

Exercício Aula 6

Prof. Frederico Coelho

23 de agosto de 2019

CLASSIFICADOR DE BAYES APLICADO A UM PROBLEMA MULTIVARIADO 2

Neste exercício o aluno deverá resolver um problema de classificação multivariado utilizando o classificador de bayes. A base de dados será uma real chamada *spambase*. As explicações para esta base podem ser encontradas no moodle.

Passos a serem seguidos:

1. Carregar os dados. Dica: utilize a função `read.csv` (`spam <- read.csv("spam-base.data", sep=" ", header = FALSE)`).
2. Separar os dados em 90% para treinamento e 10% para teste
3. Estimar a densidade multivariada para as duas classes utilizando a expressão da figura 0.1 vista em sala de aula. Dica: comando em R para sua implementação `> pdfnvar <- function(x, m, K, n)((1/(sqrt((2*pi)^n*(det(K))))) * exp(-0.5*(t(x - m)% * %(solve(K))% * %(x - m))))`, onde K é a matriz de covariâncias calculada com a função `cov()`, m é o vetor de médias e n é o número de características.
4. Aplicar o classificador ao conjunto de teste e calcular a acurácia. A acurácia é o número de classificações corretas do conjunto de testes dividido pelo número total de amostras do mesmo conjunto.
5. Deverá ser entregue um relatório em pdf com a acurácia média de acerto e seu desvio padrão para 10 experimentos do seu classificador, utilizando a técnica de validação cruzada.

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right)$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \rho_{n2}\sigma_n\sigma_2 & \cdots & \sigma_n^2 \end{bmatrix}$$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$$

Figura 0.1: PDF multivariada