

Projeto AM 2016-2

Francisco de A. T. de Carvalho¹ Cleber Zenchettin²

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

- 1) Considere a tabela de dados "abalone" do site uci machine learning repository (<http://archive.ics.uci.edu/ml/>).
 - Execute o algoritmo "VKCM-K" nessa tabela de dados 100 vezes para obter uma partição em 3 grupos e selecione o melhor resultado segundo a função objetivo. Para detalhes do algoritmo "VKCM-K" veja a seção 3 do artigo:
"M.R.P. Ferreira, F.A.T. de Carvalho and E.C. Simões, Kernel-based hard clustering methods with kernelization of the metric and automatic weighting of the variables, Pattern Recognition, 51, 310-321, 2016".

Observações:

- No algoritmo "VKCM-K", os prototipos dos grupos são calculados com a equação (12), os pesos de relevância das variáveis são calculados com a equação (16), os grupos são atualizados com as equações (11) e (17);
- Parametros: número de grupos $K = 3$
- Para o melhor resultado imprimir: i) a matriz de prototipos, ii) o vetor de pesos de relevancia das variáveis, iii) a partição (para cada grupo, a lista de objetos), iv) 0 índice de Rand corrigido.

Considere novamente a tabela de dados "abalone". Os exemplos são rotulados segundo as classes M, F, I.

- a) Use validação cruzada estratificada para avaliar e comparar classificadores descritos abaixo. Se necessário, retire do conjunto de aprendizagem, um conjunto de validação para fazer ajuste de parametros.
- b) Obtenha uma estimativa pontual e um intervalo de confiança para a taxa de erro para cada classificador;
- c) Usar Friedman test (teste não parametrico) para comparar os classificadores. Se necessário, usar também o Nemenyi test (pos teste);

2) Considere os seguintes classificadores:

- i) Classificador bayesiano. Considere a seguinte regra de decisão: afetar o exemplo \mathbf{x}_k a classe ω_l se $l = \arg \max_{1 \leq i \leq c} P(\omega_i | \mathbf{x}_k)$ com

$$P(\omega_i | \mathbf{x}_k) = \frac{p(\mathbf{x}_k | \omega_i) P(\omega_i)}{\sum_{r=1}^c p(\mathbf{x}_k | \omega_r) P(\omega_r)}$$

- a) Estime $P(\omega_i)$ pelo metodo de maxima verossimilhança.
- b) Para cada classe ω_i ($i = 1, 2, 3$) estime $p(\mathbf{x}_k | \omega_i) = p(\mathbf{x}_k | \omega_i, \theta_i)$ pelo método da máxima verossimilhança, onde $\theta_i = \begin{pmatrix} \mu_i \\ \Sigma_i \end{pmatrix}$,

$$\mu_i = \begin{pmatrix} \mu_{i1} \\ \dots \\ \mu_{ij} \\ \dots \\ \mu_{id} \end{pmatrix}, \text{ e } \Sigma_i = \begin{pmatrix} \lambda_{i1} & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \lambda_{ij} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \dots & \lambda_{id} \end{pmatrix},$$

i)) Classificador bayesiano.

b) supondo uma normal multivariada:

$$p(\mathbf{x}_k | \omega_i, \theta_i) = (2\pi)^{-\frac{d}{2}} (\prod_{j=1}^d \lambda_{ij})^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \sum_{j=1}^d \frac{(x_{kj} - \mu_{ij})^2}{\lambda_{ij}} \right\}. \text{ Use como}$$

$$\text{estimativas } \hat{\mu}_i = \begin{pmatrix} \hat{\mu}_{i1} \\ \dots \\ \hat{\mu}_{ij} \\ \dots \\ \hat{\mu}_{id} \end{pmatrix}, \text{ e } \hat{\Sigma}_i = \begin{pmatrix} \hat{\lambda}_{i1} & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \hat{\lambda}_{ij} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \dots & \hat{\lambda}_{id} \end{pmatrix}, \text{ onde}$$

$$\hat{\mu}_{ij} = \frac{1}{n} \sum_{k=1}^n x_{kj} \text{ e } \hat{\lambda}_{ij} = \frac{1}{n} \sum_{k=1}^n (x_{kj} - \hat{\mu}_{ij})^2$$

ii) Usar MLP e SVM para fazer a classificação dos dados.

iii) Voto majoritário: afetar o exemplo \mathbf{x}_k a classe ω_j se

$$\Delta_{jkB} + \Delta_{jkMLP} + \Delta_{jkSVM} = \max_{r=1}^3 (\Delta_{rkB} + \Delta_{rkMLP} + \Delta_{rkSVM}) \text{ onde}$$

$$\Delta_{rkB} = 1 \text{ se } P(\omega_r | \mathbf{x}_k) = \max_{l=1}^3 P(\omega_l | \mathbf{x}_k); \Delta_{rkB} = 0 \text{ senão;}$$

$$\Delta_{rkMLP} = \Delta_{rkSVM} = 1 \text{ se o objeto } \mathbf{x}_k \text{ é classificado na classe } \omega_r;$$

$$\Delta_{rkMLP} = \Delta_{rkSVM} = 0 \text{ senão.}$$

Observações Finais

- No Relatório e na saída da ferramenta devem estar bem claros:
 - a) 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: QUARTA-FEIRA 23/11/2016
- Enviar por email : o programa fonte, o executável (se houver), os dados e o relatório do projeto
- PASSAR NA MINHA SALA PARA ASSINAR A ATA DE ENTREGA DO TRABALHO EM 23/11/2016
- ALUNOS DE PÓS-GRADUAÇÃO: o projeto pode ser realizado com no máximo 2 alunos.
- ALUNOS DE GRADUAÇÃO: o projeto pode ser realizado com no máximo 4 alunos.