

Projeto AM 2014-2

Francisco de A. T. de Carvalho¹ Teresa B. Ludermir²

1 Centro de Informatica-CIn/UFPE
Av. Prof. Luiz Freire, s/n -Cidade Universitaria, CEP 50740-540, Recife-PE, Brasil,
{fatc,tbl}@cin.ufpe.br

Use o Toolbox do MatLab para gerar 300 objetos formando 2 classes, a classe 1 com 200 objetos e a classe 2 com 100 objetos. Esses objetos são descritos por duas variáveis quantitativas.

Os objetos de cada classe são gerados à partir de distribuições normais bi-variadas com os seguintes parâmetros:

- Classe 1-1: $\mu_1 = 60$, $\mu_2 = 30$, $\sigma_1^2 = 9$, $\sigma_2^2 = 144$, $\sigma_{12} = 0$, $\rho_{12} = 0$
- Classe 2: $\mu_1 = 45$, $\mu_2 = 22$, $\sigma_1^2 = 100$, $\sigma_2^2 = 9$, $\sigma_{12} = 0$, $\rho_{12} = 0$
- Classe 1-2: $\mu_1 = 52$, $\mu_2 = 30$, $\sigma_1^2 = 9$, $\sigma_2^2 = 9$, $\sigma_{12} = 0$, $\rho_{12} = 0$

100 objetos da classe 1 são gerados à partir da distribuição 1-1 e os outros 100 à partir da distribuição 1-2

- 1) Aplique a esses objetos o algoritmo "FCM-DFCV" 100 vezes para obter uma partição fuzzy com 3 clusters e selecione o melhor resultado segundo a função objetivo. A partir da partição fuzzy encontre a partição hard e calcule o índice de Rand corrigido.

Para detalhes do algoritmo VKFCM-K-LP ver a seção 3.2.2 do artigo:

“Francisco de A.T. de Carvalho, Camilo P. Tenório and Nicomedes Cavalcanti Junior, Partitional fuzzy clustering methods based on adaptive quadratic distances, Fuzzy Sets and Systems, 157, 2833-2857, 2006”.

Observações:

- Parâmetros do algoritmo: $m = 2$, $T = 150$, $\epsilon = 0.0000000001$
- Para o melhor resultado imprimir: a partição fuzzy (a matrix **U** dos graus de pertinencia dos objetos nos grupos), a partição hard (os grupos), a matrix de protótipos e os protótipos de cada grupo, a matrix de pesos, o índice de Rand corrigido.

2) Considere a tabela de dados “objetos \times variáveis”. Os 300 exemplos são rotulados segundo as classes 1 e 2.

Considere a seguinte regra de decisão: afetar o exemplo

\mathbf{x}_k ($k = 1, \dots, 300$) a classe ω_j ($j = 1, 2$) se

$$j = \arg \max_{i \in \{1, 2\}} P(\omega_i | \mathbf{x}_k) \text{ com } P(\omega_i | \mathbf{x}_k) = \frac{p(\mathbf{x}_k | \omega_i) P(\omega_i)}{\sum_{j=1}^c P(\mathbf{x}_k | \omega_j) P(\omega_j)}$$

Estime $P(\omega_i)$ ($i = 1, 2$) pelo método da máxima

verossimilhança e considere os seguintes casos:

- a) Para a classe 2, estime $p(\mathbf{x}_k | \omega_2) = p(\mathbf{x}_k | \omega_2, \theta_2)$ pelo método da máxima verossimilhança, supondo uma normal bi-variada.

Para a classe 1, estime usando o algoritmo EM

$p(\mathbf{x}_k | \omega_1) = p(\mathbf{x}_k | \omega_1, \theta_1)$ supondo uma mistura de duas normais bivariadas tal que

$$p(\mathbf{x}_k | \omega_1, \theta_1) = \sum_{j=1}^2 p(\mathbf{x}_k | \omega_{1j}, \theta_{1j}) P(\omega_{1j})$$

a) (cont.) Compute

$$P(\omega_i|\mathbf{x}_k) = P(\omega_i|\mathbf{x}_k, \theta_i) = \frac{p(\mathbf{x}_k|\omega_i, \theta_i)P(\omega_i)}{\sum_{j=1}^C P(\mathbf{x}_k|\omega_j, \theta_j)P(\omega_j)} \quad (i = 1, 2)$$

b) Estime $p(\mathbf{x}_k|\omega_i) = p(\mathbf{x}_k|\omega_i, \theta_i)$ ($i = 1, 2$) usando Janela de Parzen. Use a função de kernel bivariada produto. Para cada componente use a função de kernel gaussiana. Varie o parâmetro de suavização h . Compute

$$P(\omega_i|\mathbf{x}_k) = P(\omega_i|\mathbf{x}_k, \theta_i) = \frac{p(\mathbf{x}_k|\omega_i, \theta_i)P(\omega_i)}{\sum_{j=1}^C P(\mathbf{x}_k|\omega_j, \theta_j)P(\omega_j)} \quad (i = 1, 2)$$

~~c)~~ Estime diretamente $P(\omega_i|\mathbf{x}_k) = P(\omega_i|\mathbf{x}_k, \theta_i)$ ($i = 1, 2$) pelo método dos k-vizinhos mais próximos. Use a distância Euclidiana. Varie o número de vizinhos. (classificando tbm)

d) Usar MLP para fazer a classificação dos dados.

e) Use a regra da soma para classificar o exemplo \mathbf{x}_k a partir do cálculo de $P(\omega_i|\mathbf{x}_k)$ obtido pelos classificadores bayesiano, Janela de Parzen e k-vizinhos

- f) Obtenha uma estimativa pontual e um intervalo de confiança para a taxa de erro para cada classificador
 - g) usar Friedman test (teste não parametrico) para comparar os classificadores. Usar também o Nemenyi test (pos teste)
- 3) Use validação cruzada estratificada para avaliar e comparar esses classificadores

Observações

- 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.
- Data de apresentação e entrega do projeto: 04/02/2015
- Enviar por email: o programa fonte, o executável, os dados e o relatório do projeto
- PASSAR NA MINHA SALA PARA ASSINAR A ATA DE ENTREGA DO TRABALHO EM 04/02/2015
- ALUNOS DE PÓS-GRADUAÇÃO: O projeto pode ser realizado com no máximo 2 alunos.
- ALUNOS DE PÓS-GRADUAÇÃO: O projeto pode ser realizado com no máximo 4 alunos.