

Disciplina: Aprendizagem de Máquina

Prof. Francisco de A.T. de Carvalho

## Projeto

- 1) Considere no Tolbox do MatLab, um conjunto de 300 padrões formando 3 classes, uma classe com 150, uma outra com 100 e a ultima com 50 padrões. Esses padrões so descritos por 2 variaveis quantitativas. Os padrões de cada classe so gerados a partir de distribuies normais bi-variadas segundo os seguintes parmetros:

- ▶ Classe 1:  $\mu_1 = 0$ ,  $\mu_2 = 0$ ,  $\sigma_1 = 2$ ,  $\sigma_2 = 1$ ,  $\sigma_{12} = 1.7$ ,  
 $\rho_{12} = 0.85$
- ▶ Classe 2:  $\mu_1 = 0$ ,  $\mu_2 = 3$ ,  $\sigma_1 = 0.5$ ,  $\sigma_2 = 0.5$ ,  $\sigma_{12} = 0.0$ ,  
 $\rho_{12} = 0.0$
- ▶ Classe 2:  $\mu_1 = 4$ ,  $\mu_2 = 3$ ,  $\sigma_1 = 2$ ,  $\sigma_2 = 1$ ,  $\sigma_{12} = -1.7$ ,  
 $\rho_{12} = -0.85$

A partir da matriz “objetos  $\times$  atributos” use a distancia Euclideana para obter uma matriz de dissimilaridades “objetos  $\times$  objetos”. Aplique a essa matriz de dissimilaridades o algoritmo de agrupamento *SFMdd* (The partitioning fuzzy K-medoids clustering algorithm based on a single dissimilarity matrix) com 2 clusters 100 vezes e selecione o melhor resultado segundo o critério de adequao entre os clusters e seus representantes. Calcule o índice de Rand Corrigido.

Para detalhes do algoritmo *SFMdd* ver o artigo:

F.A.T. de Carvalho, Y. Lechevallier and F.M de Melo. Relational partitioning fuzzy clustering algorithms based on multiple dissimilarity matrices. Fuzzy Sets and Systems (Accepted)

- 2) Considere os dados do item 1. Considere a seguinte regra de deciso: afetar o padro  $k$  a classe  $j$  se:

$$j = \operatorname{argmax}_i P(\omega_i | x_k, \theta_i) \text{ com } P(\omega_i | x_k, \theta_i) = \frac{p(x_k | \omega_i, \theta_i) P(\omega_i)}{\sum_{j=1}^c p(x_k | \omega_j, \theta_j) P(\omega_j)}$$

Em relação à estimativa de  $P(\omega_i | x_k, \theta_i)$ , estime  $P(\omega_i)$ ,  $i=1,2$ , pelo método de máxima verossimilhança e considere os seguintes casos:

- a) Para a classe 1, estime  $P(\omega_i|x_k, \theta_i)$  pelo mtodo da mxima verossimilhana, supondo uma normal mutivariada. Em relao a classe 2, suponha a existncia de dois componentes e estime  $P(\omega_i|x_k, \theta_i)$  usando o algoritmo EM, supondo uma mistura de distribuies normais multivariadas;
  - b) Estime  $P(\omega_i|x_k, \theta_i)$ ,  $i=1,2$  usando Janela de Parzen. Une a função de kernel gaussiana. Varie o parâmetro de suavizao  $h$ .
  - c) Estime diretamente  $P(\omega_i|x_k, \theta_i)$ ,  $i=1,2$  pelo k-vizinhos. Use a distância Euclidiana. Varie o número de vizinhos.
  - d) Estime  $P(\omega_i|x_k, \theta_i)$ ,  $i=1,2$  usando a combinação desses classificadores pela soma.
- 3) Use validao cruzada estratificada (ou hold-out estratificado) para avaliar e comparar esses classificadores

## Observações

- 1) No Relatório e na saída da ferramenta devem estar bem claros:  
i) como foi realizada a combinação dos classificadores; ii) como foram organizados os experimentos de tal forma a realizar corretamente a avaliação dos modelos e a comparação entre os mesmos. Fornecer também uma descrição dos dados.
- 2) Data de Entrega do projeto: 11/11/2012
- 3) Enviar por email: o programa fonte, o executável, os dados e o relatório do projeto
- 4) PASSAR NA MINHA SALA PARA ASSINAR A ATA DE ENTREGA DO TRABALHO EM 12/11/2012
- 5) O projeto deve ser realizado sozinho ou em dupla