UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**LUIZ ALBERTO QUEIROZ CORDOVIL JÚNIOR**
**RODRIGO FARIAS ARAÚJO**

# SISTEMAS NEBULOSOS: EXERCÍCIO COMPUTACIONAL 3

**Belo Horizonte - MG, 2017**

**LUIZ ALBERTO QUEIROZ CORDOVIL JÚNIOR**
**RODRIGO FARIAS ARAÚJO**

# APLICAÇÃO DO ALGORITMO C-MEANS PARA SEGMENTAÇÃO DE IMAGENS

Relatório apresentado como requisito parcial para obtenção de aprovação na disciplina Sistemas Nebulosos do Programa de Pós-Graduação em Engenharia Elétrica, na Universidade Federal de Minas Gerais.

Prof. Dr. André Paim Lemos

**Belo Horizonte - MG, 2017**

# Sumário

# Lista de Figuras

## Resumo

Um algoritmo de clusterização organiza itens em grupos baseado no critério de similaridade. O algoritmo *fuzzy c-Means* realiza agrupamentos em que determinado item pode pertencer a mais de um grupo afim, em que o grau de pertinência para cada item é dado por uma distribuição de probabilidade sobre os clusters. Neste trabalho é apresentada aplicação deste algoritmo de agrupamento para a tarefa de segmentação de imagens e inferência de dados a partir de aprendizado não-supervisionado e raciocínio aproximado.

# 1 Introdução

*Fuzzy c-means*, é uma algoritmo para análise de agrupamentos associado à representação de uma classe ou grupo singular em um dado conjunto de dados. $c$ agrupamentos são representados por um vetor de centros *C*.

Para que se alcance este objetivo ou *clusterização*, realiza-se um número de iterações no sentido de minimizar $J_cM(\mu_h, C)$ dada por:

$$J = \sum_{i=1}^{N}\sum_{j=1}^{C} \mu_{ij}\|x_i - c_j\|^2 \tag{1.1}$$

onde:

- $N \longrightarrow$ número de dados;
- $\mathcal{C} \longrightarrow$ número de centros (*clusters*);
- $c_j$ vetor de centro do *j*;
- $\mu_{ij} \longrightarrow$ grau de pertinência do *i*-ésimo dado $x_i$ no *cluster* $j$.

A norma $\|x_i - c_j\|$ mede a similaridade (ou proximidade) de um dado $x_i$ para vetor de centro $c_j$ do *j*. Em cada iteração, o algoritmo mantém o vetor centro para cada *cluster*.

Para cada ponto $x_i$, o grau de pertinência para cada $c_j$ é calculado da seguinte forma:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C}(\frac{\|x_i - c_j\|}{\|x_i - c_k\|})^{\frac{2}{m-1}}} \tag{1.2}$$

onde *m* é coeficiente de fuzzificação.

O vetor centro $c_j$ é calculado como:

$$c_j = \frac{\sum_{i=1}^{N} \mu_{ij}^m x_i}{\sum_{1}^{N} \mu_{ij}^m} \tag{1.3}$$

No início do algoritmo, o grau de pertinência para os dados é inicializado com um valor randômico $\theta_{ij} \in [0, 1] \mid \sum_j^C \mu_{ij} = 1$.

O coeficiente de *fuzzificação* $m \in [1, \infty]$, mede a tolerância do *cluster*. Este valor determina o quanto um *cluster* pode sobrepor outro. Quanto maior este valor, maior a sobreposição entre os agrupamentos, ao passo que também se usa uma maior quantidade de dados que podem estar inseridos em subconjunto fuzzy em que o grau de pertinência não é nem 0 ou 1, mas sim, algo entre estes valores.

O critério de parada é expresso em termos da acurácia dos graus de pertinência aplicados ao conjunto de dados, no sentido de determinar o número de iterações durante o processo de minimização. Esta é calculada utilizando o grau de pertinência de uma iteração para outra, considerando o maior valor de $\mu_{ij}$ para todos os dados $x_i$ em todos os (clusters) $c_j$. A representação desta medida entre a iteração $k$ e $k+1$ é dada por:

$$\epsilon = \Delta_i^N \Delta_j^C |\mu_{ij}^{k+1} - \mu_{ij}^k| \tag{1.4}$$

onde $\mu_{ij}^k$ e $\mu_{ij}^{k+1}$ são os graus de pertinência nas respectivas iterações e operador $\Delta$, retorna o maior valor do vetor em análise.

**Algoritmo 1:** Algoritmo Fuzzy c-Means

**Agrupamento**

Determinar a quantidade de partições $c$;

Determinar o erro máximo $e$;

Inicializar os centros aleatoriamente;

Inicializar o contador de iterações $t = 0$;

**repita**

Incrementar $t$;

Atualizar $\mu_h$;

Atualizar $\mathcal{C}$;

**até que** $\|C^{(t)} - C^{(t-1)}\| < (e)$

$$Regra\ R_j:\ Se\ x_1 \text{ é } A_{j1}\ e\ ...\ e\ x_n \text{ é } A_{jn} \text{ então } Classe\ C_j,\ j = 1, 2, ..., N \tag{1.5}$$

onde:

- $x = x_1, ..., x_n$: $n-$dimensional vetor de padrões;
- $A_{ij}$: valor linguístico antecedente, $(i = 1, 2, ..., n)$;
- $C_j$: classe consequente;
- $N$: número de regras fuzzy SE-ENTÃO.

Na abordagem do artigo a ponderação das regras SE-ENTÃO com os graus de certeza é dada por:

$$Regra\ R_j:\ Se\ x_i \text{ é } A_{j1}\ e...e\ x_n \text{ é } A_{jn} \text{ então } Classe\ C_j \text{ com } CF_j, j = 1, 2, ..., N \tag{1.6}$$

onde $CF_j$ é o grau de certeza do consequente, $C_j$, de cada regra fuzzy SE-ENTÃO $R_j$, é um número real no intervalo [0,1].

Para esta aplicação foi utilizada a base de dados *Iris* (1936), do biólogo e estatístico britânico Ronald Fisher, para três espécies da flor *Iris* (*setosa, virginica, versicolor*), com 50 amostras de cada espécie totalizando 150 amostras no conjunto de dados. Considera-se o conjunto de dados separáveis pela discriminação das seguintes características:

- comprimento e largura da sépala;
- comprimento e largura da pétala.

A Figura 1, ilustra uma representação bi-dimensional da base dados *Iris* considerando apenas duas variáveis dos mesmos, $x_1$ e $x_2$, as quais representam comprimento e largura da sépala, respectivamente.

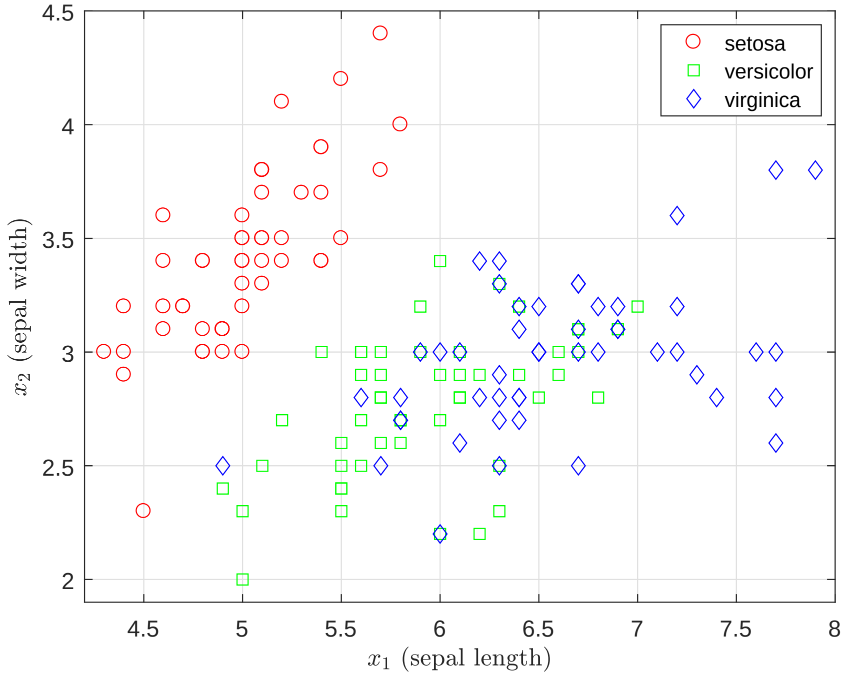


Figura 1: Representação bi-dimensional dos dados considerando as variáveis $x_1$ e $x_2$, comprimento e largura da sépala, respectivamente.

## 2 Metodologia

O método implementado consiste de uma abordagem de seleção de apenas uma classe, ou classe vencedora, durante a fase de classificação, além de, quando na observação da área de decisão haver fronteiras de classificação tendo em vista as regras fuzzy. Ao ponderar-se estas pelo grau de certeza $C_j$:

$$\sum_{p \in Class\ C_j} \mu_j(x_p) = \max\{ \sum_{p \in Class\ k} \mu(x_p) : k = 1, 2, ..., c\} \tag{2.1}$$

onde: $x_p$: número de padrões e $c$: número de classes.

Como pode ser visto na Eq. (2.1) a classe consequente $C_j$ é especificada como a classe dominante no espaço fuzzy correspondente ao antecedente de cada regra fuzzy SE-ENTÃO. Aplicando-se a definição demonstrada na Eq. (1.6), um novo padrão $xp = (x_{p1}, ..., x_{pn})$, pode ser classificado a partir de:

$$\mu_j^*(x_p).CF_j = \max\{\mu_j(x_p).CF_j : \ j = 1, 2, ..., N\} \tag{2.2}$$

A classe consequente pode ser determinada a partir de padrões de treinamento, assim como se esta é a classe dominante em determinado subespaço do espaço fuzzy. Como toda regra tem sua própria área de decisão, cujo tamanho é determinado pelo grau de certeza e pelo antecedente linguístico das funções de pertinência, a abordagem realiza mudança na dimensão da área de decisão por ponderação. Neste sentido o grau de certeza pode ser obtido a partir da Eq. (2.3).

$$CF_j = \frac{\beta_{Classe\ C_j}(R_j) - \overline{\beta}}{\sum_{k=1}^{c} \beta_{Classe\ k}(R_j)} \tag{2.3}$$

onde $C_j$ é a classe consequente, e:

$$\overline{\beta} = \frac{\sum_{k \neq C_j} \beta_{Classe\ k}(R_j)}{(c-1)} \tag{2.4}$$

As formulações indicadas nas equações (2.3) e (2.4) estendem a determinação do grau de certeza para um problema de classificação com $c$ classes.

# 3 Resultados

Inicialmente o modelo foi descrito através de apenas duas variáveis do conjunto de dados, as quais correspondem ao comprimento e largura da sépala de todas as classes, afim de se visualizar as delimitações das regiões de classificação.

Em seguida, todas as quatro variáveis do conjunto de dados são utilizadas para criação de um modelo de classificação geral. A abordagem empregada consiste na técnica de validação cruzada, onde são utilizados de maneira aleatória $70\%$ dos dados para treinamento e os $30\%$ dos dados restantes para validação do modelo.

Além disso, o número de funções de pertinência para descrição das variáveis linguísticas é alterado de $2$ até $15$ funções de pertinências. Todo o processo é repetido $25$ vezes para um número fixo de funções de pertinência, afim de se obter o erro médio de classificação, desconsiderando a natureza estocástica da escolha dos dados de treinamento de validação.

A escolha do tipo de função de pertinência é a mesma apresentada em [1], funções triangulares igualmente espaçadas. Na Figura 2 são ilustradas 6 funções de pertinência para cada uma das variáveis da base de dados *Iris*, considerando as variáveis já normalizadas.

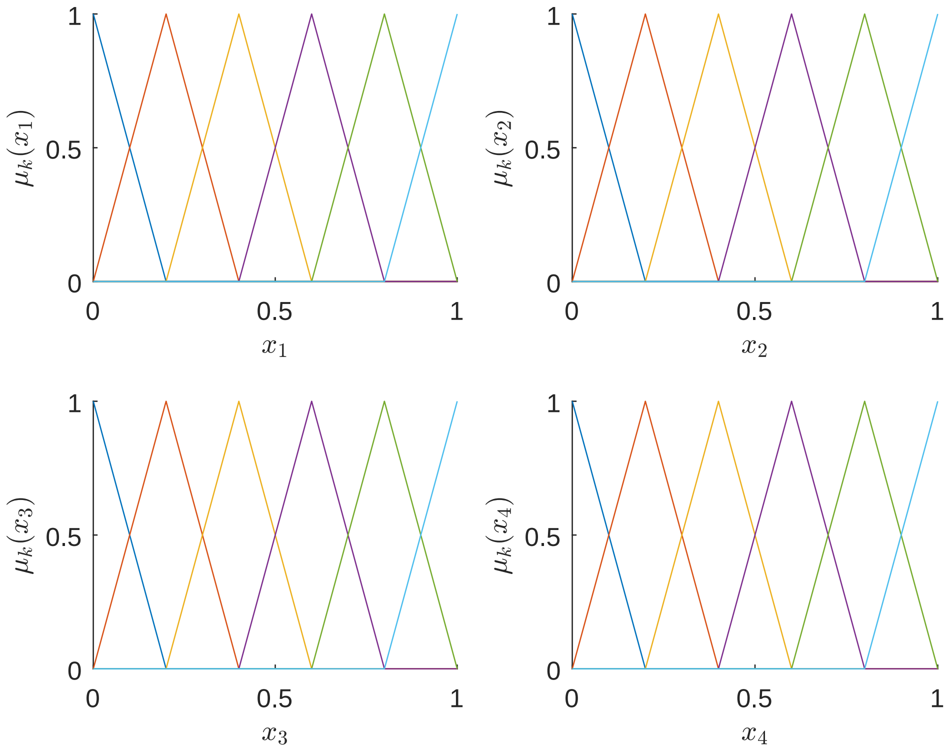


Figura 2: Funções de pertinência para variável do banco de dados *Iris*.

## 3.1 Regiões de Classificação

Como os dados possuem quatro variáveis de entradas, a visualização das regiões de classificação é prejudicada, para exemplificar, apenas duas variáveis, comprimento e largura da sépala, foram utilizadas no treinamento afim de se obter regiões de classificação bi-dimensionais, conforme dito anteriormente.

Nesta etapa, a partir da definição do grau de certeza na observação das características do conjunto de dados, o sistema de classificação converge para uma discriminação das classes (espécies), como uma região de classificação, a partir da especificação de cada regra e da fronteira de classificação.

As Figuras 3 e 4 ilustram as regiões de classificação obtidas a partir da utilização das *t-normas*, produto e mínimo, respectivamente. Em ambas as Figuras foram utilizadas 8 (oito) funções de pertinência triangulares para definição dos antecedentes das variáveis, $x_1$ e $x_2$. O motivo dessa escolha será abordado na Subseção 3.2.

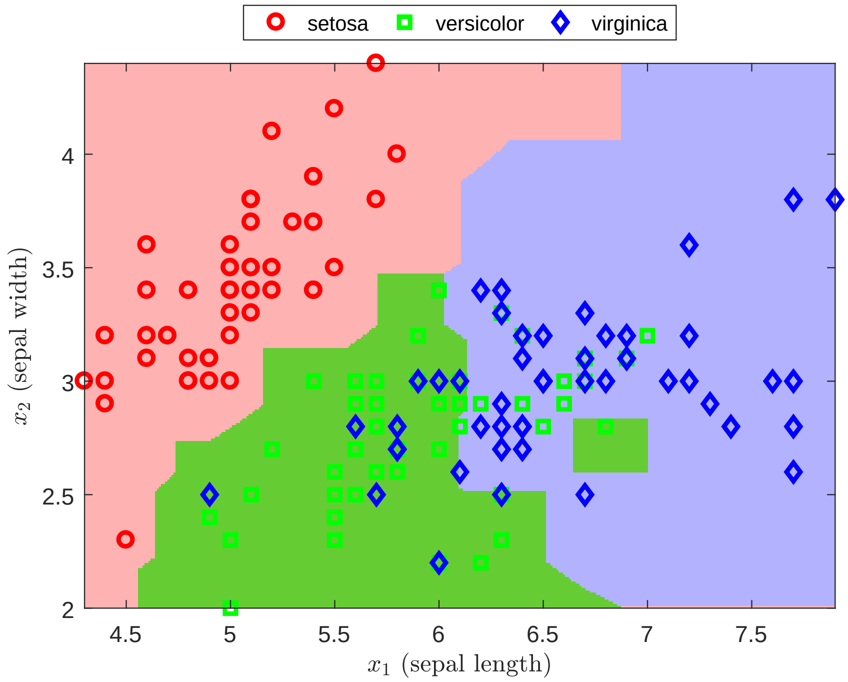


Figura 3: Região de classificação usando a *t-norma* produto.

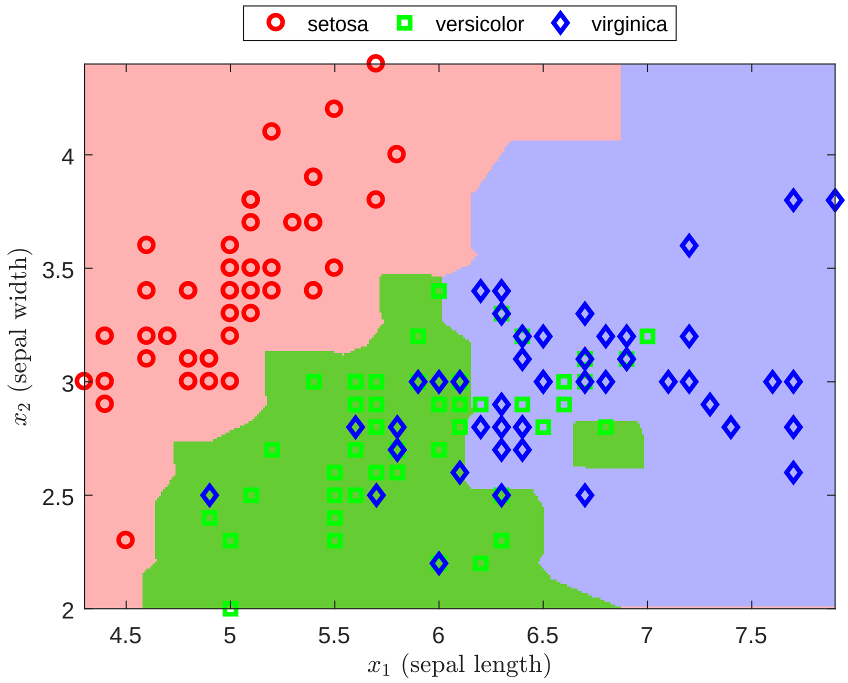


Figura 4: Região de classificação usando a *t-norma* mínimo.

O particionamento do espaço de entrada (variáveis) fornece o suporte para formulação de regras que estejam relacionadas a este subespaço. Em termos práticos, a utilização destes dois tipos de conjunções lógicas, implica na análise das regras, sejam elas condicionais (relação interna entre antecedentes e consequentes) e incondicionais, de maneira geral afirmativas.

Nota-se que o processo de inferência, ou seja, o processo de avaliação das regras para cada subconjunto fuzzy acaba por estabelecer relações. Neste caso, ao combinar o conjunto de dados de entrada no sistema de classificação, ao menos uma regra deve ser ativada ponderada pelo grau de certeza.

Além disso, os operadores mínimo e produto algébrico, utilizados como *t-norma* no que se refere a inferência composicional de regras, estabelecem a interseção entre regras para a dedução de uma consequência lógica.

## 3.2 Treinamento e Validação

Na abordagem computacional, por meio da técnica de validação cruzada, cada conjunto de espécies foi separado de forma aleatória em subconjuntos de treinamento e validação representados, com dimensão de $70\%$ e $30\%$, respectivamente de cada classe de padrões.

Para todas as classes, no total, quatro dados de entrada aplicados a três espécies, com $50$ amostras. Como ferramenta de aprendizagem, os conjuntos de treinamento e validação, foram determinados observando-se os percentuais em $35$ e $15$ amostras, respectivamente, para cada classe $C_j$.

Neste sentido, na etapa de treinamento, a partir de aprendizado supervisionado, houve a classificação inicial das amostras, como função da quantidade de regras e dos graus de certeza, no sentido de se estabelecer as condições e parâmetros iniciais do sistema de classificação.

A partir destes, houve a implementação das características de contexto adquiridas na etapa de treinamento, no sentido de aferir-se o desempenho do sistema de classificação.

Os experimentos foram realizados $25$ vezes, para cada número de funções de pertinência de $2$ até $15$, com o propósito de se observar o limiar de convergência e paridade entre o erro do conjunto de treinamento e erro no conjunto de validação e eliminar a natureza estocástica da escolha do dados de treinamento de validação.

As Figuras 5 e 6 apresentam a evolução do erro de classificação, quando avaliados os dados de treinamento e validação, em função do número de funções de pertinência.

Nota-se que, para ambas as *t-normas* o número dito ideal de funções de pertinência é $8$. De modo que, para um número menor de funções de pertinência ocorre o chamado *under-*

*fitting*, e para um número maior de funções de pertinência ocorre o chamado *overfitting*.

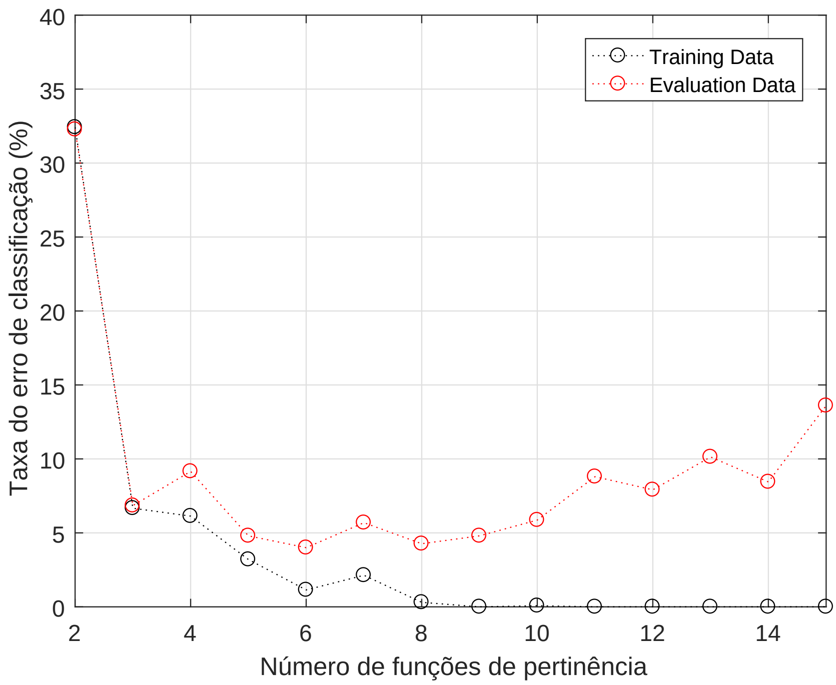


Figura 5: Trajetória do erro em relação ao número de funções de pertinência, considerando a avaliação dos dados de treinamento e validação para a *t-norma* produto.

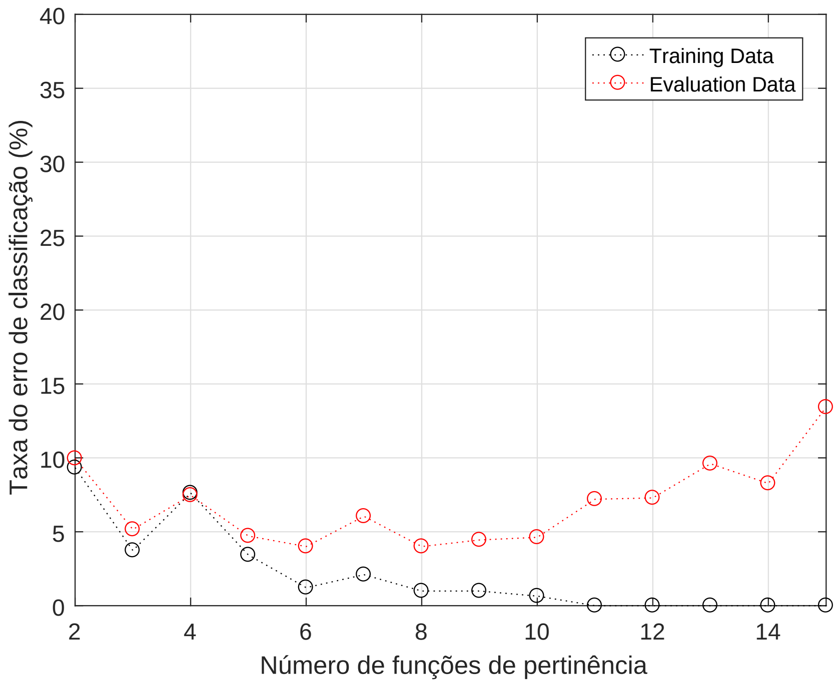


Figura 6: Trajetória do erro em relação ao número de funções de pertinência, considerando a avaliação dos dados de treinamento e validação para a *t-norma* mínimo.

# 4 Conclusões

A utilização de sistemas de classificação baseados em regras ponderadas através de graus de certeza, oferece uma solução baseada em conhecimento dedutivo, o que em termos práticos, implica na sobreposição de uma regra dominante sobre outra no espaço fuzzy, ou seja, na combinação de $n$ regras, pelo menos uma, por inferência composicional, forma a consequência lógica e tendência.

Todos os procedimentos realizados no artigo e replicados nas simulações descritas compõem base de conhecimento por raciocínio aproximado, pelo fato da definição de consequentes a partir de relações de compatibilidade. Neste sentido a conclusão das conjunções lógicas, seja pelo operador mínimo ou produto como *t-norma*, são baseadas no domínio de uma regra sobre a outra no espaço de decisão, sendo este ponderado pelo grau de certeza.

Assim, o procedimento apresentado em [1], o qual consiste na ponderação de regras SE-ENTÃO a partir de graus de certeza substitui a necessidade de otimização das funções de pertinências empregadas para descrição das variáveis linguísticas.

# Referências

[1] ISHIBUCHI, Hisao; NAKASHIMA, Tomoharu. Effect of rule weights in fuzzy rule-based classification systems. IEEE Transactions on Fuzzy Systems, v. 9, n. 4, p. 506-515, 2001.