



escola
britânica de
artes criativas
& tecnologia

Profissão: Cientista de Dados

A Maldição da Dimensionalidade



Fenômeno de Huges

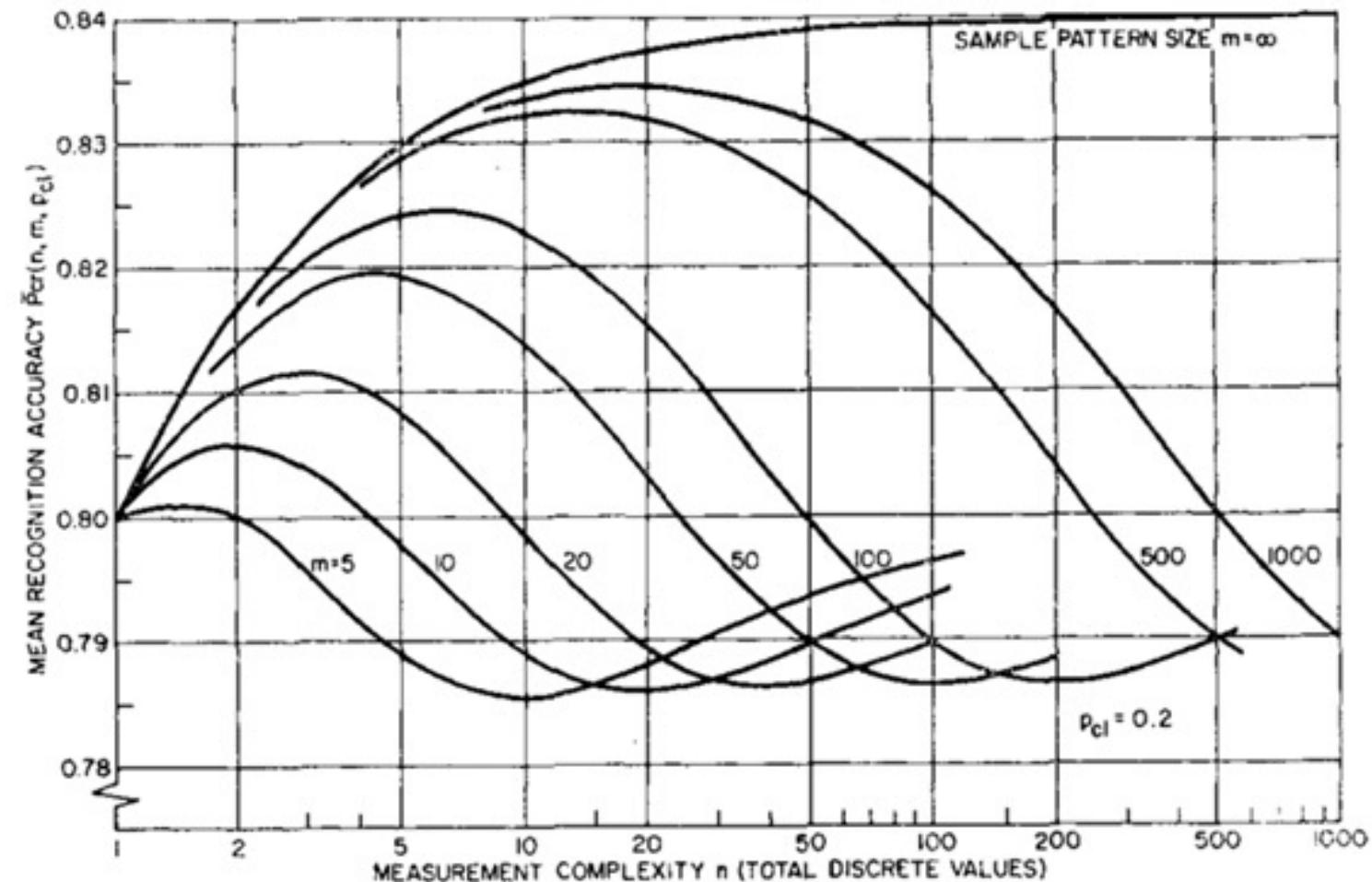


Fig. 4. Finite data set accuracy ($p_{c1} = \frac{1}{5}$).

Redução de dimensionalidade



Redução de dimensionalidade

Reducir o número
de variáveis
substancialmente

Perdendo o
mínimo possível
de informação

Pra que usa?

- Visualização de dados
- Eficiência em algoritmos de *machine learning*
- Ordenação de indivíduos
- Eficiência em transmissões (imagem áudio etc)

Exemplos classicos



escola
britânica de
artes criativas
& tecnologia

Profissão: Cientista de Dados

Análise de Componentes Principais

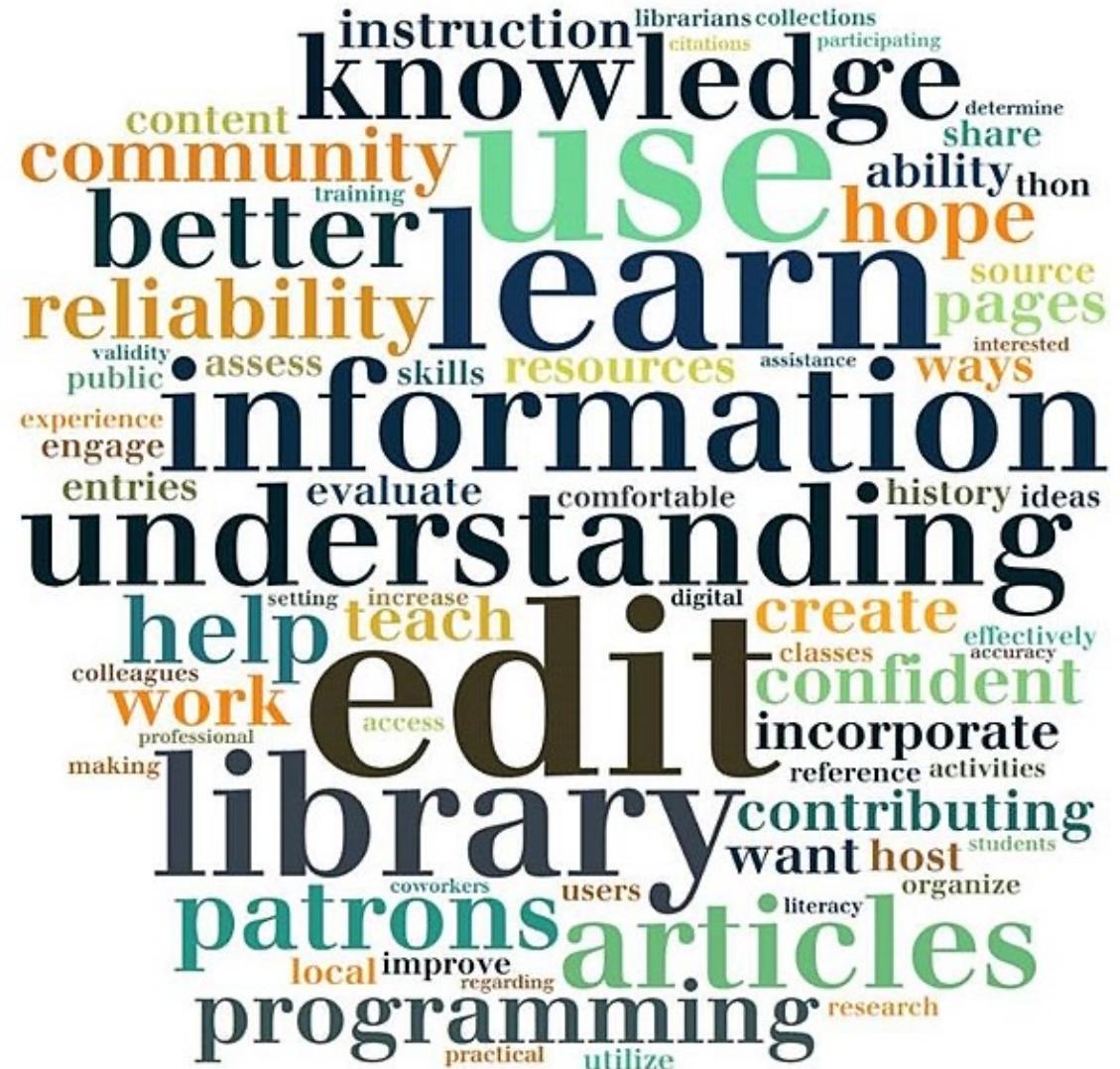
Reconhecimento de imagens





Dados de
acelerômetro e
giroscópio

Processamento de linguagem natural



instruction librarians collections citations participating
knowledge determine share
content community training ability than
better reliability source pages
experience engage skills resources assistance interested ways
entries evaluate comfortable history ideas
evaluate setting increase digital
colleagues teach create effectively accuracy
work professional access classes activities
making digital contributing students
library patrons users literacy organize
articles local improve regarding research
programming practical utilize

Situações

- Machine learning
- Interpretação secundária
- Muitas variáveis
- Muitos dados vs recursos

Handwritten mathematical notes and diagrams on a blackboard:

- $P_1 = 60000$
- $18.50 + 63.50 \cdot 1.1 = x^2 + 4.1132$
- $260 = 5.63 \cdot 113.4$
- $c = 552$
- $D(x) = 2 + 3 + 4.31447$
- $\text{P} \rightarrow \text{H} = 5.45250$
- $\text{V} = 22$
- $\text{C}(x, y) \begin{cases} xy = c \\ cx - cy = 33^2 \\ \pi = c \end{cases}$
- $A \cap B, \frac{24+x}{y} + \frac{2^2+3^2}{c} + \frac{x}{2} = 9$
- $\text{men} = 584. + n^{30} (x^2 + 34x + c)$
- $x = 9.25$
- $\sum_{x=2}^{u=14} N_{50} \cdot x - \frac{1}{2} [964 + xg + \rho]$
- $\beta = 9 + x^2 + y + \rho$
- $\begin{bmatrix} 010112 \\ 010002 \\ 111002 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix}$

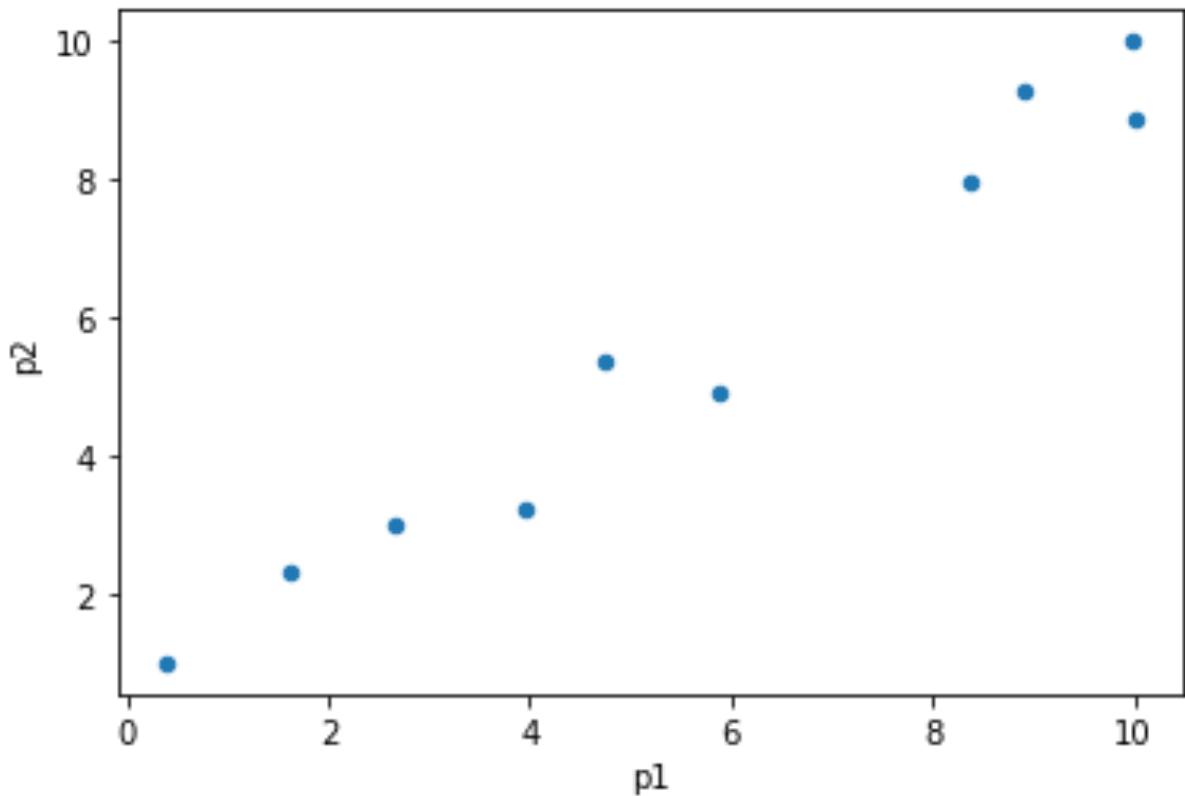
Objetivos

- Conhecer a técnica
- Saber utilizar com o *Scikit Learn*
- Saber decidir se ela é adequada
- Conhecer os elementos e passos básicos

Matéria prima

Covariância - correlação

Correlação e covariância



$$Cov(X, Y) = \sigma_{x,y} = \frac{1}{N} \sum_{i=1}^N (X - \bar{X})(Y - \bar{Y})$$

$$Cor(X, Y) = \rho_{x,y} = \frac{Cov(X, Y)}{Var(X)Var(Y)}$$

Correlação e covariância

$$Cor(X, Y) = \rho_{x,y} = \frac{Cov(X, Y)}{Var(X)Var(Y)}$$

$$x_{pad} = \frac{x - \hat{x}}{DesvPad(x)}$$

Matriz de covariância e de correlação

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix} \quad R = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix}$$