

## Trabalho Computacional II - Inferência Nebulosa

---

Prof. Cristiano Leite de Castro

23 de outubro de 2018

### 1 TAREFAS

1. Para resolver o Exercício 2 recomenda-se a leitura do Capítulo 4 do livro texto: Jyh-Shing Roger Jang and Chuen-Tsai Sun. 1996. Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. Para resolver o Exercício 3, que se refere a um problema de classificação de padrões, recomenda-se a leitura do artigo "*Effect of Rule Weights in Fuzzy Rule-Based Classification Systems*" que pode ser obtido neste link:  
<https://ieeexplore.ieee.org/document/940964>
2. Seja a função  $y = \cos(x)$ , para  $x$  definido no intervalo de  $[-\pi/2, 3\pi/2]$ , conforme ilustra a Figura 2. Pede-se:
  - (a) Empregue o mecanismo de inferência de Sugeno com consequentes de ordem 1 e obtenha uma expressão analítica para aproximar esta função. Dica: use funções de pertinência do tipo triangular para "fuzzificação" da variável  $x$ .
  - (b) Mostrar o gráficos da aproximação (regressão) e calcular o Erro Quadrático Médio  $EQM = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$ , onde  $y_i$  é a saída real da função e  $\hat{y}_i$  é a saída obtida pelo sistema nebuloso.
3. Projete um classificador binário (duas classes) baseado em regras nebulosas. O número de regras ( $K$ ) e as funções de pertinência para os antecedentes e consequentes das regras devem ser definidas com base no algoritmo de agrupamento *Fuzzy K-Means*, da seguinte forma:

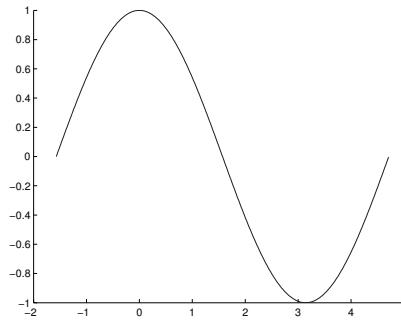


Figura 1.1: Função  $y = \text{cosseno}(x)$  no intervalo de  $[-\pi/2, 3\pi/2]$ .

- Dado o conjunto de dados formado por uma matriz  $X$  com dimensões  $n \times d$  (onde  $n$  é o número de padrões e  $d$  é a dimensão do espaço de entrada) e uma matriz  $Y$  com dimensões  $n \times 1$  (indicando os rótulos para os padrões), divida  $(X, Y)$  em duas partições: 70% dos padrões para treinamento  $(X_t, Y_t)$  e o restante (30%) para validação  $(X_v, Y_v)$ .
- Aplique o algoritmo *Fuzzy K-Means* sobre  $X_t$ . O centróide do  $j$ -ésimo grupo (*cluster*) deve corresponder a uma regra do tipo:  
regra  $j$ : se  $x_1$  é  $A_{1j}$  e  $x_2$  é  $A_{2j}$  e ... e  $x_d$  é  $A_{dj}$  então  $y_j = c$  com  $c \in \{0, 1\}$
- Defina o antecedente  $A_{ij}$  como uma função de pertinência Gaussiana com centro  $c_{ij}$  igual à projeção do centróide do grupo  $j$  na variável de entrada  $i$ .
- Proponha uma forma de calcular a dispersão ( $\sigma_{ij}$ ) de  $A_{ij}$  a partir dos valores de pertinência para o grupo  $j$  obtidos com o algoritmo *Fuzzy K-Means* (coluna  $j$  da matriz  $U$ ).
- O valor  $c$  para o conseqüente ( $y_j$ ) deve ser definido como a classe (0 ou 1) que fornece o valor máximo para a soma dos valores de pertinência do grupo  $j$ . Para isso, some todos os valores de pertinência do grupo  $j$  por classe. Use a matriz  $Y_t$  para descobrir qual é a classe de cada padrão.
- A saída do classificador nebuloso para um dado padrão de entrada  $\mathbf{x} = \{x_1, x_2, \dots, x_d\}$  deve ser calculada por:
  - para cada regra  $j$ , calcule os valores de pertinência  $\mu_{A_{ij}}(x_i)$  para cada variável de entrada  $x_i$ .
  - para cada regra  $j$ , obtenha o grau de ativação:  $\omega_j = \prod_{i=1}^d \mu_{A_{ij}}(x_i)$ .
  - agregue os graus de ativação das regras que possuem o mesmo conseqüente (mesma classe) utilizando o operador soma probabilística (s-norma) e escolha como saída para o padrão  $\mathbf{x}$  a classe que fornece o maior valor agregado.
- Calcule a acurácia do classificador nebuloso para o conjunto de validação

$(X_v, Y_v)$ , considerando diferentes valores para  $K = 2, 3, \dots, 8$  (número de regras). Mostre um gráfico da acurácia em função de  $K$ .

- Mostre a superfície de separação gerada pelo classificador no espaço de entrada para 3 diferentes valores de  $K$ .
- Descreva a estratégia proposta por você para calcular  $\sigma_{ij}$ .
- Teste seu algoritmo de classificação utilizando as seguintes bases de dados:
  - a) **dataset\_2d.mat**: base de dados sintética que está disponível no moodle. Neste arquivo  $x$  se refere aos dados de entrada e  $y$  as respectivas saídas (classes).
  - b) Pima Diabetes: base de dados real que pode ser obtida neste link:  
<https://www.kaggle.com/uciml/pima-indians-diabetes-database>