

Tratamento e Mineração de Dados

Naive Bayes

Marcos Pereira

4 de maio de 2020

1 Sumário

2 Abordagem Teórica

- Naive Bayes

3 Tratamento de Dados

4 Modelo - Algoritmo

5 Resultados

6 Curvas de Calibração de Probabilidade

$$\mathbf{R} = \begin{matrix} & \text{classe} \\ \left[\begin{array}{ccc|c} x_1 & x_2 & x_3 & y_1 \\ x'_1 & x'_2 & x'_3 & y_2 \\ x_1 & x'_2 & x_3 & y_3 \end{array} \right] & \left. \vphantom{\begin{array}{ccc|c} x_1 & x_2 & x_3 & y_1 \\ x'_1 & x'_2 & x'_3 & y_2 \\ x_1 & x'_2 & x_3 & y_3 \end{array}} \right\} \text{instâncias} \end{matrix} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix}$$

Desse modo os vetores instâncias são os vetores \mathbf{x}_i , é importante notar que as relações equivalem aos nossos data sets.

- Conjunto de Algoritmos de aprendizado supervisionado;
- Utiliza inferência Bayesiana
 - Avaliação de hipóteses pela máxima verossimilhança
- Teorema de Bayes¹:

$$\begin{aligned} P(y|x_1, \dots, x_n) &= \prod_{i=1}^n \frac{P(y)P(x_i|y)}{P(x_i)} \\ &\equiv \frac{P(y) (P(x_1|y)P(x_2|y)\dots P(x_n|y))}{P(x_1)P(x_2)\dots P(x_n)} \end{aligned} \quad (1)$$

- $P(\mathbf{x}) \in \mathbb{R}$ (é constante), então utilizamos a seguinte regra de classificação:

$$P(y|\mathbf{x}) \propto \prod_{i=1}^n P(y)P(x_i|y) \quad (2)$$

¹O símbolo \prod representa um produtório. 

- Máxima Probabilidade A Posteriori para realizar predição de dados:

$$Y = \arg \max_y \prod_{i=1}^n P(y)P(x_i|y) \quad (3)$$

- Naive Bayes Gaussiano. Verossimilhança[1]:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\left(\frac{x_i - \mu_y}{2\sigma_y}\right)^2}$$

- Predição de dados:

$$Y = \arg \max_y P(y) \exp \left[-\frac{1}{2\sigma_y} \sum_{i=1} (x_i - \mu_y)^2 \right]$$



F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.