentadas quatro versões para regressão logística.

Métodos não supervisionados também vêm ganhando destaque, Wang *et al.* (2019) propõem um método de agrupamento baseado em uma nova representação na forma de uma sequência de 3-tuplas para séries temporais com intervalos; Carvalho e Simões (2017) adaptam o algoritmo *Fuzzy C-Means Clustering* baseado em distâncias adaptáveis de City-Block e Hausdorff; em Carvalho e Tenorio (2010) foi modelado o *Fuzzy K-means Clustering* para dados intervalares. Ainda em aprendizagem não-supervisionada, algoritmos de redução de dimensionalidade como Análise dos Componentes Principais e *Multidimensional Scaling* são remodeladas em (Lauro e Palumbo (2000), Douzal-Chouakria *et al.* (2011)) e (Denoeux e Masson (2000), Groenen *et al.* (2006)), respectivamente.

Máquina de Aprendizagem Mínima (*Minimal Learning Machine*, MLM) (Souza Junior *et al.*, 2013) é um método de aprendizado supervisionado que tem atraído muito interesse nos últimos anos devido ao seu desempenho e à necessidade de um único hiperparâmetro. Em poucas palavras, o modelo MLM constrói uma relação não linear entre entradas e saídas usando um mapeamento linear entre matrizes de distância nos espaços de entrada e saída. Aplicações bem-sucedidas do MLM foram relatadas em Mesquita *et al.* (2017b), Coelho *et al.* (2014), Marinho *et al.* (2017). Embora o MLM tenha alcançado bons resultados em muitos domínios, a formulação original não permite sua aplicação em problemas que envolvem dados intervalares.

Com o objetivo de estender a aplicabilidade do modelo MLM original, neste trabalho são propostos variantes do MLM para conjuntos de dados intervalares, tanto para problemas de classificação quanto para regressão. As formulações são baseadas nas abordagens que utilizam o limite superior e inferior ou o centro e raio dos intervalos como entradas dos problemas, como visto em Neto e Carvalho (2010) e Souza *et al.* (2011).

1.1 Justificativa e Objetivos

Adaptar ou criar algoritmos que consigam manipular dados intervalares tem se mostrado eficaz nos últimos anos. Manipular os próprios intervalos dos dados, ao invés de adaptá-los para serem entradas de métodos clássicos de aprendizagem, possibilita manter sua

integridade, evita perda de informação e pode-se conseguir resultados mais precisos ou mais coerentes com os problemas. Assim, estender o MLM que é um método simples e tem alcançado resultados expressivos na literatura é uma contribuição relevante. Além disso, problemas de classificação com dados intervalares ainda é uma área menos presente, este trabalho propõe uma alternativa para esse tipo de problema.

1.2 Objetivos

Objetivo Geral: estender o MLM para resolver problemas com dados intervalares.

Além desse, são objetivos específicos:

- formular uma variante do MLM considerando o centro e raio dos intervalos como entrada, porém o processo de treinamento e estimação das saídas são feitas de forma independentes como visto em (NETO; CARVALHO, 2010);
- formular uma variante do MLM considerando o limite superior e inferior dos intervalos como entrada, resolvendo o problema de forma análoga a dita acima;
- desenvolver as duas abordagens acima, mas solucionando as otimizações necessárias de forma única;
- Adaptar dois modelos do MLM para classificação: i) utilizando o centro e raio e ii) utilizando limite superior e inferior como as entradas do problema.

Este trabalho está organizado da seguinte forma: o Capítulo 2 detalha a versão original do MLM; Capítulo 3 desenvolve os quatro modelos propostos de MLM para dados intervalares; o Capítulo 4 descreve a metodologia utilizada nos experimentos de validação dos métodos de regressão; Capítulo 5 explica a metodologia usada nos experimentos de classificação; Capítulo 6 apresenta os resultados obtidos em experimentos numéricos (em conjuntos de dados reais e sintéticos) e os discute; e o Capítulo 7 conclui o trabalho.

2 MÁQUINA DE APRENDIZAGEM MÍNIMA

O modelo MLM relaciona as distâncias entre os pontos no espaço de entrada com as distâncias entre eles no espaço de saída. Para realizar essa tarefa, um conjunto de K pontos do conjunto de dados é selecionado como os pontos de referência. Em seguida, é construído um mapeamento linear das distâncias entre cada ponto e cada ponto de referência no espaço de entrada e as distâncias entre cada ponto e cada ponto de referência no espaço de saída. Na etapa de predição, o modelo calcula as distâncias até os pontos de referência no espaço de entrada e as projeta usando o modelo linear para encontrar uma estimativa da matriz de distância do espaço de saída. Finalmente, a matriz de distância de saída, juntamente com os pontos de referência, é usada para estimar a saída final resolvendo um problema de otimização convexa. As principais etapas do MLM são discutidas em mais detalhes nas próximas seções.

2.1 Formulação Básica

O problema de aprendizagem pode ser definido como a tarefa de aproximar uma função contínua, $f : \mathcal{X} \rightarrow \mathcal{Y}$, de um conjunto de dados $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i = f(\mathbf{x}_i) + \varepsilon_i)\}_{i=1}^N$, com $\mathbf{x}_i \in \mathcal{X}$ e $\mathbf{y}_i \in \mathcal{Y}$. Este trabalho refere-se $\mathcal{X} = \mathbb{R}^p$ como o espaço de entrada e $\mathcal{Y} = \mathbb{R}^d$ como o espaço de saída.

A fim de aproximar a função objetivo, f , MLM obtém as transformações $\delta_k : \mathcal{Y} \rightarrow \mathbb{R}_+$ e $d_k : \mathcal{X} \rightarrow \mathbb{R}_+$ dadas pelas funções de distância calculadas para pontos fixos $\{(\mathbf{m}_k, \mathbf{t}_k = f(\mathbf{m}_k)) \in \mathcal{D}\}_{k=1}^K$ que são denotados por ponto de referência. Formalmente, $d_k(\mathbf{x})$ representa a distância Euclidiana entre um ponto \mathbf{x} e um ponto de referência \mathbf{m}_k , enquanto $\delta_k(\mathbf{y})$ representa a distância Euclidiana entre o rótulo correspondente \mathbf{y} e o rótulo do ponto de referência \mathbf{t}_k .

O algoritmo MLM assume que existe um mapeamento linear, g_k , entre os espaços induzidos pelas funções de distância d e δ . Considerando agora o conjunto de dados \mathcal{D} , as distâncias no espaço de entrada entre os pontos de dados e os pontos de referência de entrada são computados em uma matriz $\mathbf{D} \in \mathbb{R}_+^{N \times K}$. De forma similar, a matriz de distância no espaço de saída entre os N pontos de dados e os pontos de referência de saída são representados por $\mathbf{\Delta} \in \mathbb{R}_+^{N \times K}$. Usando esses dados, o MLM encontra o mapeamento g_k de acordo com o modelo $\Delta_{i,k} = g_k(\mathbf{D}_{i,\cdot}) + \varepsilon_i$ para todo $i = 1, \dots, N$. Nesse modelo, ε_n representa os resíduos enquanto $\mathbf{D}_{n,-}$ denota a n -ésima linha da matriz \mathbf{D} . Similarmente, $\Delta_{n,k}$ é o elemento na n -ésima linha e k -ésima coluna de $\mathbf{\Delta}$.

As distâncias no espaço de saída podem ser aproximadas por uma combinação linear de distâncias no espaço de entrada. Colocando os mapeamentos para todos N pontos de dados juntos, o modelo de regressão de distância do MLM é representado em uma matriz dada por $\mathbf{\Delta} = \mathbf{DB} + \boldsymbol{\varepsilon}$, onde a matriz $\mathbf{B} \in \mathbb{R}^{K \times K}$ abrange os K vetores de coeficientes \mathbf{b}_k em suas colunas.

Dado que, o MLM calcula uma função $h_{\mathbf{B}}(\mathbf{x}) : \mathcal{X} \rightarrow \mathcal{Y}$ dado por

$$h_{\mathbf{B}}(\mathbf{x}) = \arg \min_{\mathbf{y}} \sum_{k=1}^K \left[\delta_k^2(\mathbf{y}) - \left(\sum_{i=1}^K d_i(\mathbf{x}) B_{i,k} \right)^2 \right]^2 \quad (2.1)$$

onde $\delta_k(\mathbf{y}) = \|\mathbf{y} - \mathbf{t}_k\|$ é a distância Euclidiana entre \mathbf{y} e o k -ésimo ponto de referência de saída \mathbf{t}_k , $d_i(\mathbf{x}) = \|\mathbf{x} - \mathbf{m}_i\|$ é a distância Euclidiana entre \mathbf{x} e o i -ésimo ponto de referência de entrada \mathbf{m}_i , e K denota o número de pontos de referência.

2.2 Algoritmo de Aprendizagem

O algoritmo de treinamento do MLM pode ser dividido em duas etapas:

1. selecionar os pontos de referência $\{(\mathbf{m}_k; \mathbf{t}_k)\}$;
2. calcular a matriz de regressão \mathbf{B} .

Em relação à seleção de pontos de referência, o MLM original sugere que isso pode ser feito aleatoriamente. Entretanto, procedimentos alternativos podem ser usados para selecionar os pontos de referência, tais como os trabalhos de Alencar (2017) e Dias *et al.* (2018).

O segundo passo do algoritmo consiste em encontrar os regressores lineares \mathbf{B} . Isso pode ser feito usando o método dos mínimos quadrados, dado por

$$\hat{\mathbf{B}} = (\mathbf{D}^\top \mathbf{D})^{-1} \mathbf{D}^\top \mathbf{\Delta}, \quad (2.2)$$

onde \mathbf{D} e $\mathbf{\Delta}$ são as matrizes de distância entre os pontos no conjunto de dados e os pontos de referência no espaço de entrada e saída, respectivamente.

2.3 Estimação da Saída

A estimativa da saída também pode ser dividida em duas etapas:

1. estimar as distâncias para os pontos de referência no espaço de saída;
2. estimar a saída $\hat{\mathbf{y}}$ a partir das distâncias para os pontos no espaço de saída.

Na primeira etapa, utiliza-se o novo ponto \mathbf{x} e computa-se sua distância para cada ponto de referência \mathbf{m}_k , construindo o vetor $[d_1(\mathbf{x}), \dots, d_K(\mathbf{x})]$. Então, estima-se as distâncias

entre a nova amostra e os pontos de referência no espaço de saída usando Eq. (2.3).

$$\hat{\delta}_k(\mathbf{y}) = \sum_{i=1}^K d_i(\mathbf{x}) \hat{B}_{i,k}, \quad \forall k = 1, \dots, K \quad (2.3)$$

As estimativas $\hat{\delta}_1(\mathbf{y}) \dots \hat{\delta}_K(\mathbf{y})$ são utilizadas para encontrar \mathbf{y} , o que pode ser entendido como um problema de multilateração. A Figura 1 mostra este problema graficamente. A localização de \mathbf{y} pode ser estimada a partir da minimização da função objetivo apresentada na Eq. (2.4).

$$J(\mathbf{y}) = \sum_{k=1}^K \left((\mathbf{y} - \mathbf{t}_k)^\top (\mathbf{y} - \mathbf{t}_k) - \hat{\delta}_k^2(\mathbf{y}) \right)^2 \quad (2.4)$$

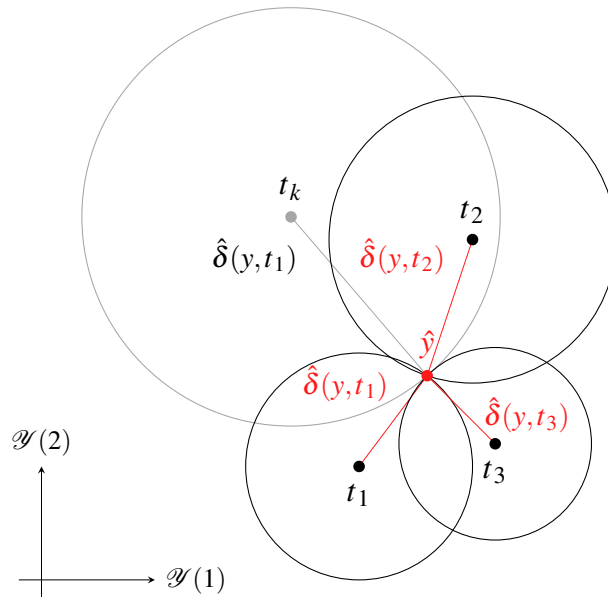


Figura 1 – Exemplificação do problema de multilateração

O MLM original usa o algoritmo Levenberg-Marquardt (LM) para resolver o problema de otimização. Outras opções são discutidas em (ALENCAR, 2017).

2.3.1 MLM para classificação

O MLM pode ser estendido para problemas de classificação de forma direta representando a resposta de cada amostra como um vetor binário através do esquema de codificação 1-de- S , onde S é o número de classes de um conjunto C . Nessa abordagem, apenas uma das componentes do vetor estará "ligada". Matematicamente, o conjunto de saídas $Y = \{\mathbf{y}_i\}_{i=1}^N$ com $\mathbf{y}_i \in \mathbb{R}^S$, que correspondem aos pontos de entrada \mathbf{X} , é então definido de tal forma que a j -ésima componente de \mathbf{y}_i é α se $\mathbf{x}_i = C_j$, senão, será β , onde α e β são escalares inteiros tal que $\alpha > \beta$. Normalmente, escolhas usuais são $\alpha = 1$ e $\beta = 0$ ou $\alpha = 1$ e $\beta = -1$.

Pode-se notar, que para problemas de classificação binária, é possível simplesmente utilizar os escalares inteiros para representar a saída, sendo dispensável a abordagem 1-de-S. Como visto em (MESQUITA *et al.*, 2017a), é apresentando a MLM de Vizinho Mais Próximo (*Nearest Neighborhood Minimal Learning Machine*, NN-MLM), que explora características inerentes ao problema de classificação para acelerar o processo de estimação de saída do MLM. Uma característica particular dos problemas de classificação é que a saída pertence a um conjunto finito de opções, que são as classes. Assim, na estimação da saída não é necessário explorar todos os valores de classe possíveis para determinar aquela que minimiza o erro entre a distância e as estimativas obtidas. Dessa forma, consegue-se estimar a saída apenas verificando qual a classe do ponto de referência que possui a menor diferença entre as distâncias estimadas e a distância que teríamos em cada caso.

Como provado em Mesquita *et al.* (2017a), a classe estimada é dada pelo rótulo do k -ésimo ponto de referência que satisfaz a equação:

$$k^* = \underset{k=1, \dots, K}{\operatorname{argmin}} \delta^2(\mathbf{y}, \mathbf{t}_k), \quad (2.5)$$

onde k^* é o índice do ponto de referência mais próximo da distância estimada, ou seja, a saída é dada por

$$\hat{\mathbf{y}} = \mathbf{t}_{k^*} \quad (2.6)$$

3 MLM PARA DADOS INTERVALARES

Neste capítulo, são apresentadas quatro extensões propostas do MLM para lidar com dados intervalares. Seja $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ um conjunto de N vetores de entrada. Cada vetor de entrada \mathbf{x}_i é representado por um vetor de características intervalar $\mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{ip})$, onde $\mathbf{x}_{ij} = [x_{ij}^L, x_{ij}^U] \in \zeta = \{[x^L, x^U] \mid x^L, x^U \in \mathbb{R}, x^L \leq x^U\}$, com $j \in \{1, 2, \dots, p\}$. Dado $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N$ ser um conjunto de N vetores de saída. Cada vetor de saída \mathbf{y}_i é também representado por um vetor de características intervalar, com $\mathbf{y}_i = (\mathbf{y}_{i1}, \mathbf{y}_{i2}, \dots, \mathbf{y}_{id})$, onde $\mathbf{y}_{is} = [y_{is}^L, y_{is}^U] \in \zeta$, com $s \in \{1, 2, \dots, d\}$.

Assumindo a existência de um mapeamento contínuo $f : \mathcal{X} \rightarrow \mathcal{Y}$ entre o espaço de entrada e de saída, o objetivo é estimar f a partir dos dados com modelos multirespostas

$$\mathbf{Y} = f(\mathbf{X}) + \mathbf{E},$$

onde \mathbf{E} é a matriz de resíduos com N linhas e D colunas.

3.1 MLM com Limite Superior e Inferior Independentes

Neste primeiro método proposto, abreviado de MLM-LU, é considerada uma abordagem utilizando o limite superior e inferior para estimar f . Sejam $\mathbf{X}^L = \{x_{i1}^L, x_{i2}^L, \dots, x_{ip}^L\}_{i=1}^N$ e $\mathbf{X}^U = \{x_{i1}^U, x_{i2}^U, \dots, x_{ip}^U\}_{i=1}^N$ os limites inferiores e superiores dos vetores de entrada, respectivamente. Também, sejam $\mathbf{Y}^L = \{y_{i1}^L, y_{i2}^L, \dots, y_{id}^L\}_{i=1}^N$ e $\mathbf{Y}^U = \{y_{i1}^U, y_{i2}^U, \dots, y_{id}^U\}_{i=1}^N$ os limites inferiores e superiores dos vetores de saída, respectivamente.

Dado um conjunto de K pontos de referência no espaço de entrada, $\mathbf{R}^L = \{\mathbf{m}_k\}_{k=1}^K$ e $\mathbf{R}^U = \{\mathbf{n}_k\}_{k=1}^K$ para os limites inferiores e superiores, respectivamente, com $\mathbf{R}^L \subseteq \mathbf{X}^L$ e $\mathbf{R}^U \subseteq \mathbf{X}^U$, e as saída correspondentes $\mathbf{T}^L = \{\mathbf{t}_k\}_{k=1}^K$ e $\mathbf{T}^U = \{\mathbf{c}_k\}_{k=1}^K$, é definido $\mathbf{D}_{x^L} \in \mathbb{R}^{N \times K}$ como a matriz de distância entre os conjuntos \mathbf{X}^L e \mathbf{R}^L , onde a k -ésima coluna de \mathbf{D}_{x^L} contém as distâncias $d(\mathbf{X}^L, \mathbf{m}_k)$ entre os limites inferiores das entradas \mathbf{x}_i^L e o k -ésimo ponto de referência do espaço de entrada \mathbf{m}_k . Também é definido $\mathbf{\Delta}_{y^L} \in \mathbb{R}^{N \times K}$ como a matriz de distância entre os conjuntos \mathbf{Y}^U e \mathbf{T}^U , de tal forma que a k -ésima coluna de $\mathbf{\Delta}_{y^L}$ contém as distâncias $\delta(\mathbf{Y}^L, \mathbf{t}_k)$ entre os limites inferiores de saída \mathbf{y}_i^L e o k -ésimo ponto de referência no espaço de saída \mathbf{t}_k .

Similarmente, define-se $\mathbf{D}_{x^U} \in \mathbb{R}^{N \times K}$ como a matriz de distância entre os conjuntos \mathbf{X}^U e \mathbf{R}^U . Define-se também $\mathbf{\Delta}_{y^U} \in \mathbb{R}^{N \times K}$ como a matriz de distância entre \mathbf{Y}^U e \mathbf{T}^U .

Como na formulação original do MLM, assume-se que existem mapeamentos entre as matrizes de distância de entrada e saída, e que eles têm uma estrutura linear para cada resposta.

O modelo de regressão tem a forma:

$$\mathbf{\Delta}_{y^L} = \mathbf{D}_{x^L} \mathbf{B}_L + \mathbf{E}_L \quad (3.1)$$

$$\mathbf{\Delta}_{y^U} = \mathbf{D}_{x^U} \mathbf{B}_U + \mathbf{E}_U \quad (3.2)$$

As colunas das matrizes de regressão $\mathbf{B}_L \in \mathbb{R}^{K \times K}$ e $\mathbf{B}_U \in \mathbb{R}^{K \times K}$ correspondem aos coeficientes para a K -ésima resposta. Considerando que as matrizes de regressão \mathbf{B}_L e \mathbf{B}_U são independentes, pode-se estimá-las separadamente através da minimização da soma residual multivariada dos quadrados:

$$RSS(\mathbf{B}_L) = \text{tr}((\mathbf{\Delta}_{y^L} - \mathbf{D}_{x^L} \mathbf{B}_L)^\top (\mathbf{\Delta}_{y^L} - \mathbf{D}_{x^L} \mathbf{B}_L)) \quad (3.3)$$

$$RSS(\mathbf{B}_U) = \text{tr}((\mathbf{\Delta}_{y^U} - \mathbf{D}_{x^U} \mathbf{B}_U)^\top (\mathbf{\Delta}_{y^U} - \mathbf{D}_{x^U} \mathbf{B}_U)) \quad (3.4)$$

Novamente, como no MLM original, se o problema for sobredeterminado sem solução, usa-se a solução aproximada fornecida pela estimativa dos mínimos quadrados. Se Eq. (3.1) é sobredeterminada, então \mathbf{B}_L é estimado a partir de

$$\hat{\mathbf{B}}_L = (\mathbf{D}_{x^L}^\top \mathbf{D}_{x^L})^{-1} \mathbf{D}_{x^L}^\top \mathbf{\Delta}_{y^L} \quad (3.5)$$

Se Eq. (3.1) é unicamente determinada, existe uma única solução dada por

$$\hat{\mathbf{B}}_L = \mathbf{D}_{x^L}^{-1} \mathbf{\Delta}_{y^L} \quad (3.6)$$

De forma semelhante, se Eq. (3.2) é sobredeterminada, então a matriz \mathbf{B}_U é estimada a partir de:

$$\hat{\mathbf{B}}_U = (\mathbf{D}_{x^U}^\top \mathbf{D}_{x^U})^{-1} \mathbf{D}_{x^U}^\top \mathbf{\Delta}_{y^U} \quad (3.7)$$

Se Eq. (3.2) é unicamente determinada, existe uma única solução dada por

$$\hat{\mathbf{B}}_U = \mathbf{D}_{x^U}^{-1} \mathbf{\Delta}_{y^U} \quad (3.8)$$

Depois de estimar as matrizes $\hat{\mathbf{B}}_L$ e $\hat{\mathbf{B}}_U$, considera-se um ponto de teste de entrada $\mathbf{x} = [\mathbf{x}^L, \mathbf{x}^U]$, onde $\mathbf{x}^L \in \mathbb{R}^p$ é um vetor contendo todos os limites inferiores e $\mathbf{x}^U \in \mathbb{R}^p$ é um vetor contendo todos os limites superiores. As distâncias entre \mathbf{x}^L e $\{\mathbf{m}_k\}_{k=1}^K$ são coletadas em um vetor $\mathbf{d}_L(\mathbf{x}^L, \mathbf{R}^L) = [d(\mathbf{x}^L, \mathbf{m}_1), \dots, d(\mathbf{x}^L, \mathbf{m}_k)]$ e as distâncias entre \mathbf{x}^U e $\{\mathbf{n}_k\}_{k=1}^K$ são adicionadas em um vetor $\mathbf{d}_U(\mathbf{x}^U, \mathbf{R}^U) = [d(\mathbf{x}^U, \mathbf{n}_1), \dots, d(\mathbf{x}^U, \mathbf{n}_k)]$. As distâncias correspondentes entre as saídas desconhecidas, \mathbf{y}^L e \mathbf{y}^U , e os pontos de referência no espaço de saída conhecidos, $\{\mathbf{t}_k\}_{k=1}^K$ e $\{\mathbf{c}_h\}_{h=1}^H$, são estimados como

$$\hat{\delta}(\mathbf{y}^L, \mathbf{T}^L) = \mathbf{d}_L(\mathbf{x}^L, \mathbf{R}^L) \hat{\mathbf{B}}_L \quad (3.9)$$

$$\hat{\delta}(\mathbf{y}^U, \mathbf{T}^U) = \mathbf{d}_U(\mathbf{x}^U, \mathbf{R}^U) \hat{\mathbf{B}}_U \quad (3.10)$$

O problema de estimar as saídas \mathbf{y}^L e \mathbf{y}^U a partir das distâncias $\hat{\delta}(\mathbf{y}^L, \mathbf{T}^L)$ e $\hat{\delta}(\mathbf{y}^U, \mathbf{T}^U)$ pode ser entendido como uma multilateração da estimativa das localizações em \mathcal{D} . Similar ao MLM, a localização de \mathbf{y}^L pode ser estimada a partir da minimização da função objetivo:

$$J(\mathbf{y}^L) = \sum_{k=1}^K \left((\mathbf{y}^L - \mathbf{t}_k)^\top (\mathbf{y}^L - \mathbf{t}_k) - \hat{\delta}(\mathbf{y}^L, \mathbf{t}_k) \right)^2 \quad (3.11)$$

Um problema semelhante deve ser resolvido para estimar \mathbf{y}^U . No entanto, uma restrição deve ser incluída no problema de otimização para garantir matematicamente a integridade do intervalo de saída, isto é, $\mathbf{y}^L \leq \mathbf{y}^U$. Então, o propósito é resolver o seguinte problema de otimização com restrição:

$$\begin{aligned} \text{minimize}_{\mathbf{y}^U} \quad & \sum_{k=1}^K \left((\mathbf{y}^U - \mathbf{c}_k)^\top (\mathbf{y}^U - \mathbf{c}_k) - \hat{\delta}(\mathbf{y}^U, \mathbf{c}_k) \right)^2 \\ \text{s.t.} \quad & \mathbf{y}^L \leq \mathbf{y}^U \end{aligned} \quad (3.12)$$

Esse problema de otimização pode ser resolvido com os Mínimos Quadrados Não Lineares (NLS - *Nonlinear Least Squares*).

3.2 MLM com Centro e Raio Independentes

Nesta segunda variante proposta do MLM (referenciada como MLM-CR), utiliza-se uma abordagem com o centro e raio dos intervalos. Considerando as seguintes definições: $\mathbf{X}^C = \{x_{i1}^C, x_{i2}^C, \dots, x_{ip}^C\}_{i=1}^N$, $\mathbf{X}^R = \{x_{i1}^R, x_{i2}^R, \dots, x_{ip}^R\}_{i=1}^N$, $\mathbf{Y}^C = \{y_{i1}^C, y_{i2}^C, \dots, y_{id}^C\}_{i=1}^N$ e $\mathbf{Y}^R = \{y_{i1}^R, y_{i2}^R, \dots, y_{id}^R\}_{i=1}^N$, onde $x_{ij}^C = \frac{x_{ij}^L + x_{ij}^U}{2}$, $x_{ij}^R = \frac{x_{ij}^U - x_{ij}^L}{2}$, $y_{is}^C = \frac{y_{is}^L + y_{is}^U}{2}$ e $y_{is}^R = \frac{y_{is}^U - y_{is}^L}{2}$, com $j \in \{1, 2, \dots, p\}$ e $s \in \{1, 2, \dots, d\}$.

Dado um conjunto de K pontos de referência de entrada, $\mathbf{R}^C = \{\mathbf{m}_k\}_{k=1}^K$ e $\mathbf{R}^R = \{\mathbf{n}_k\}_{k=1}^K$ são o centro e o raio, respectivamente, com $\mathbf{R}^C \subseteq \mathbf{X}^C$ e $\mathbf{R}^R \subseteq \mathbf{X}^R$, e as saídas correspondentes $\mathbf{T}^C = \{\mathbf{t}_k\}_{k=1}^K$ e $\mathbf{T}^R = \{\mathbf{c}_k\}_{k=1}^K$, sendo $\mathbf{D}_{x^C} \in \mathbb{R}^{N \times K}$ a matriz de distância entre os conjuntos \mathbf{X}^C e \mathbf{R}^C , onde a k -ésima coluna de \mathbf{D}_{x^C} contém as distâncias $d(\mathbf{X}^C, \mathbf{m}_k)$ entre os centros dos pontos de entrada \mathbf{x}_i^C e o k -ésimo centro do ponto de referência de entrada \mathbf{m}_k . Define-se $\mathbf{\Delta}_C \in \mathbb{R}^{N \times K}$ tal que sua k -ésima coluna contém as distâncias $\delta(\mathbf{Y}^C, \mathbf{t}_k)$ entre os centros dos pontos de saída \mathbf{y}_i^C e o k -ésimo centro do ponto de referência de saída \mathbf{t}_k . Analogamente, sendo $\mathbf{D}_{x^R} \in \mathbb{R}^{N \times K}$ a matriz de distância entre os conjuntos \mathbf{X}^R e \mathbf{R}^R , e $\mathbf{\Delta}_{y^R} \in \mathbb{R}^{N \times K}$ sendo a matriz de distância entre os conjuntos \mathbf{Y}^R e \mathbf{T}^R .

Assume-se que as matrizes de distância de entrada e de saída são mapeadas por uma função e que elas têm uma estrutura linear para cada resposta. Então, o modelo de regressão tem

o formato:

$$\mathbf{\Delta}_{y^C} = \mathbf{D}_{x^C} \mathbf{B}_C + \mathbf{E}_C \quad (3.13)$$

$$\mathbf{\Delta}_{y^R} = \mathbf{D}_{x^R} \mathbf{B}_R + \mathbf{E}_R \quad (3.14)$$

Como no modelo proposto anteriormente, as colunas das matrizes de regressão $\mathbf{B}_C \in \mathbb{R}^{K \times K}$ e $\mathbf{B}_R \in \mathbb{R}^{K \times K}$ correspondem aos coeficientes das respostas. Assumindo que as matrizes de regressão \mathbf{B}_C e \mathbf{B}_R são independentes, é possível estimar separadamente através da minimização da soma dos quadrados dos resíduos multivariados como função de perda:

$$RSS(\mathbf{B}_C) = \text{tr}((\mathbf{\Delta}_{y^C} - \mathbf{D}_{x^C} \mathbf{B}_C)^\top (\mathbf{\Delta}_{y^C} - \mathbf{D}_{x^C} \mathbf{B}_C)) \quad (3.15)$$

$$RSS(\mathbf{B}_R) = \text{tr}((\mathbf{\Delta}_{y^R} - \mathbf{D}_{x^R} \mathbf{B}_R)^\top (\mathbf{\Delta}_{y^R} - \mathbf{D}_{x^R} \mathbf{B}_R)) \quad (3.16)$$

Pode-se estimar \mathbf{B}_C e \mathbf{B}_R a partir de

$$\hat{\mathbf{B}}_C = (\mathbf{D}_{x^C}^\top \mathbf{D}_{x^C})^{-1} \mathbf{D}_{x^C}^\top \mathbf{\Delta}_{y^C} \quad (3.17)$$

$$\hat{\mathbf{B}}_R = (\mathbf{D}_{x^R}^\top \mathbf{D}_{x^R})^{-1} \mathbf{D}_{x^R}^\top \mathbf{\Delta}_{y^R} \quad (3.18)$$

Depois de computar $\hat{\mathbf{B}}_C$ e $\hat{\mathbf{B}}_R$, considera-se um ponto de teste de entrada $\mathbf{x} = [\mathbf{x}^C, \mathbf{x}^R]$, onde $\mathbf{x}^C \in \mathbb{R}^p$ é um vetor contendo todos p centros e $\mathbf{x}^R \in \mathbb{R}^p$ é um vetor contendo todos p raios. Também, considera-se que as distâncias entre \mathbf{x}^C e $\{\mathbf{m}_k\}_{k=1}^K$ estão reunidas em um vetor $\mathbf{d}_C(\mathbf{x}^C, \mathbf{R}^C) = [d(\mathbf{x}^C, \mathbf{m}_1), \dots, d(\mathbf{x}^C, \mathbf{m}_k)]$ e que as distâncias entre \mathbf{x}^R e $\{\mathbf{n}_k\}_{k=1}^K$ são coletadas em um vetor $\mathbf{d}_R(\mathbf{x}^R, \mathbf{R}^R) = [d(\mathbf{x}^R, \mathbf{n}_1), \dots, d(\mathbf{x}^R, \mathbf{n}_k)]$. As distâncias correspondentes das saídas desconhecidas, \mathbf{y}^C e \mathbf{y}^R , para os pontos de referência de saída, $\{\mathbf{t}_k\}_{k=1}^K$ e $\{\mathbf{c}_k\}_{k=1}^K$, podem ser estimadas como

$$\hat{\delta}(\mathbf{y}^C, \mathbf{T}^C) = \mathbf{d}_C(\mathbf{x}^C, \mathbf{R}^C) \hat{\mathbf{B}}_C \quad (3.19)$$

$$\hat{\delta}(\mathbf{y}^R, \mathbf{T}^R) = \mathbf{d}_R(\mathbf{x}^R, \mathbf{R}^R) \hat{\mathbf{B}}_R \quad (3.20)$$

Mais uma vez, enfrenta-se um problema de multilateração que pode ser resolvido de forma semelhante à solução adotada na abordagem anterior. Na primeira etapa, pode-se utilizar a abordagem LM para encontrar \mathbf{y}^C minimizando a seguinte função objetivo:

$$J(\mathbf{y}^C) = \sum_{k=1}^K \left((\mathbf{y}^C - \mathbf{t}_k)^\top (\mathbf{y}^C - \mathbf{t}_k) - \hat{\delta}(\mathbf{y}^C, \mathbf{t}_k) \right)^2 \quad (3.21)$$

A fim de obter \mathbf{y}^R , é necessário adicionar uma restrição para garantir que $\mathbf{y}^R \geq 0$. Como resultado, consegue-se \mathbf{y}^R resolvendo o seguinte problema de otimização com restrição:

$$\begin{aligned} \underset{\mathbf{y}^R}{\text{minimize}} \quad & \sum_{k=1}^K \left((\mathbf{y}^R - \mathbf{c}_k)^\top (\mathbf{y}^R - \mathbf{c}_k) - \hat{\delta}(\mathbf{y}^R, \mathbf{c}_k) \right)^2 \\ \text{s.t.} \quad & \mathbf{y}^R \geq 0 \end{aligned} \quad (3.22)$$

Novamente, pode-se usar NLS para resolver esse problema.

3.3 MLM com Limite Superior e Inferior Dependentes

Nesta terceira abordagem proposta (abreviada como MLM-LUT), utiliza-se também os limites inferiores e superiores como entrada, mas diferente da primeira variante, o problema não é considerado independente, assim, é possível tratar casos em que os intervalos são relacionados entre si. Como na Seção 3.1, sejam $\mathbf{X}^L = \{x_{i1}^L, x_{i2}^L, \dots, x_{ip}^L\}_{i=1}^N$ e $\mathbf{X}^U = \{x_{i1}^U, x_{i2}^U, \dots, x_{ip}^U\}_{i=1}^N$ os limites inferiores e superiores dos vetores de entrada, respectivamente. Também, sejam $\mathbf{Y}^L = \{y_{i1}^L, y_{i2}^L, \dots, y_{id}^L\}_{i=1}^N$ e $\mathbf{Y}^U = \{y_{i1}^U, y_{i2}^U, \dots, y_{id}^U\}_{i=1}^N$ os limites inferiores e superiores dos vetores de saída, respectivamente. Agora é definida uma nova matriz $\mathbf{X}^{LU}_{N \times 2p}$, onde seus valores são gerados a partir da concatenação de \mathbf{X}^L e \mathbf{X}^U . Analogamente, também é definido o conjunto de saída $\mathbf{Y}^{LU}_{N \times 2d}$ a partir da concatenação de \mathbf{Y}^L e \mathbf{Y}^U .

Dado um conjunto de K pontos de referência no novo espaço de entrada, \mathbf{X}^{LU} , $\mathbf{R}^{LU} = \{\mathbf{m}_k\}_{k=1}^K$, com $\mathbf{R}^{LU} \subseteq \mathbf{X}^{LU}$, e as saídas correspondentes $\mathbf{T}^{LU} = \{\mathbf{t}_k\}_{k=1}^K$, é definido $\mathbf{D}_{x^{LU}} \in \mathbb{R}^{N \times K}$ como a matriz de distância entre os conjuntos \mathbf{X}^{LU} e \mathbf{R}^{LU} , onde a k -ésima coluna de $\mathbf{D}_{x^{LU}}$ contém as distâncias $d(\mathbf{x}_i^{LU}, \mathbf{m}_k)$. Também é definido $\mathbf{\Delta}_{y^{LU}} \in \mathbb{R}^{N \times K}$ como a matriz de distância entre os conjuntos \mathbf{Y}^{LU} e \mathbf{T}^{LU} de tal forma que a k -ésima coluna de $\mathbf{\Delta}_{y^{LU}}$ contém as distâncias $\delta(\mathbf{y}_i^{LU}, \mathbf{t}_k)$.

Como na formulação original do MLM e nas anteriores, assume-se a existência de um mapeamento entre as matrizes de distância de entrada e saída, e que eles possuem uma estrutura linear para cada resposta. Assim, o novo modelo de regressão é

$$\mathbf{\Delta}_{y^{LU}} = \mathbf{D}_{x^{LU}} \mathbf{B}_{LU} + \mathbf{E}_{LU} \quad (3.23)$$

Como visto anteriormente, as colunas de matrizes de regressão $\mathbf{B}_{LU} \in \mathbb{R}^{K \times K}$ correspondem aos coeficientes das respostas. Aqui, é considerado que os limites inferiores e superiores dos intervalos de saída são relacionados com os limites inferiores e superiores das entradas, assim, os coeficientes não são mais calculados de forma independente. \mathbf{B}_{LU} é estimado a partir

da minimização da soma dos quadrados dos resíduos multivariados como a seguir:

$$RSS(\mathbf{B}_{LU}) = \text{tr}((\mathbf{\Delta}_{y^{LU}} - \mathbf{D}_{x^{LU}}\mathbf{B}_{LU})^\top (\mathbf{\Delta}_{y^{LU}} - \mathbf{D}_{x^{LU}}\mathbf{B}_{LU})) \quad (3.24)$$

Embora os coeficientes não sejam calculados separadamente, ainda é possível computá-los a partir de

$$\hat{\mathbf{B}}_{LU} = (\mathbf{D}_{x^{LU}}^\top \mathbf{D}_{x^{LU}})^{-1} \mathbf{D}_{x^{LU}}^\top \mathbf{\Delta}_{y^{LU}} \quad (3.25)$$

Após computar $\hat{\mathbf{B}}_{LU}$, considera-se um ponto de teste de entrada $\mathbf{x} = \mathbf{x}^{LU}$, onde $\mathbf{x}^{LU} \in \mathbb{R}^{2p}$ é um vetor contendo a concatenação dos limites superiores e inferiores dos intervalos de \mathbf{x} . Também, considera-se que as distâncias entre \mathbf{x}^{LU} e $\{\mathbf{m}_k\}_{k=1}^K$ estão em um vetor $\mathbf{d}_{LU}(\mathbf{x}^{LU}, \mathbf{R}^{LU}) = [d(\mathbf{x}^{LU}, \mathbf{m}_1), \dots, d(\mathbf{x}^{LU}, \mathbf{m}_k)]$. As distâncias correspondentes da saída desconhecida, \mathbf{y}^{LU} para os pontos de referência de saída, $\{\mathbf{t}_k\}_{k=1}^K$, podem ser estimadas como

$$\hat{\delta}(\mathbf{y}^{LU}, \mathbf{T}^{LU}) = \mathbf{d}_{LU}(\mathbf{x}^{LU}, \mathbf{R}^{LU}) \hat{\mathbf{B}}_{LU} \quad (3.26)$$

Por fim, chega-se a um problema de multilateração onde a saída é $\mathbf{y}^{LU} = [\mathbf{y}^L, \mathbf{y}^U]$. Agora, é necessário que o vetor de saída tenha suas d primeiras componentes menores ou iguais que as d últimas, ou seja, os limites inferiores estimados tem que ser menores ou iguais aos limites superiores estimados para que a integridade do vetor de intervalos de saída seja mantida. Para garantir isso matematicamente, é adicionada uma nova restrição e a saída é estimada minimizando a seguinte função objetivo:

$$\begin{aligned} & \underset{\mathbf{y}^{LU}}{\text{minimize}} \sum_{k=1}^K \left((\mathbf{y}^{LU} - \mathbf{t}_k^{LU})^\top (\mathbf{y}^{LU} - \mathbf{t}_k^{LU}) - \hat{\delta}(\mathbf{y}^{LU}, \mathbf{t}_k^{LU}) \right)^2 \\ & \text{s.t. } \mathbf{y}^L \leq \mathbf{y}^U \end{aligned} \quad (3.27)$$

Esse é um problema com uma função não linear e restrita que pode ser minimizada através de vários métodos, neste trabalho foi utilizado otimização de pontos interiores.

3.4 MLM com centro e raio Dependentes

Nesta seção, é apresentada a quarta extensão do MLM (MLM-CRT) para regressão intervalar, aqui, como visto na Seção 3.3, também é assumido que os intervalos são relacionados entre si, ou seja, os limites superiores e inferiores da saída estão relacionados com os limites inferiores e superiores das entradas.

Como na Seção 3.2, sejam $\mathbf{X}^C = \{x_{i1}^C, x_{i2}^C, \dots, x_{ip}^C\}_{i=1}^N$, $\mathbf{X}^R = \{x_{i1}^R, x_{i2}^R, \dots, x_{ip}^R\}_{i=1}^N$, $\mathbf{Y}^C = \{y_{i1}^C, y_{i2}^C, \dots, y_{id}^C\}_{i=1}^N$ e $\mathbf{Y}^R = \{y_{i1}^R, y_{i2}^R, \dots, y_{id}^R\}_{i=1}^N$, onde $x_{ij}^C = \frac{x_{ij}^L + x_{ij}^U}{2}$, $x_{ij}^R = \frac{x_{ij}^U - x_{ij}^L}{2}$, $y_{is}^C = \frac{y_{is}^L + y_{is}^U}{2}$ e $y_{is}^R = \frac{y_{is}^U - y_{is}^L}{2}$, com $j \in \{1, 2, \dots, p\}$ e $s \in \{1, 2, \dots, d\}$. É construído uma matriz $\mathbf{X}_{N \times 2p}^{CR}$ tal que os elementos de suas p primeiras colunas são \mathbf{X}^C e as p últimas são \mathbf{X}^R . De forma similar, é definido o conjunto de saída \mathbf{Y}^{CR} a partir de \mathbf{Y}^C e \mathbf{Y}^R .

Sendo $\mathbf{R}^{CR} \subseteq \mathbf{X}^{CR}$ um conjunto de pontos de referências no novo espaço de entrada, \mathbf{X}^{CR} , tal que $\mathbf{R}^{CR} = \{\mathbf{m}_k\}_{k=1}^K$ e as saídas correspondentes $\mathbf{T}^{CR} = \{\mathbf{t}_k\}_{k=1}^K$, é definido $\mathbf{D}_{x^{CR}} \in \mathbb{R}^{N \times K}$ como a matriz de distância entre os conjuntos \mathbf{X}^{CR} e \mathbf{R}^{CR} , onde a k -ésima coluna de $\mathbf{D}_{x^{CR}}$ contém as distâncias $d(\mathbf{x}_i^{CR}, \mathbf{m}_k)$. A matriz de distâncias entre os conjuntos \mathbf{Y}^{CR} e \mathbf{T}^{CR} é definida como $\mathbf{\Delta}_{y^{CR}} \in \mathbb{R}^{N \times K}$ tal que a k -ésima coluna de $\mathbf{\Delta}_{y^{CR}}$ contém as distâncias $\delta(\mathbf{y}_i^{CR}, \mathbf{t}_k^{CR})$.

Assumindo o mapeamento linear entre as matrizes de distância de entrada e saída, o novo modelo de regressão é

$$\mathbf{\Delta}_{y^{CR}} = \mathbf{D}_{x^{CR}} \mathbf{B}_{CR} + \mathbf{E}_{CR} \quad (3.28)$$

As colunas da matriz de regressão $\mathbf{B}_{CR} \in \mathbb{R}^{K \times K}$ correspondem aos coeficientes das respostas. Agora, os limites superiores e inferiores de saída tem dependência dos limites de entrada. Assim, os coeficientes não são mais calculados de forma independente. \mathbf{B}_{CR} é estimado a partir da minimização de

$$RSS(\mathbf{B}_{CR}) = \text{tr}((\mathbf{\Delta}_{y^{CR}} - \mathbf{D}_{x^{CR}} \mathbf{B}_{CR})^\top (\mathbf{\Delta}_{y^{CR}} - \mathbf{D}_{x^{CR}} \mathbf{B}_{CR})) \quad (3.29)$$

Computa-se os coeficientes da seguinte forma:

$$\hat{\mathbf{B}}_{CR} = (\mathbf{D}_{x^{CR}}^\top \mathbf{D}_{x^{CR}})^{-1} \mathbf{D}_{x^{CR}}^\top \mathbf{\Delta}_{y^{CR}} \quad (3.30)$$

Com $\hat{\mathbf{B}}_{CR}$ calculado, um ponto de teste de entrada $\mathbf{x} = \mathbf{x}^{CR}$, em que $\mathbf{x}^{CR} \in \mathbb{R}^{2p}$ é um vetor contendo a concatenação dos centros e raios dos intervalos de \mathbf{x} . Tem-se também que as distâncias entre \mathbf{x}^{CR} e $\{\mathbf{m}_k^{CR}\}_{k=1}^K$ são alocados em um vetor $\mathbf{d}_{CR}(\mathbf{x}^{CR}, \mathbf{R}^{CR}) = [d(\mathbf{x}^{CR}, \mathbf{m}_1), \dots, d(\mathbf{x}^{CR}, \mathbf{m}_k^{CR})]$. As distâncias correspondentes da saída desconhecida, \mathbf{y}^{CR} , para os pontos de referência de saída, $\{\mathbf{t}_k^{CR}\}_{k=1}^K$, são estimadas como

$$\hat{\delta}(\mathbf{y}^{CR}, \mathbf{T}^{CR}) = \mathbf{d}_{CR}(\mathbf{x}^{CR}, \mathbf{R}^{CR}) \hat{\mathbf{B}}_{CR} \quad (3.31)$$

Agora, têm-se um problema de multilateração onde a saída é $\mathbf{y}^{CR} = [\mathbf{y}^C, \mathbf{y}^R]$. Para manter a integridade matemática do intervalo de saída, é necessário que o vetor de saída tenha

suas d últimas colunas maiores que 0, ou seja, os raios dos intervalos estimados tem que ser maiores ou iguais a 0. Para garantir isso matematicamente, é adicionada uma nova restrição e a saída é estimada minimizando a seguinte função objetivo:

$$\begin{aligned} & \underset{\mathbf{y}}{\text{minimize}} e_{\mathbf{y}}^{CR} \sum_{k=1}^K \left((\mathbf{y}^{CR} - \mathbf{t}_k^{CR})^\top (\mathbf{y}^{CR} - \mathbf{t}_k^{CR}) - \hat{\delta}(\mathbf{y}^{CR}, \mathbf{t}_k^{CR}) \right)^2 \\ & \text{s.t. } \mathbf{y}^R \geq 0 \end{aligned} \quad (3.32)$$

De novo, a função objetivo é não linear e com restrições; neste trabalho foi utilizado otimização de pontos interiores.

3.5 Classificação

Os modelos vistos nas Seções 3.3 e 3.4 também podem ser facilmente estendidos para problemas de classificação; a única alteração necessária é na estimação dos rótulos, pois as saídas são vetores de classes no esquema $1 - de - S$ e não intervalos, assim não sendo preciso resolver um problema de multilateração. Esse novo problema pode ser solucionado como proposto em Mesquita *et al.* (2017a) (NN-MLM), pois é mais simples e menos custoso, assim:

$$\hat{\mathbf{y}} = \mathbf{t}_{k^*} \quad (3.33)$$

onde k^* é o índice do ponto de referência mais próximo da distância estimada:

$$k^* = \underset{k=1, \dots, K}{\text{argmin}} \delta^2(\mathbf{y}, \mathbf{t}_k), \quad (3.34)$$

4 CONFIGURAÇÕES DOS EXPERIMENTOS NA TAREFA DE REGRESSÃO

Para validar as extensões do MLM, foram realizados experimentos com 3 tipos de dados intervalares diferentes: o primeiro é sintético, não linear e gerado a partir dos centros e raios; o segundo também é sintético e não linear, mas agora considera-se que os limites inferiores e superiores de saída são diretamente relacionados com os limites das entradas; e o terceiro são dados intervalares reais. Com o intuito de comparação, foram implementados cinco algoritmos para regressão não linear: três variantes do *Extreme Learning Machine* (ELM) e dois da Regressão de Kernel. Para medir o desempenho dos métodos, o *Root Mean Square Error Bounds* é utilizado. Nas seções a seguir, a metodologia é melhor detalhada.

4.1 Métodos Regressão

A fim de avaliar os métodos propostos, este trabalho os comparou com cinco algoritmos de regressão não-linear. A seguir, é descrito brevemente cada método considerado neste estudo.

4.1.1 *Extreme Learning Machine*

Uma extensão para dados intervalares do *Extreme Learning Machine* (abreviado por iELM) é visto em ??). São propostos três modelos iELM, o primeiro compreende dois ELMs independentes para aproximar os valores do centro e do raio dos intervalos de saída. Uma abordagem semelhante é usada no segundo modelo, em que dois ELMs são utilizados para aproximar os limites superiores e inferiores da saída. O terceiro modelo é uma versão modificada da abordagem inferior-superior, onde uma restrição é adicionada para garantir que o limite inferior seja sempre menor ou igual ao limite superior.

4.1.1.1 *Modelo 1*

O primeiro modelo (iELM-1) utiliza os centros e raios dos intervalos de entrada para prever os intervalos de saída. Em ??), assume-se a independência entre as duas variáveis, matematicamente o novo ELM intervalar é dado por:

$$\begin{cases} \sum_{i=1}^L \beta_i g(\mathbf{x}_j^C) = \sum_{i=1}^L \beta_i g(w_i \mathbf{x}_j^C + p_i) = \mathbf{y}_j^C \\ \sum_{i=1}^L \beta_i g(\mathbf{x}_j^R) = \sum_{i=1}^L \beta_i g(w_i \mathbf{x}_j^R + p_i) = \mathbf{y}_j^R \end{cases} \quad (4.1)$$

onde $j = \{1, \dots, N\}$, w_i é o vetor de pesos que conecta o i -ésimo nó da camada oculta com os nós da camada de entrada, β_i é o vetor de pesos que conecta o i -ésimo nó da camada oculta com os nós da camada de saída, p_i é o *threshold* do i -ésimo nó oculto, L é o número de neurônios da rede e g é a função de ativação, normalmente utiliza-se a função sigmoide.

Dadas as matrizes $\mathbf{H}_1 = g(\mathbf{X}_{N \times p}^C \mathbf{W}_{p \times L})$, $\mathbf{H}_2 = g(\mathbf{X}_{N \times p}^R \mathbf{W}_{p \times L})$, $\boldsymbol{\beta}_{L \times d}$, a Eq. 4.2 pode ser reescrita de forma reduzida como

$$\begin{cases} \mathbf{H}_1 \boldsymbol{\beta} = \mathbf{Y}^C \\ \mathbf{H}_2 \boldsymbol{\beta} = \mathbf{Y}^R \end{cases} \quad (4.2)$$

Sejam $\mathbf{A} = \begin{pmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{pmatrix}$ e $\mathbf{B} = \begin{pmatrix} \mathbf{Y}^C \\ \mathbf{Y}^R \end{pmatrix}$, temos que

$$\mathbf{A} \boldsymbol{\beta} = \mathbf{B}, \quad (4.3)$$

desta forma, é possível resolver esse sistema linear utilizando a técnica de otimização dos mínimos quadrados:

$$\hat{\boldsymbol{\beta}} = \mathbf{A}^\dagger \mathbf{B}. \quad (4.4)$$

4.1.1.2 Modelo 2

Esta abordagem é semelhante à primeira, porém são considerados como entrada os limites superiores e inferiores dos intervalos, a saída gerada também são os limites. Essa abordagem será abreviada como (iELM-2). Matematicamente, o modelo é dado por

$$\begin{cases} \sum_{i=1}^L \beta_i g(\mathbf{x}_j^L) = \sum_{i=1}^L \beta_i g(w_i \mathbf{x}_j^L + p_i) = \mathbf{y}_j^L \\ \sum_{i=1}^L \beta_i g(\mathbf{x}_j^U) = \sum_{i=1}^L \beta_i g(w_i \mathbf{x}_j^U + p_i) = \mathbf{y}_j^U \end{cases} \quad (4.5)$$

A resolução da Eq. 4.5 é análoga à vista no primeiro modelo.

4.1.1.3 Modelo 3

Nos dois primeiros modelos apresentados acima, não existe garantia que os intervalos de saída vão ser íntegros, ou seja, pode acontecer casos em que $\mathbf{y}_i^R < 0$ ou $\mathbf{y}_i^L > \mathbf{y}_i^U$. Para contornar esse problema, esta terceira variante além de usar os limites dos intervalos como entrada, adiciona ao problema uma restrição de que o limite inferior de saída não pode ser maior que o limite

superior. O modelo é formulado como:

$$\begin{cases} \sum_{i=1}^L \beta_i g(\mathbf{x}_j^L) = \sum_{i=1}^L \beta_i g(w_i \mathbf{x}_j^L + p_i) = \mathbf{y}_j^L \\ \sum_{i=1}^L \beta_i g(\mathbf{x}_j^U) = \sum_{i=1}^L \beta_i g(w_i \mathbf{x}_j^U + p_i) = \mathbf{y}_j^U \\ \sum_{i=1}^L \beta_i g(\mathbf{x}_j^L) \leq \sum_{i=1}^L \beta_i g(\mathbf{x}_j^U) \end{cases} \quad (4.6)$$

Esse é um problema de otimização quadrática com restrições lineares.

4.1.2 Interval Kernel Regression

Em Fagundes *et al.* (2014), são propostos versões da regressão de kernel que usa dados intervalares como entrada. Dois modelos são considerados, o primeiro estima os limites dos intervalos de saída em relação a uma função suave dos centros dos intervalos de entrada (abreviado como RKCI). O segundo modelo estima os limites dos intervalos de saída através de duas funções suaves, uma para o centro e outra para o raio dos intervalos de entrada (RKCIA).

4.1.2.1 Modelo 1

Neste primeiro modelo, a relação não linear entre as entradas e saída são as seguintes:

$$E(Y/\mathbf{X}) = \left[E(Y^L/\mathbf{X}^C), E(Y^U/\mathbf{X}^C) \right] \quad (4.7)$$

$$= \left[m^L(\mathbf{X}^C), m^U(\mathbf{X}^C) \right] \quad (4.8)$$

onde m^L e m^U são funções desconhecidas.

Este modelo baseia-se no estimador de Nadayara-Watson, assim, as funções de ponderação para o centro dos intervalos são dadas por

$$\hat{m}^L(\mathbf{x}^C) = \sum_{i=1}^N \omega_i^C \mathbf{y}_i^L \quad (4.9)$$

$$\hat{m}^U(\mathbf{x}^C) = \sum_{i=1}^N \omega_i^C \mathbf{y}_i^U \quad (4.10)$$

onde ω_i^C é definido utilizando um *kernel* K como

$$\omega_i^C = \frac{K(d(\mathbf{x}, \mathbf{x}_i))}{\sum_{i=1}^N K(d(\mathbf{x}, \mathbf{x}_i))}, \quad (4.11)$$

em que K é uma função de *kernel* gaussiana entre os centros dos intervalos e $d(\mathbf{x}, \mathbf{x}_i)$ é a distância Euclidiana entre \mathbf{x} e \mathbf{x}_i .

4.1.2.2 Modelo 2

No segundo modelo, são consideradas funções diferentes para os limites inferiores e superiores. A relação não linear é dada por

$$E(Y/\mathbf{X}) = \left[E(Y^C/\mathbf{X}^C) - \frac{1}{2}E(Y^R/\mathbf{X}^R), E(Y^C/\mathbf{X}^C) + \frac{1}{2}E(Y^R/\mathbf{X}^R) \right] \quad (4.12)$$

$$= \left[m^C(\mathbf{X}^C) - \frac{1}{2}m^R(\mathbf{X}^R), m^C(\mathbf{X}^C) + \frac{1}{2}m^R(\mathbf{X}^R) \right]. \quad (4.13)$$

Como no método anterior, as funções desconhecidas são baseadas no estimador proposto por Nadaraya-Watson, então

$$\hat{m}^C(\mathbf{x}^C) = \sum_{i=1}^N \omega_i^C y_i^C \quad (4.14)$$

$$\hat{m}^R(\mathbf{x}^R) = \sum_{i=1}^N \omega_i^R y_i^R \quad (4.15)$$

onde $\hat{m}^R(\mathbf{x}^R)$ é calculado de forma análoga ao $\hat{m}^C(\mathbf{x}^C)$.

4.2 Conjuntos de dados

Nas próximas subseções, são descritos os conjuntos de dados que foram utilizados no experimentos. São mostrados três tipos de conjuntos de dados intervalares: i) sintéticos em que os limites inferiores e superiores das saídas tem relação fraca com os dois limites de entrada; ii) os sintéticos em que os intervalos de saída estão diretamente relacionados com os intervalos de entrada; iii) e os conjunto de dados reais intervalares

4.2.1 Conjuntos sintéticos esquema 1

Esses conjuntos foram gerados de acordo com duas estratégias diferentes apresentadas em Neto e Carvalho (2017). Para cada uma, geram-se três variantes usando diferentes conjuntos de parâmetros. Cada um desses esquemas é descrito abaixo:

4.2.1.1 Conjunto de dados sintético 1 (DS1)

Nesta abordagem, usa-se o centro e raio de um conjunto de intervalos. O centro e o raio de cada variável de entrada foram gerados aleatoriamente seguindo as distribuições uniformes $\mathcal{U}(a, b)$ e $\mathcal{U}(c, d)$. As relações não lineares entre as entradas e saídas são apresentadas

nas Eqs. (4.16) e (4.17).

$$y_i^c = \frac{\theta_0^c}{\theta_1^c + e^{\theta_2^c x_i^c}} + \varepsilon_i^c \quad (4.16)$$

$$y_i^r = \theta_0^r + e^{-\theta_1^r x_i^r} + \varepsilon_i^r \quad (4.17)$$

Os parâmetros das Eqs. (4.16) e (4.17) são selecionados de acordo com $\theta_0^c \sim \mathcal{U}(1.9, 2.1)$, $\theta_1^c \sim \mathcal{U}(2.9, 3.1)$, $\theta_2^c \sim \mathcal{U}(0.9, 1.1)$, $\theta_0^r \sim \mathcal{U}(0.0, 0.5)$ e $\theta_1^r \sim \mathcal{U}(0.9, 1.1)$. Os resíduos são obtidos, para cada amostra, a partir de uma distribuição Gaussiana $\varepsilon^C \sim \mathcal{N}(0, \sigma^2)$.

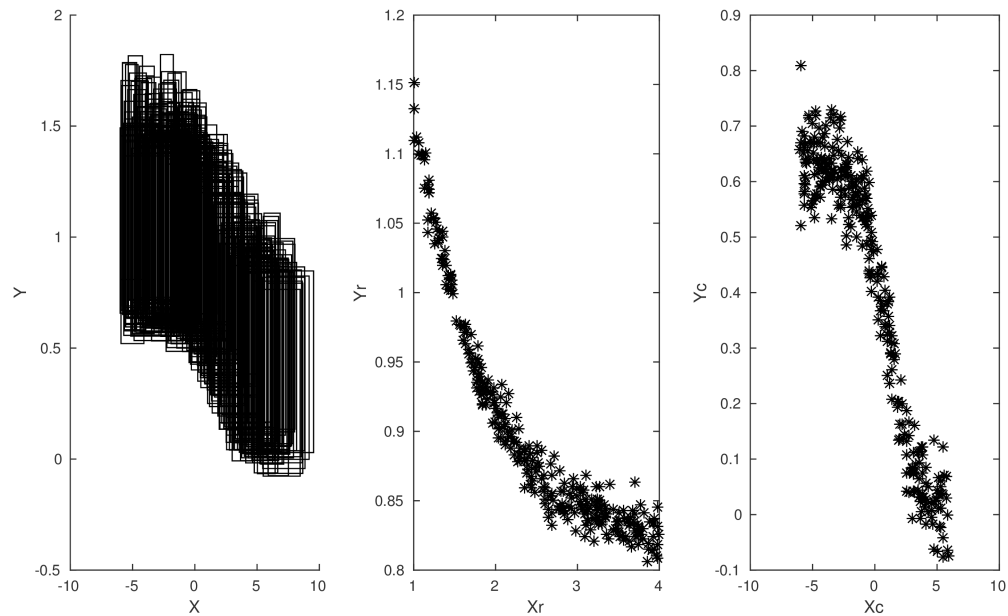
A Figura 2 mostra um exemplo de um conjunto DS1.

Figura 2 – Ilustração do conjunto sintético $DS1_1$

(a) Relação não linear

(b) Função Exponencial

(c) Função Logística



4.2.1.2 Conjuntos de dados sintéticos 2 (DS2)

A segunda estratégia tem uma relação não linear diferente entre entradas e saídas. Utiliza-se o modelo de Michaelis-Menten apresentado em Eq. (4.18) e o modelo Antoine visto em Eq. (4.19) para expressar as relações de centro e raio, respectivamente.

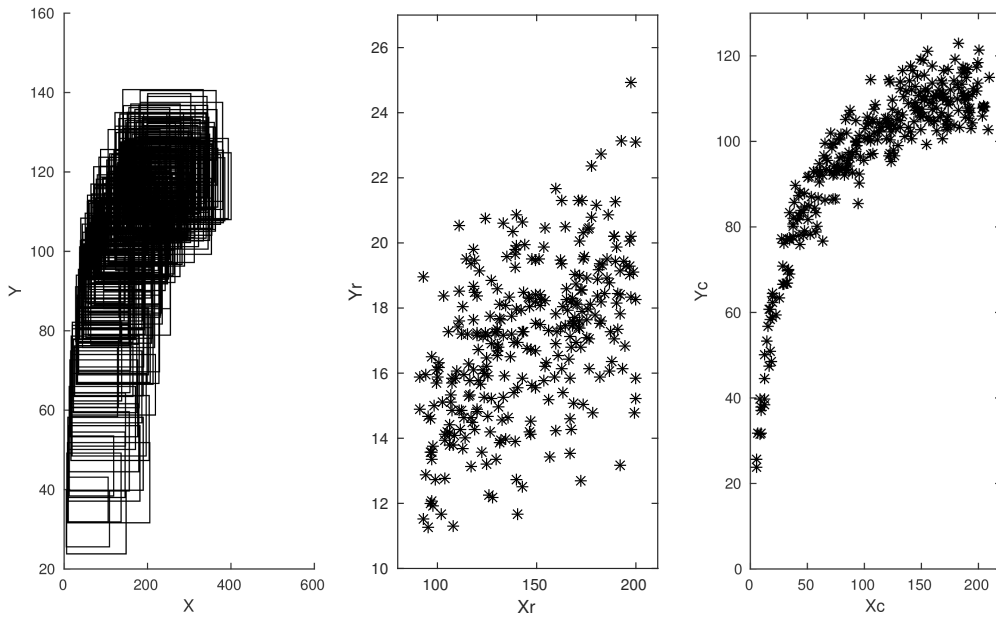
$$y_i^c = \frac{\theta_1^c x_i^c}{x_i^c \theta_0^c + \theta_1^c} + \varepsilon_i^c \quad (4.18)$$

$$\log(y_i^r) = \theta_0^r - \frac{\theta_1^r}{\theta_2^r + x_i^r} + \varepsilon_i^r \quad (4.19)$$

Os parâmetros das Eqs. (4.18) e (4.19) são selecionados de acordo com $\theta_0^C \sim \mathcal{U}(19, 21)$, $\theta_1^C \sim \mathcal{U}(117, 123)$, $\theta_0^R \sim \mathcal{U}(18, 20)$, $\theta_1^R \sim \mathcal{U}(490, 510)$ e $\theta_2^R \sim \mathcal{U}(-17, -15)$. Os resíduos são obtidos para cada amostra a partir de uma distribuição Gaussiana, $\varepsilon^C \sim \mathcal{N}(0, \sigma^2)$. A Figura 3 ilustra um conjunto de dados baseado na configuração DS2.

Figura 3 – Ilustração do conjunto sintético $DS2_2$

(a) Relação não linear (b) Modelo de Antoine (c) Modelo de MichaelisMenten



Considerando os dois esquemas mencionados acima, foram gerados seis conjuntos de dados (dois conjuntos de dados para o esquema DS1 e quatro para o esquema DS2), cada um com 300 amostras, escolhendo valores diferentes para os parâmetros a , b , c e d . A Tabela 1 mostra as configurações utilizadas nos experimentos.

Tabela 1 – Parâmetros utilizados para geração dos conjuntos de dados sintéticos

config	X^C	X^R	ε^C	ε^R
$DS1_1$	$\sim \mathcal{U}(-6, 6)$	$\sim \mathcal{U}(1, 4)$	$\sim \mathcal{N}(0, 0.05)$	$\sim \mathcal{N}(0, 0.01)$
$DS1_2$	$\sim \mathcal{U}(-6, 6)$	$\sim \mathcal{U}(1, 4)$	$\sim \mathcal{N}(0, 0.5)$	$\sim \mathcal{N}(0, 0.01)$
$DS2_1$	$\sim \mathcal{U}(5, 210)$	$\sim \mathcal{U}(90, 200)$	$\sim \mathcal{N}(0, 0.5)$	$\sim \mathcal{N}(0, 0.01)$
$DS2_2$	$\sim \mathcal{U}(5, 210)$	$\sim \mathcal{U}(90, 200)$	$\sim \mathcal{N}(0, 0.5)$	$\sim \mathcal{N}(0, 2)$
$DS2_3$	$\sim \mathcal{U}(5, 210)$	$\sim \mathcal{U}(90, 200)$	$\sim \mathcal{N}(0, 50)$	$\sim \mathcal{N}(0, 0.01)$
$DS2_4$	$\sim \mathcal{U}(5, 210)$	$\sim \mathcal{U}(90, 200)$	$\sim \mathcal{N}(0, 50)$	$\sim \mathcal{N}(0, 2)$

4.2.2 Conjuntos Sintéticos Esquema 2

Esses conjuntos foram gerados baseados na ideia que os intervalos de saídas podem estar diretamente relacionados com os limites superiores e inferiores das entradas, diferente do anterior. Aqui, os centros e raios dos intervalos de entrada foram selecionados através de distribuições uniformes ($\mathbf{X}^C \sim \mathcal{U}(a, b)$ e $\mathbf{X}^R \sim \mathcal{U}(c, d)$) e as saídas obtidas a partir de funções não lineares que dependem dos limites das entradas. Foram utilizadas 2 estratégias (DSR1 e DSR2) para obtenção das saídas:

1. $y^L = (\sum_{i=1}^p (x_i^L - x_i^U))^2$ e $y^U = (\sum_{i=1}^p (x_i^L + x_i^U))^2$
2. $y^L = \sqrt{\sum_{i=1}^p (x_i^U - x_i^L)}$ e $y^U = \sqrt{\sum_{i=1}^p (x_i^L + x_i^U)}$

Foram gerados 8 conjuntos, quatro de cada estratégia, em que quatro possuem dez dimensões e quatro duas dimensões. A Tabela 2 mostra as configurações utilizadas na criação dos conjuntos, as estratégias citadas acima utilizam essas mesmas configurações.

Tabela 2 – Parâmetros utilizados para geração do segundo esquema dos conjuntos de dados sintéticos

Config	\mathbf{X}^C	\mathbf{X}^R	Dimensão
DSR	$\sim \mathcal{U}(1, 10)$	$\sim \mathcal{U}(1, 3)$	10
DSR	$\sim \mathcal{U}(10, 100)$	$\sim \mathcal{U}(10, 30)$	2
DSR	$\sim \mathcal{U}(1, 10)$	$\sim \mathcal{U}(1, 3)$	2
DSR	$\sim \mathcal{U}(10, 100)$	$\sim \mathcal{U}(10, 30)$	10

4.2.3 Conjuntos Reais

Foi realizado experimentos com seis conjuntos de dados reais que podem ser encontrados em (FAGUNDES, 2013). Uma breve descrição de cada conjunto de dados é apresentada abaixo.

- **Cardiológico:** O conjunto de dados intervalar cardiológico tem cinquenta e nove pacientes que são descritos por três características. Duas variáveis independentes baseadas no intervalo, Pressão Arterial Sistólica e Pressão Arterial Diastólica são as entradas do problema e são usadas para prever a Frequência de Pulso.
- **Carro:** O conjunto de dados intervalar Carro tem trinta e três modelos de carros descritos por oito recursos. Nos testes, foi considerado apenas duas variáveis de entrada (velocidade máxima e a quantidade de cilindradas do motor) para prever o intervalo do preço.
- **Carro 7:** É o mesmo conjunto Carro, porém com mais atributos, como aceleração, passo,

comprimento, largura e altura.

- **Cogumelos:** O conjunto de dados intervalar Cogumelo consiste em vinte e três espécies que são descritas por três variáveis intervalares. As entradas são o comprimento do estipe e a espessura do estipe, e a saída é a largura da capa do píleo.
- **Futebol:** O conjunto Futebol fornece informações de jogadores profissionais de futebol de 20 equipes em França. Cada jogador é descrito por duas variáveis independentes, altura e idade, e uma variável dependente: peso.
- **NASA:** Um MDP (*Metrics Data Program*) é uma fundação que reúne as métricas de treze projetos de software e gera um repositório dessas informações. Cada projeto de software deve estar associado a trinta e três métricas de software. O processo de generalização transforma conjuntos de dados clássicos em intervalos pela execução da biblioteca ISDA que foi desenvolvida em (FAGUNDES, 2013). Finalmente, treze objetos intervalares foram gerados e três variáveis intervalares foram escolhidas. As duas variáveis independentes são o número de operandos e o número de operadores, que são usados para prever o Número de Linhas.

4.3 *Root Mean Square Error Bounds*

A avaliação de desempenho dos modelos de regressão é dada a partir da raiz do erro quadrático médio dos limites (*Root Mean Square Error Bounds* - RMSEB), que é obtido a partir das saídas originais e dos valores de saída estimados. O RMSEB é definido da seguinte forma:

$$\begin{aligned}
 RMSE^L &= \sqrt{\frac{1}{n}(\mathbf{y}^L - \hat{\mathbf{y}}^L)^\top (\mathbf{y}^L - \hat{\mathbf{y}}^L)} \\
 RMSE^U &= \sqrt{\frac{1}{n}(\mathbf{y}^U - \hat{\mathbf{y}}^U)^\top (\mathbf{y}^U - \hat{\mathbf{y}}^U)} \\
 RMSEB &= RMSE^L + RMSE^U
 \end{aligned} \tag{4.20}$$

5 CONFIGURAÇÕES DOS EXPERIMENTOS NA TAREFA DE CLASSIFICAÇÃO

As variantes do MLM foram validadas utilizando dois tipos de conjuntos, sintéticos (linearmente separáveis e não-linearmente separáveis) e reais. A fim de comparar o desempenho, foram implementados 3 abordagens da regressão logística vista em (SOUZA *et al.*, 2011). Para verificar a performance dos modelos, é utilizada a acurácia das classificações. Os métodos utilizados para comparação e conjuntos de dados são melhores explicados nas seções seguintes.

5.1 Métodos Classificação

Nesta seção, são apresentados superficialmente os modelos utilizados para comparação, sendo três variantes da regressão logística.

5.1.1 Regressão Logística para Dados Intervalares

Dado um conjunto intervalar \mathbf{X} , são construídos três modelos de regressão logística, a primeira utiliza os centros como entrada, a segunda concatena os limites inferiores e superiores e usa como entrada do problema, e a terceira combina dois modelos de regressão para classificar novas amostras. Estes modelos serão superficialmente explicados a seguir.

5.1.1.1 Modelo 1 (LOGIT-C) e 2 (LOGIT-LU)

Dado um conjunto dos centros de intervalos, \mathbf{X}^C , seja $\boldsymbol{\beta}_s$ um vetor de parâmetros desconhecidos de dimensão $p + 1$ para classe $s (s = 1 \dots S)$ e $\mathbf{x}^C = (1, x_1^C, \dots, x_p^C)$ um vetor de entrada. Para $s = 1, \dots, S$, o modelo logístico para os dados intervalares baseado em uma probabilidade a posteriori definida pelos centros dos intervalos é da seguinte forma:

$$\pi_1 = Pr(Y_k = 1/\mathbf{x}^C) = \frac{\exp(\boldsymbol{\beta}^\top \mathbf{x}^C)}{1 + \exp(\boldsymbol{\beta}^\top \mathbf{x}^C)} \quad (5.1)$$

podendo ser reescrita como

$$\log \frac{Pr(Y_k = 1/\mathbf{x}^C)}{Pr(Y_k = 0/\mathbf{x}^C)} = \boldsymbol{\beta}^\top \mathbf{x}^C \quad (5.2)$$

No segundo modelo, a entrada da regressão são os limites inferiores e superiores dos intervalos. Dado um conjunto $\mathbf{X}^{LU} = [\mathbf{X}^L \mathbf{X}^U]$, o modelo logístico é dado por

$$\pi_2 = Pr(Y_k = 1/\mathbf{x}^{LU}) = \frac{\exp(\boldsymbol{\beta}^\top \mathbf{x}^{LU})}{1 + \exp(\boldsymbol{\beta}^\top \mathbf{x}^{LU})} \quad (5.3)$$

podendo ser reescrita como

$$\log \frac{Pr(Y_k = 1/\mathbf{x}^{LU})}{Pr(Y_k = 0/\mathbf{x}^C)} = \boldsymbol{\beta}^\top \mathbf{x}^{LU} \quad (5.4)$$

Os vetores $\boldsymbol{\beta}$ podem ser estimados através do método do máximo *likelihood*. Uma nova amostra é classificada considerando a maior probabilidade a posteriori entre as classes.

5.1.1.2 Modelo 3 (LOGIT_L_U)

A terceira variante considera como entrada os limites inferiores e superiores, mas separadamente. Este método consiste em ajustar duas regressões logísticas para cada classe. Os resultados dessas regressões ajustadas (probabilidades a posteriori) são combinados para obter uma probabilidade posterior agrupada. Assim, de forma semelhante as anteriores, sendo \mathbf{X}^L e \mathbf{X}^U as matrizes contendo os limites inferiores e superiores das entradas, o modelo é construído da seguinte forma:

$$\frac{Pr(Y_k = 1/\mathbf{x}^L) + Pr(Y_k = 1/\mathbf{x}^U)}{2}, \quad (5.5)$$

onde

$$Pr(Y_k = 1/\mathbf{x}^L) = \frac{\exp(\boldsymbol{\beta}^\top \mathbf{x}^L)}{1 + \exp(\boldsymbol{\beta}^\top \mathbf{x}^L)} \quad (5.6)$$

$$Pr(Y_k = 1/\mathbf{x}^U) = \frac{\exp(\boldsymbol{\alpha}^\top \mathbf{x}^U)}{1 + \exp(\boldsymbol{\alpha}^\top \mathbf{x}^U)} \quad (5.7)$$

Novamente, os parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\alpha}$ podem ser estimados através do método de máximo *likelihood*.

5.2 Conjunto de Dados

Neste trabalho foram utilizados quinze conjuntos de dados sintéticos e cinco reais. Os conjuntos artificiais linearmente separáveis foram gerados a partir da abordagem vista em Souza *et al.* (2011), já os não linearmente separáveis foram obtidos a partir de Queiroz *et al.* (2018). Esses conjuntos de dados serão melhores explanados a seguir.

5.2.1 Conjuntos de Dados Sintéticos

Nos experimentos foram utilizados quatro modelos de dados, dois linearmente separáveis e dois não linearmente separáveis.

5.2.1.1 Conjunto Linearmente Separáveis (DSL)

Dois padrões de conjuntos de dados em \mathbb{R}^2 são considerados. São gerados 450 pontos dividido em três classes com tamanhos distintos: uma classe com um formato de elipse com 200 pontos e duas classes com formatos esféricos com 150 e 100 pontos, respectivamente. Cada classe foi gerada a partir de duas distribuições normais diferentes. Cada ponto conseguido, $z = (z_1, z_2)$, é transformado em um intervalo a partir de um raio γ selecionado aleatoriamente; assim, um ponto intervalar é gerado a partir de $([z_1 - \gamma_1, z_1 + \gamma_1], [z_2 - \gamma_2, z_2 + \gamma_2])$.

O primeiro padrão de dados foi gerado a partir das seguintes configurações vistas na Tabela 3:

Tabela 3 – Configurações conjunto de dados linearmente separável 1 (DSL1)

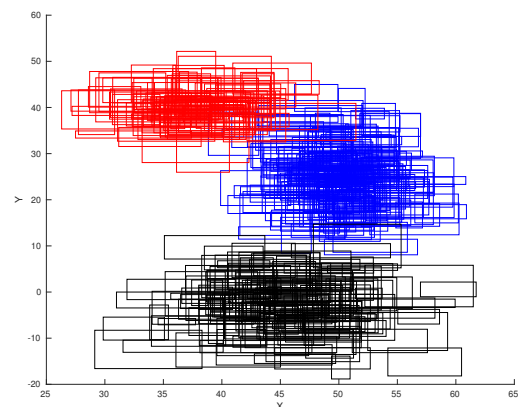
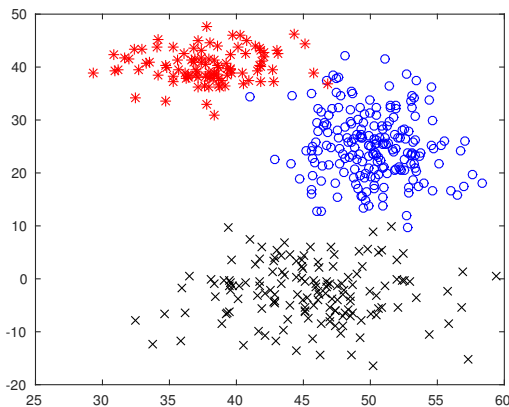
	μ_1	σ_1^2	μ_2	σ_2^2
Classe 1	50	9	25	36
Classe 2	45	25	-2	25
Classe 3	38	9	40	9

A Figura 4, mostra um exemplo do conjunto gerado com γ_1 e γ_2 aleatoriamente selecionados a partir de $\mathcal{U} \sim (1, 10)$. É notado que neste caso, não ocorre muita sobreposição das classes.

Figura 4 – Ilustração do conjunto sintético linear 1 com $\gamma_1, \gamma_2 = \mathcal{U} \sim (1, 10)$

(a) Relação entre X_1^C e X_2^C

(b) Relação entre X^C e X^R



O segundo padrão foi gerado a partir das configurações vista na Tabela 4 e a Figura 5 exemplifica um conjunto com com γ_1 e γ_2 randomicamente selecionados a partir de $\mathcal{U} \sim (1, 10)$. Nestes conjuntos, a sobreposição já é bem mais presente que no padrão 1.

Para os dois padrões de dados, os γ_1 e γ_2 foram selecionados aleatoriamente a partir de cinco distribuições uniformes: $\mathcal{U} \sim (1, 10)$, $\mathcal{U} \sim (1, 20)$, $\mathcal{U} \sim (1, 30)$, $\mathcal{U} \sim (1, 40)$ e $\mathcal{U} \sim (1, 50)$.

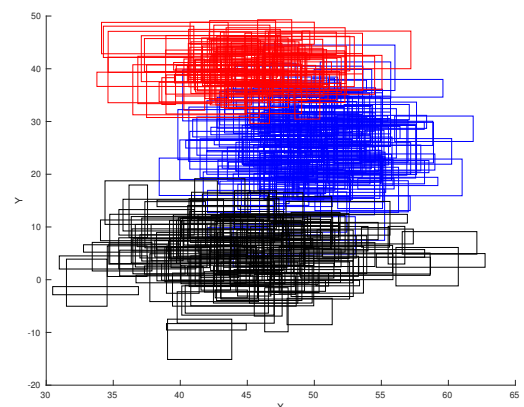
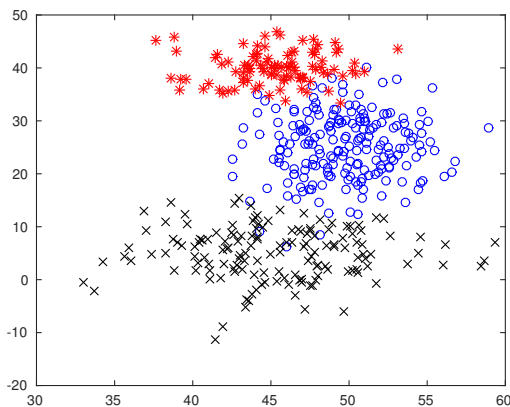
Tabela 4 – Configurações conjunto de dados linearmente separável 2 (DSL2)

	μ_1	σ_1^2	μ_2	σ_2^2
Classe 1	50	9	25	36
Classe 2	45	25	5	25
Classe 3	45	9	40	9

Figura 5 – Ilustração do conjunto sintético linear 2 com $\gamma_1, \gamma_2 = \mathcal{U} \sim (1, 10)$

(a) Relação entre X_1^C e X_2^C

(b) Relação entre X^C e X^R



5.2.1.2 Conjuntos Não-Linearmente Separáveis (DSNL)

Para obter esses conjuntos, é utilizada uma ideia semelhante à descrita acima: é gerado um conjunto de dados clássicos e criado raios a partir de γ . Dois padrões são usados para construir-lós.

O primeiro conjunto de dados tem 100 pontos distribuídos em duas classes. Cada classe é definida como uma semicircunferência gerada a partir de dados com a mesma distribuição

uniforme. Este conjunto é elaborado a partir de

$$\text{Classe 1 } X_1 \sim \mathcal{U}(5, 25)$$

$$X_2 = \sqrt{100 - (X_1 - 15)^2} + 20$$

$$e_1 \sim \mathcal{N}(0, 1)$$

$$S_{X_1} = 10$$

$$S_{X_2} = -3$$

$$\text{Classe 2 } X_1 \sim \mathcal{U}(5, 25)$$

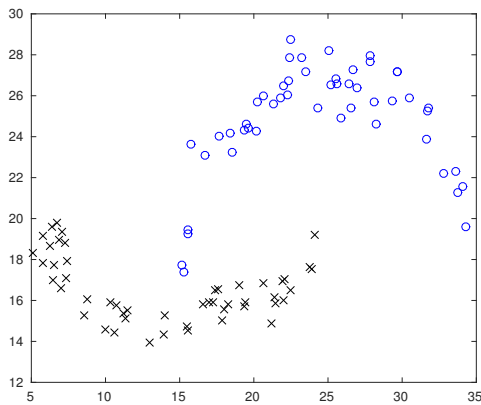
$$X_2 = \sqrt{100 - (X_1 - 15)^2} + 5$$

$$e_2 \sim \mathcal{N}(0, 1)$$

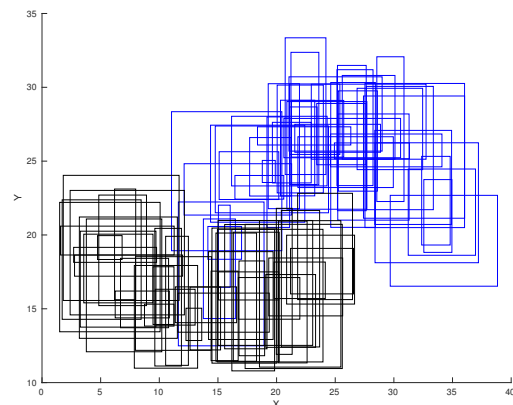
onde X_1 é a primeira coordenada, X_2 é a segunda coordenada dada pela equação da semicircunferência, S_{X_1} e S_{X_2} são constantes adicionadas para forçar a aproximação da Classe 1 com a Classe 2, e e_1 e e_2 são ruídos adicionados às coordenadas X_2 . A Figura 6 mostra um exemplo do conjunto com γ_1 e γ_2 randomicamente selecionados a partir de $\mathcal{U} \sim (1, 10)$.

Figura 6 – Ilustração do conjunto sintético não linearmente separável 1 (DSL N1) com $\gamma_1, \gamma_2 = \mathcal{U} \sim (1, 10)$

(a) Relação entre \mathbf{X}_1^C e \mathbf{X}_2^C



(b) Relação entre X^C e X^R



O segundo conjunto de dados tem 150 pontos distribuídos em duas classes de forma distinta, a primeira tem 100 pontos e a segunda 50. Ambas as classes foram desenvolvidas como circunferências com a mesma origem, mas com raios diferentes; as Classe 1 e 2 são construídas

a partir das Equações:

$$(X_1 - 20)^2 + (X_2 - 20)^2 = 20^2$$

$$(X_1 - 20)^2 + (X_2 - 20)^2 = 5^2$$

Assim, o segundo modelo é gerado por

$$\text{Classe 1 } X_1 \sim \mathcal{U}(0, 40)$$

$$X_2' = \frac{40 + \sqrt{-4X_1^2 + 160X_1}}{2}$$

$$X_2'' = \frac{40 - \sqrt{-4X_1^2 + 160X_1}}{2}$$

$$e_1 \sim \mathcal{N}(0, 1)$$

$$\text{Classe 2 } X_1 \sim \mathcal{U}(15, 25)$$

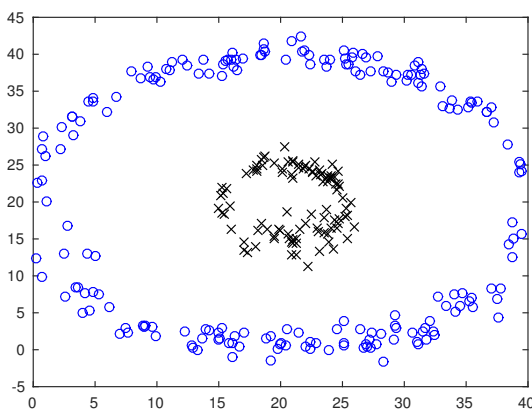
$$X_2' = \frac{40 + \sqrt{-4X_1^2 + 160X_1 + 1500}}{2}$$

$$X_2'' = \frac{40 - \sqrt{-4X_1^2 + 160X_1 + 1500}}{2}$$

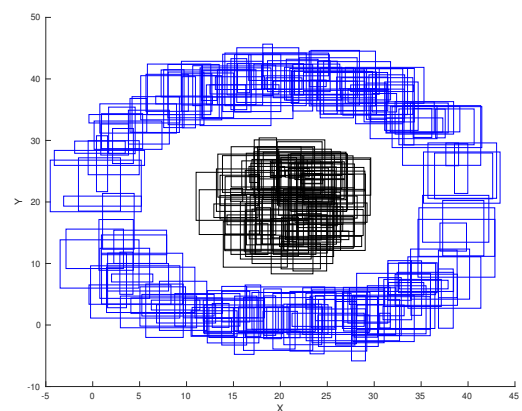
A Figura 7 mostra um exemplo do conjunto com γ_1 e γ_2 randomicamente selecionados a partir de $\mathcal{U} \sim (1, 10)$.

Figura 7 – Ilustração do conjunto sintético não linearmente separável 2 (DSNL2) com $\gamma_1, \gamma_2 = \mathcal{U} \sim (1, 10)$

(a) Relação entre \mathbf{X}_1^C e \mathbf{X}_2^C



(b) Relação entre X^C e X^R



Nesta abordagem, γ_1 e γ_2 foram selecionados randomicamente a partir de três distribuições uniformes: $\mathcal{U} \sim (1, 5)$, $\mathcal{U} \sim (1, 10)$ e $\mathcal{U} \sim (1, 15)$.

5.2.2 Conjuntos de dados reais

Foram utilizados três conjuntos de dados reais, o carro e cogumelo (os mesmos do problema de regressão, porém adaptados para classificação) e uma conjunto com informações das temperaturas de cidades no decorrer de doze meses. Esses conjuntos são descritos a seguir:

- **Carro:** consiste em 33 modelos de carros descritos por 8 variáveis intervalares (Preço, Cilindrada, Velocidade Máxima, Aceleração, Passo, Comprimento, Largura e Altura), a variável que determina a classe é a categoria do carro. Foi utilizada duas abordagens, uma considerando todas as quatro categorias (Utilitário, Berlina, Esporte e Luxo) e outra agrupando Utilitário e Berlina, e Esporte e Luxo, ficando com duas classes.
- **Cidades:** Este conjunto de dados intervalar (CARVALHO; SOUZA, 2010) diz respeito a 37 cidades. Cada cidade é descrita por 12 variáveis intervalares que são as temperaturas máximas e mínimas de 12 meses em graus centígrados. São quatro classes, que não são balanceadas; a primeira tem quinze amostras, a segunda tem vinte, a terceira uma e a quarta uma.

6 RESULTADOS

6.1 Resultados Experimentos de Regressão

Nestes experimentos, todos os hiperparâmetros de cada modelo foram definidos usando validação cruzada por *10-fold*. Em conjuntos de dados sintéticos, foram calculados os RMSEBs para todos os modelos de regressão usando simulação de Monte Carlo com 30 execuções independentes. Em conjuntos de dados reais, foi utilizado uma abordagem de validação cruzada *leave-one-out*, pela pouca quantidade de amostras dos conjuntos. Os resultados obtidos para os conjuntos de dados sintéticos com relação fraca é mostrado na Tabela 5, os resultados dos conjuntos de dados diretamente relacionados são vistos nas Tabelas 6 e 7, e os RMSEB obtidos nos conjuntos reais estão na Tabela 8.

Tabela 5 – RMSEB obtidos nos conjuntos de dados sintéticos com relação fraca

	$DS1_1$	$DS1_2$	$DS2_1$	$DS2_2$	$DS2_3$	$DS2_4$
MLM-CR	0.256 ± 0.017	0.802 ± 0.044	0.952 ± 0.244	1.102 ± 0.302	73.77 ± 5.041	67.15 ± 2.86
MLM-LU	0.269 ± 0.018	0.814 ± 0.042	18.06 ± 2.075	15.71 ± 2.039	76.35 ± 6.327	72.92 ± 3.444
MLM-CRT	0.996 ± 0.063	1.429 ± 0.088	162.0 ± 3.07	161.1 ± 3.090	162.0 ± 12.89	169.3 ± 13.37
MLM-LUT	0.263 ± 0.019	0.877 ± 0.036	1.967 ± 0.633	1.697 ± 0.443	83.59 ± 5.982	77.13 ± 3.992
ELM 1	0.553 ± 0.027	0.953 ± 0.047	78.44 ± 2.234	77.18 ± 2.206	103.3 ± 5.601	100.8 ± 4.897
ELM 2	1.185 ± 0.029	1.432 ± 0.050	21.20 ± 2.972	17.29 ± 2.477	73.54 ± 5.751	70.27 ± 3.365
ELM 3	1.189 ± 0.030	1.443 ± 0.054	20.31 ± 2.592	17.30 ± 2.467	73.27 ± 5.737	69.93 ± 3.468
RKCI	0.289 ± 0.016	0.783 ± 0.037	2.329 ± 0.599	2.233 ± 0.354	72.40 ± 6.351	67.27 ± 3.873
RKCIA	0.275 ± 0.014	0.776 ± 0.037	1.353 ± 0.701	1.296 ± 0.462	72.39 ± 6.347	67.25 ± 3.886

Tabela 6 – RMSEB obtidos na primeira configuração dos dados sintéticos com relação direta

	$DSR1_1$	$DSR1_2$	$DSR1_3$	$DSR1_4$
MLM-CR	2904.90 ± 189.20	24572.33 ± 1444.22	247.68 ± 17.16	257408.15 ± 20212.48
MLM-LU	1708.14 ± 147.27	7831.70 ± 829.92	144.77 ± 11.10	94420.46 ± 6734.53
MLM-CRT	3694.45 ± 173.98	14404.96 ± 531.30	174.32 ± 9.66	341919.72 ± 15672.19
MLM-LUT	1509.55 ± 118.19	2093.87 ± 475.65	103.70 ± 8.97	101801.30 ± 7583.10
ELM 1	2865.90 ± 205.21	24861.15 ± 1604.85	251.32 ± 14.32	257516.99 ± 20242.90
ELM 2	2146.86 ± 171.42	12057.08 ± 1214.04	187.98 ± 15.00	178640.60 ± 18269.02
ELM 3	2156.68 ± 155.31	3907721.4 ± 21332158.21	197.54 ± 18.51	180177.16 ± 18591.89
RKCI	2685.38 ± 239.84	3338.65 ± 215.71	173.48 ± 14.57	176712.44 ± 17837.55
RKCIA	3296.49 ± 246.37	24092.14 ± 1553.08	260.92 ± 16.49	289911.35 ± 23140.71

Como pode ser notado, as variantes propostas de MLM alcançaram os melhores resultados em quatro dos seis conjuntos de dados sintéticos com relação fraca. Nos conjuntos de dados com relação mais forte, o MLM-LUT foi melhor em quase todas as oito variações, o que poderia ser esperado já que ele é o único método que relaciona diretamente os limites inferiores e superiores de entrada com os de saídas. Nos conjuntos reais, as extensões do MLM

Tabela 7 – RMSEB obtidos na segunda configuração dos dados sintéticos com relação direta

	$DSR2_1$	$DSR2_2$	$DSR2_3$	$DSR2_4$
MLM-CR	2.794 ± 0.112	3.342 ± 0.178	1.750 ± 0.096	7.079 ± 0.365
MLM-LU	2.185 ± 0.089	2.296 ± 0.121	1.622 ± 0.080	4.184 ± 0.304
MLM-CRT	39.44 ± 0.187	24.95 ± 0.263	7.947 ± 0.173	125.65 ± 0.403
MLM-LUT	2.085 ± 0.114	1.648 ± 0.111	1.598 ± 0.066	2.519 ± 0.188
ELM 1	5.006 ± 0.359	4.764 ± 0.400	2.325 ± 0.170	12.501 ± 1.042
ELM 2	3.409 ± 0.205	3.597 ± 0.543	1.767 ± 0.094	6.710 ± 0.645
ELM 3	3.396 ± 0.197	3.558 ± 0.302	1.785 ± 0.102	6.665 ± 0.641
RKCI	2.704 ± 0.127	2.162 ± 0.121	1.583 ± 0.078	6.065 ± 0.436
RKCIA	3.040 ± 0.150	3.262 ± 0.136	1.710 ± 0.088	8.180 ± 0.502

Tabela 8 – RMSEB obtidos nos conjuntos reais

	Cardiologia	Carro	Cogumelo	Carro 7	Futebol	NASA
MLM-CR	18.529 ± 10.95	43652.10 ± 41409.59	2.811 ± 1.609	42276.12 ± 42890.15	4.231 ± 3.036	1650.95 ± 1317.55
MLM-LU	17.92 ± 13.04	41972.42 ± 47358.27	3.139 ± 1.437	43588.88 ± 45915.99	4.655 ± 2.913	1360.83 ± 1611.13
MLM-CRT	20.65 ± 10.84	47146.43 ± 49216.49	3.399 ± 1.950	37478.16 ± 36024.51	4.743 ± 2.441	863.248 ± 702.125
MLM-LUT	18.49 ± 10.60	42559.24 ± 45657.13	4.046 ± 1.775	42713.90 ± 47084.59	4.836 ± 3.688	953.369 ± 987.399
ELM 1	17.77 ± 11.82	87138.48 ± 92683.12	6.206 ± 17.53	154179.81 ± 109222.00	3.593 ± 2.318	966.071 ± 1472.901
ELM 2	22.22 ± 34.42	64481.83 ± 48739.61	4.947 ± 5.967	1804312.74 ± 843200.54	4.024 ± 3.192	945.845 ± 914.384
ELM 3	51.00 ± 245.24	1170267.2 ± 587480	2.873 ± 1.914	465659.67 ± 1508449.54	3.965 ± 2.139	7550.61 ± 22085.91
RKCI	16.124 ± 9.795	46264.38 ± 52219.28	3.673 ± 2.003	57831.53 ± 72304.61	3.680 ± 2.492	887.889 ± 1007.2
RKCIA	16.368 ± 9.422	51467.75 ± 53694.8	3.368 ± 1.993	166112.28 ± 76765.65	3.317 ± 2.272	855.481 ± 1007.70

tiveram os menores erros em três dos seis. Mais especificamente, pode-se notar que o MLM-CR alcançou bons resultados, ficando na frente da maioria nos dados sintéticos sem relação e obtendo resultados bem próximos dos melhores (quando não obteve o melhor), nos conjuntos reais. Vale ressaltar que se existir uma relação entre os limites inferiores e superiores de entrada e saída, o MLM-LUT seria a melhor opção. É importante notar também que nos demais conjuntos de dados, os RMSEBs dos modelos MLM foram similares aos melhores resultados obtidos pelos outros métodos.

6.2 Resultados Experimentos de Classificação

Nesta seção são apresentadas as acurácias obtidas com as extensões do MLM e regressão logística nos conjuntos de dados sintéticos e reais. A Tabela 9 mostra os resultados obtidos nos conjuntos de dados sintéticos (tanto linearmente separáveis, quanto não linearmente separável). A regressão logística que utiliza apenas os centro como entrada não foi considerada nestes experimentos, pois a adição dos raios não a afeta. Em todos os conjuntos, as variantes do MLM conseguiram acurácias melhores que as da regressão logística. É notado que o aumento do raio dos intervalos interferiu pouco nos modelos MLM, diferente dos modelos logísticos, em que o aumento do raio prejudicou o desempenho, principalmente nos dados linearmente separáveis.

Além disso, pelos resultados, vale ressaltar que os modelos MLM obtiveram as taxas de acerto bem próximas.

Tabela 9 – Acurácia dos métodos nos Dados Sintéticos

	MLM-CRT	MLM-LUT	LOGIT-LU	LOGIT_L_U
DSL1-1-10	0.9888 ± 0.010	0.9874 ± 0.010	0.8937 ± 0.096	0.9700 ± 0.021
DSL1-1-20	0.9933 ± 0.007	0.9929 ± 0.007	0.8822 ± 0.121	0.9596 ± 0.030
DSL1-1-30	0.9896 ± 0.006	0.9892 ± 0.007	0.8851 ± 0.112	0.8792 ± 0.070
DSL1-1-40	0.9811 ± 0.016	0.9796 ± 0.015	0.8959 ± 0.119	0.8611 ± 0.086
DSL1-1-50	0.9866 ± 0.010	0.9874 ± 0.009	0.9003 ± 0.124	0.7903 ± 0.100
DSL2-1-10	0.9655 ± 0.016	0.9670 ± 0.014	0.8774 ± 0.110	0.9062 ± 0.054
DSL2-1-20	0.9522 ± 0.019	0.9529 ± 0.022	0.8251 ± 0.094	0.8318 ± 0.108
DSL2-1-30	0.9600 ± 0.017	0.9588 ± 0.015	0.7737 ± 0.152	0.7851 ± 0.122
DSL2-1-40	0.9496 ± 0.023	0.9474 ± 0.022	0.8466 ± 0.136	0.701 ± 0.097
DSL2-1-50	0.9444 ± 0.020	0.9418 ± 0.020	0.8137 ± 0.135	0.6355 ± 0.106
DSNL1-1-5	1.00 ± 0.0	1.00 ± 0.0	0.7550 ± 0.111	0.5883 ± 0.120
DSNL1-1-10	0.9900 ± 0.020	0.9900 ± 0.020	0.6883 ± 0.119	0.5850 ± 0.131
DSNL1-1-15	0.9983 ± 0.009	0.9983 ± 0.009	0.6050 ± 0.129	0.5266 ± 0.123
DSNL2-1-5	1.00 ± 0.0	1.00 ± 0.0	0.5605 ± 0.121	0.6350 ± 0.088
DSNL2-1-10	1.00 ± 0.0	1.00 ± 0.0	0.5616 ± 0.121	0.5416 ± 0.126
DSNL2-1-15	1.00 ± 0.0	1.00 ± 0.0	0.6038 ± 0.114	0.5900 ± 0.1331

A Tabela 10 mostra os resultados conseguidos nos dados reais, novamente os modelos propostos obtiveram acurácias melhores, ou similares, aos modelos logísticos. Para verificar a vantagem de utilizar dados intervalares ao invés de adaptá-los utilizando apenas os centros, o MLM-C e a LOGIT-C foram consideradas nestes experimentos. Na maioria dos casos, as abordagens com os intervalos tiveram ganho sobre as abordagens usando apenas o centro. Aqui, nota-se que o MLM-CRT se destacou em três dos cinco conjuntos. A partir destes resultados, percebe-se que as variantes do MLM conseguem taxas melhores ou próximas as variantes da regressão logística, se tornando uma opção preferível.

Tabela 10 – Acurácia obtida nos dados reais

	MLM-C	MLM-CRT	MLM-LUT	LOGIT-C	LOGIT-LU	LOGIT_L_U
Carro 2 Classes	0.8181 ± 0.391	0.8484 ± 0.364	0.7878 ± 0.415	0.8181 ± 0.391	0.7575 ± 0.435	0.8484 ± 0.364
Carro 4 classes	0.6666 ± 0.478	0.6060 ± 0.496	0.5757 ± 0.501	0.4848 ± 0.488	0.5454 ± 0.505	0.3636 ± 0.488
Cidades	0.9189 ± 0.276	0.9459 ± 0.229	0.9459 ± 0.229	0.7567 ± 0.434	0.7567 ± 0.434	0.7837 ± 0.417
Cogumelo 2 Classes	0.6521 ± 0.486	0.6521 ± 0.486	0.6956 ± 0.470	0.6486 ± 0.483	0.7297 ± 0.450	0.7027 ± 0.463
Cogumelo 3 Classes	0.6521 ± 0.486	0.7826 ± 0.421	0.7391 ± 0.448	0.6756 ± 0.474	0.7027 ± 0.463	0.7567 ± 0.434

7 CONCLUSÕES E TRABALHOS FUTUROS

Algoritmos de Aprendizagem de Máquina geralmente são desenvolvidos para trabalhar com conjunto de dados clássicos onde cada amostra é representada como um único ponto. Porém em muitas situações esses dados podem ser intervalos e utilizar a forma clássica pode não representar bem o problema, causando perda de informação ou dados que não são consistentes.

Este trabalho propõe quatro modelos baseados em MLM para conjuntos de dados intervalares. Os modelos propostos, denominados MLM-CR e MLM-LU, baseiam-se na ideia de construir dois modelos de MLM independentes para aproximar as saídas definidas por intervalos (limites inferior-superior e centro-raio). Os métodos MLM-CRT e MLM-LUT partem da ideia dos limites dos intervalos de entrada serem diretamente relacionados com os limites do intervalos de saída. As propostas também garantem que as restrições necessárias para construir intervalos viáveis sejam satisfeitas, ou seja, todos os intervalos de saída respeitam as restrições: $y^L \leq y^U$ e $y^R \geq 0$.

A comparação de desempenho entre as propostas e outros métodos de regressão não-linear mostrou que as variantes de MLM podem superar os métodos previamente propostos na maioria dos conjuntos de dados testados. Este fato mostra que o MLM-CR, MLM-LU, MLM-CRT e o MLM-LUT podem ser considerados como uma alternativa válida para problemas de regressão intervalar. Ainda em regressão, uma diferença crucial entre os MLM-CRT e MLM-LUT para os métodos já existentes na literatura está na forma de tratar os intervalos; geralmente eles são considerados independentes. Nas propostas os intervalos são considerados dependentes, assim o problema não é decomposto e, dependendo do relacionamento dos dados, podem ter performance melhor como visto nos experimentos.

Os resultados mostram que na tarefa de classificação, os modelos baseados no MLM tiveram acurácias superiores em quase todos os conjuntos de dados quando comparado com os modelos logísticos, assim, mostrando-se uma opção viável. Além disso, na maioria dos casos, os modelos que utilizavam os intervalos como entrada (tanto MLM quanto regressão logística) obtiveram ganhos sobre os modelos que utilizavam apenas o centro, o que reforça a importância de estudos na área de dados intervalares.

REFERÊNCIAS

- ALENCAR, A. S. de C. **Máquina de aprendizagem mínima: aspectos teóricos e práticos**. Dissertação (Mestrado) — Universidade Federal do Ceará, 2017.
- BILLARD, L. Symbolic data analysis: what is it? In: **Compstat 2006-Proceedings in Computational Statistics**. [S. l.]: Springer, 2006. p. 261–269.
- BILLARD, L.; DIDAY, E. Regression analysis for interval-valued data. In: **Data Analysis, Classification, and Related Methods**. [S. l.]: Springer, 2000. p. 369–374.
- CARVALHO, F. d. A. de; SOUZA, R. M. de. Unsupervised pattern recognition models for mixed feature-type symbolic data. **Pattern Recognition Letters**, Elsevier, v. 31, n. 5, p. 430–443, 2010.
- CARVALHO, F. de A. T. de; SIMÕES, E. C. Fuzzy clustering of interval-valued data with city-block and hausdorff distances. **Neurocomputing**, v. 266, p. 659–673, 2017.
- CARVALHO, F. de A. T. de; TENORIO, C. P. Fuzzy k-means clustering algorithms for interval-valued data based on adaptive quadratic distances. **Fuzzy Sets and Systems**, v. 161, n. 23, p. 2978–2999, 2010.
- COELHO, D. N.; BARRETO, G. D. A.; MEDEIROS, C. M. S.; SANTOS, J. D. A. Performance comparison of classifiers in the detection of short circuit incipient fault in a three-phase induction motor. In: **2014 IEEE Symposium on Computational Intelligence for Engineering Solutions, CIES 2014, Orlando, FL, USA, December 9-12, 2014**. [S. l.: s. n.], 2014. p. 42–48.
- DENOEU, T.; MASSON, M. Multidimensional scaling of interval-valued dissimilarity data. **Pattern Recognition Letters**, v. 21, n. 1, p. 83–92, 2000.
- DIAS, M. L.; SOUSA, L. S. de; NETO, A. R. da R.; JÚNIOR, A. H. de S. Opposite neighborhood: a new method to select reference points of minimal learning machines. In: **26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning-ESANN 2018**. [S. l.: s. n.], 2018. p. 201–206.
- DOUZAL-CHOUAKRIA, A.; BILLARD, L.; DIDAY, E. Principal component analysis for interval-valued observations. **Statistical Analysis and Data Mining: The ASA Data Science Journal**, Wiley Online Library, v. 4, n. 2, p. 229–246, 2011.
- FAGUNDES, R. A.; SOUZA, R. M.; CYSNEIROS, F. J. A. Interval kernel regression. **Neurocomputing**, Elsevier, v. 128, p. 371–388, 2014.
- FAGUNDES, R. A. A. **Métodos de regressão robusta e kernel para dados intervalares**. Tese (Doutorado) — Universidade Federal de Pernambuco, 2013.
- GROENEN, P. J. F.; WINSBERG, S.; RODRÍGUEZ, O.; DIDAY, E. I-scal: Multidimensional scaling of interval dissimilarities. **Computational Statistics & Data Analysis**, v. 51, n. 1, p. 360–378, 2006.
- KOWALSKI, P. A.; KULCZYCKI, P. Interval probabilistic neural network. **Neural Computing and Applications**, Springer, v. 28, n. 4, p. 817–834, 2017.

- LAURO, C. N.; PALUMBO, F. Principal component analysis of interval data: a symbolic data analysis approach. **Computational statistics**, Citeseer, v. 15, n. 1, p. 73–87, 2000.
- LI, F.; LI, S.; TANG, N.; DENŒUX, T. Constrained interval-valued linear regression model. In: IEEE. **2017 20th International Conference on Information Fusion (Fusion)**. [S. l.], 2017. p. 1–8.
- MARINHO, L. B.; ALMEIDA, J. S.; SOUZA, J. W. M.; ALBUQUERQUE, V. H. C. de; FILHO, P. P. R. A novel mobile robot localization approach based on topological maps using classification with reject option in omnidirectional images. **Expert Syst. Appl.**, v. 72, p. 1–17, 2017.
- MESQUITA, D. P.; GOMES, J. P.; JUNIOR, A. H. S. Ensemble of efficient minimal learning machines for classification and regression. **Neural Processing Letters**, Springer, v. 46, n. 3, p. 751–766, 2017.
- MESQUITA, D. P. P.; GOMES, J. P. P.; Souza Junior, A. H. Ensemble of Efficient Minimal Learning Machines for Classification and Regression. **Neural Processing Letters**, Springer US, 2017. ISSN 1370-4621.
- NETO, E. A. L.; CARVALHO, F. A. Constrained linear regression models for symbolic interval-valued variables. **Computational Statistics & Data Analysis**, Elsevier, v. 54, n. 2, p. 333–347, 2010.
- NETO, E. A. L.; CARVALHO, F. A. Nonlinear regression applied to interval-valued data. **Pattern Analysis and Applications**, Springer, v. 20, n. 3, p. 809–824, 2017.
- NETO, E. d. A. L.; CARVALHO, F. d. A. de. Constrained linear regression models for symbolic interval-valued variables. **Computational Statistics & Data Analysis**, Elsevier, v. 54, n. 2, p. 333–347, 2010.
- QUEIROZ, D.; SOUZA, R.; CYSNEIROS, F. d. A.; ARAUJO, M. Kernelized inner product-based discriminant analysis for interval data. **Pattern Analysis and Applications**, Springer, v. 21, n. 3, p. 731–740, 2018.
- ROBERTA, A. d. A.; FILHO, R. J. Q.; SOUZA, R. M. de; CYSNEIROS, F. J. A. An interval nonparametric regression method. In: IEEE. **The 2013 International Joint Conference on Neural Networks (IJCNN)**. [S. l.], 2013. p. 1–7.
- ROQUE, A. M. S.; MATÉ, C.; ARROYO, J.; SARABIA, Á. implp: Applying multi-layer perceptrons to interval-valued data. **Neural Processing Letters**, Springer, v. 25, n. 2, p. 157–169, 2007.
- Souza Junior, A. H.; CORONA, F.; MICHE, Y.; LENDASSE, A.; BARRETO, G. A.; SIMULA, O. Minimal Learning Machine: A New Distance-Based Method for Supervised Learning. In: _____. **Advances in Computational Intelligence: 12th International Work-Conference on Artificial Neural Networks, IWANN 2013, Puerto de la Cruz, Tenerife, Spain, June 12-14, 2013, Proceedings, Part I**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. p. 408–416.
- SOUZA, R. M. de; QUEIROZ, D. C.; CYSNEIROS, F. J. A. Logistic regression-based pattern classifiers for symbolic interval data. **Pattern Analysis and Applications**, Springer, v. 14, n. 3, p. 273, 2011.

WANG, X.; YU, F.; PEDRYCZ, W.; YU, L. Clustering of interval-valued time series of unequal length based on improved dynamic time warping. **Expert Systems with Applications**, Elsevier, 2019.

YANG, Z.; LIN, D. K.; ZHANG, A. Interval-valued data prediction via regularized artificial neural network. **Neurocomputing**, Elsevier, v. 331, p. 336–345, 2019.

ZHAO, S.; WANG, X. Extreme learning machine for interval-valued data. In: **SPRINGER. International Conference on Machine Learning and Cybernetics**. [S. l.], 2014. p. 388–399.