

2o. Trabalho Computacional (TIP8311 – Reconhecimento de Padrões)

Programa de Pós-Graduação em Engenharia de Teleinformática (PPGETI)
Departamento de Engenharia de Teleinformática (DETI)
Universidade Federal do Ceará (UFC)

Responsável: Prof. Guilherme de Alencar Barreto

Data: 13/11/2017

Questão 1 (classificadores gaussianos): Implementar os quatro classificadores gaussianos estudados em sala de aula usando o conjunto de dados escolhidos para o 1º. trabalho computacional.

Realizar o treinamento e o teste dos classificadores de acordo com a seguinte metodologia:

(i) Separar aleatoriamente os dados em um conjunto de treinamento (80%) e outro de teste (20%) e treinar os classificadores. Repetir este procedimento por 100 rodadas.

(ii) Reportar os valores médio, mínimo, máximo, bem como a mediana e o desvio padrão da taxa de acerto, para cada um dos classificadores. Mostrar também o boxplot da taxa de acerto para cada um dos classificadores.

(iii) Determinar a matriz de confusão para o melhor e o pior resultado de cada um dos classificadores implementados. No total são 6 matrizes de confusão, duas por classificador.

(iv) Avaliar a evolução da taxa de acerto médio, mínimo e máximo em função da proporção de exemplos de treinamento/teste. Ou seja, testar para 10/90, 20/80, 40/60, 50/50, 60/40, 80/20, 90/10.

Obs.: Os 4 classificadores gaussianos a serem implementados são: Caso 1 ($\Sigma_i = \sigma^2 \mathbf{I}$), Caso 2 ($\Sigma_i = \Sigma$ – *pooled* ou global), Caso 3 ($\Sigma_i = \Sigma_i$) e Caso 4 ($\Sigma_i = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$ – Naive Bayes).

Questão 2 (teórica): Demonstrar matematicamente que a técnica PCA, quando aplicada para redução de dimensionalidade, promove uma reconstrução ótima dos dados originais segundo o critério dos mínimos quadrados.

Questão 3 (extração de atributos): Usando apenas o melhor classificador obtido na Questão 1, aplicar PCA ao conjunto escolhido com o objetivo de reduzir a dimensão dos vetores de atributos originais, mantendo nível de classificação equivalente. Em outras palavras, encontrar uma dimensão reduzida “q” de modo que o desempenho do classificador seja pelo menos equivalente ao desempenho do classificador que use os dados originais.

Questão 4 (seleção de atributos): A partir do valor de “q” escolhido na Questão 3, selecionar os “q” atributos mais relevantes do conjunto de dados original. Treinar e testar o classificador em questão usando os “q” atributos selecionados.

Boa Sorte!