

## Classificador Bayesiano aplicado a um problema multivalorado

Hernane Braga Pereira - 2014112627

### 1. Introdução

Este relatório tem como objetivo exemplificar o uso do classificador bayesiano em um problema multivalorado e de classificação binária.

### 2. Classificador Bayesiano em um problema multivalorado utilizando 90% dos dados como treinamento

Para este exercício foi utilizada a base de dados *heart*, um modelo que possui 13 variáveis de entrada e uma saída binária, que representa presença, ou ausência de doença no coração. O classificador utilizado foi o bayesiano para  $n$  variáveis de entrada e o modelo foi treinado com 90%, 70% e 20% dos dados de entrada, para se observar as mudanças na acurácia dos resultados.

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

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

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

Figura 1. PDF multivariada utilizada no classificador

| % Amostra de Treinamento por classe | Taxa de acurácia para a Classe 1 | Taxa de acurácia para a Classe 2 | Taxa de acurácia Total |
|-------------------------------------|----------------------------------|----------------------------------|------------------------|
| 90%                                 | 93.3%                            | 83.3%                            | 88%                    |
| 70%                                 | 77.7%                            | 83.3%                            | 80%                    |
| 20%                                 | 84.1%                            | 55.2%                            | 71.2%                  |

Quadro 1. Comparação de acurácia para diferentes tamanhos de amostras de testes

Ao analisar os resultados, conclui-se que ao diminuir a quantidade de dados utilizadas para treinamento, ocorre queda de acurácia total no modelo e em pelo menos uma das classes. A queda de acurácia entre as classes ocorre aleatoriamente, não seguindo um padrão de decaimento, ou preferência de qual categoria sofrerá maiores perdas. Este comportamento se deve a escolha aleatória de dados para treinamento. De toda forma, apesar de haver queda na taxa de acerto total, o modelo consegue classificar corretamente em 71% dos casos, usando apenas 20% dos dados para treinamento, o que indica que as variáveis de entrada do problema conseguem descrever razoavelmente bem seu comportamento final.

### 3. Referências

[1] Classificador Bayesiano:, Notas de aula, agosto de 2019.