

Introdução à Classificação



INFORMAÇÃO,
TECNOLOGIA
& INOVAÇÃO

Classificação

Y é uma variável qualitativa



EXEMPLOS



Classificação

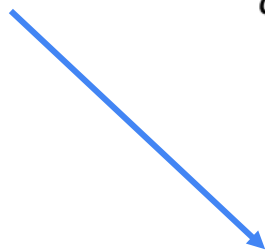
$R(g) = \mathbb{E}[(Y - g(\mathbf{X}))^2]$ não faz mais sentido

Alternativa: $R(g) := \mathbb{E}[\mathbb{I}(Y \neq g(\mathbf{X}))] = \mathbb{P}(Y \neq g(\mathbf{X}))$



Quem minimiza R?

$$g(\mathbf{x}) = \arg \max_d \mathbb{P}(Y = d|\mathbf{x})$$



Classificador de Bayes

$$g(\mathbf{x}) = 1 \iff \mathbb{P}(Y = 1|\mathbf{x}) \geq \frac{1}{2}$$



Plug-in Classifier

(1) Estimar $\mathbb{P}(Y = c|\mathbf{x})$, para cada c .

(2) Tomar

$$g(\mathbf{x}) = \arg \max_{c \in \mathcal{C}} \hat{P}(Y = c|\mathbf{x})$$



REGRESSÃO LOGÍSTICA



Regressão Logística

$$\mathbb{P}(Y = 1|\mathbf{x}) = \frac{e^{\beta_0 + \sum_{i=1}^d \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^d \beta_i x_i}}$$

Interpretação:

$$\beta = \frac{\text{Chance(Comprar|Idade} = 20\text{)}}{\text{Chance(Comprar|Idade} = 19\text{)}}$$



Estimação

$$L(y; (\mathbf{x}, \beta)) = \prod_{k=1}^n (\mathbb{P}(Y_k = 1 | \mathbf{x}_k, \beta))^{y_k} (1 - \mathbb{P}(Y_k = 1 | \mathbf{x}_k, \beta))^{1-y_k}$$
$$\prod_{k=1}^n \left(\frac{e^{\beta_0 + \sum_{i=1}^p \beta_i x_{k,i}}}{1 + e^{\beta_0 + \sum_{i=1}^p \beta_i x_{k,i}}} \right)^{y_k} \left(\frac{1}{1 + e^{\beta_0 + \sum_{i=1}^p \beta_i x_{k,i}}} \right)^{1-y_k}$$

+ Penalização



Amazon

