> Marcos Pereira

Sumário

Abordagei

Consideraçõe Naive Bayes

Tratamento de Dados

Modelo -Algoritmo

Resultados

Curvas de Calibração de Probabi-

## Tratamento e Mineração de Dados Naive Bayes

Marcos Pereira

4 de maio de 2020

## Sumário

2 Abordagem Teórica

■ Naive Bayes

1 Sumário

Tratamento de Dados

4 Modelo - Algoritmo

Resultados

6 Curvas de Calibração de Probabilidade

Marcos Pereira

Sumário

Abordage Teórica

Consideraçõe Naive Bayes

Tratamento de Dados

Modelo -Algoritmo

Resultados

Curvas de Calibração de Probabilidade

$$\mathbf{R} = \begin{bmatrix} x_1 & x_2 & x_3 & y_1 \\ x_1' & x_2' & x_3' & y_2 \\ x_1 & x_2' & x_3 & y_3 \end{bmatrix} \right\}_{\text{instâncias}} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix}$$

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

Marcos Pereira

Sumário

Consideraçõe

Tratamento

Modelo -

Resultados

Curvas de Calibração de Probabi lidade Conjunto de Algoritmos de aprendizado supervisionado;

- Utiliza inferência Bayesiana
  - Avaliação de hipóteses pela máxima verossimilhança
- Teorema de Bayes¹:

$$P(y|x_{1},\dots,x_{2}) = \prod_{i=1}^{n} \frac{P(y)P(x_{i}|y)}{P(x_{i})}$$

$$\equiv \frac{P(y)(P(x_{1}|y)P(x_{2}|y)\dots P(x_{n}|y))}{P(x_{1})P(x_{2})\dots P(x_{n})}$$
(1)

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

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

 $<sup>^1{\</sup>rm O}$ símbolo  $\prod$  representa um produtório.  $\square$  >  $< \varnothing$  > <  $\geqslant$  > <  $\geqslant$  > < > <math display="inline">< >

Resultados

Curvas de Calibração de ProbabiMáxima Probabilidade A Posteriori para realizar predição de dados:

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

■ Naive Bayes Gaussiano. Verossimilhança[1]:

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

■ Predição de dados:

$$Y = \arg\max_{y} P(y) \exp \left[ -\frac{1}{2\sigma_{y}} \sum_{i=1}^{n} (x_{i} - \mu_{y})^{2} \right]$$

> Marcos Pereira

Sumário

Abordager

Considerações

Naive Bayes

Tratamento de Dados

Modelo -Algoritme

Resultados

Curvas de Calibração

lidada

> Marcos Pereira

Sumário

Abordager

Consideraçõe

Naive Bayes

Tratamento de Dados

Modelo -Algoritmo

Resultados

Curvas de Calibração

lidada

> Marcos Pereira

Sumário

Abordage

Consideraçõe

Naive Baye

Tratamento de Dados

Modelo -Algoritmo

Resultados

Curvas de Calibração

de Proi

> Marcos Pereira

Sumário

Abordage

Consideraçõe

Tratament de Dados

Modelo -

Resultados

Curvas de Calibração de Probabilidade

**◆□▶ ◆□▶ ◆■▶ ◆■▶ ■ り**९0

Marcos Pereira

Sumário

Considerações

Tratamento de Dados

Modelo -Algoritmo

Resultados

Curvas de Calibração de Probabilidade



- 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.