



escola
britânica de
artes criativas
& tecnologia

Profissão: Cientista de Dados

A Maldição da Dimensionalidade



Fenômeno de Huges

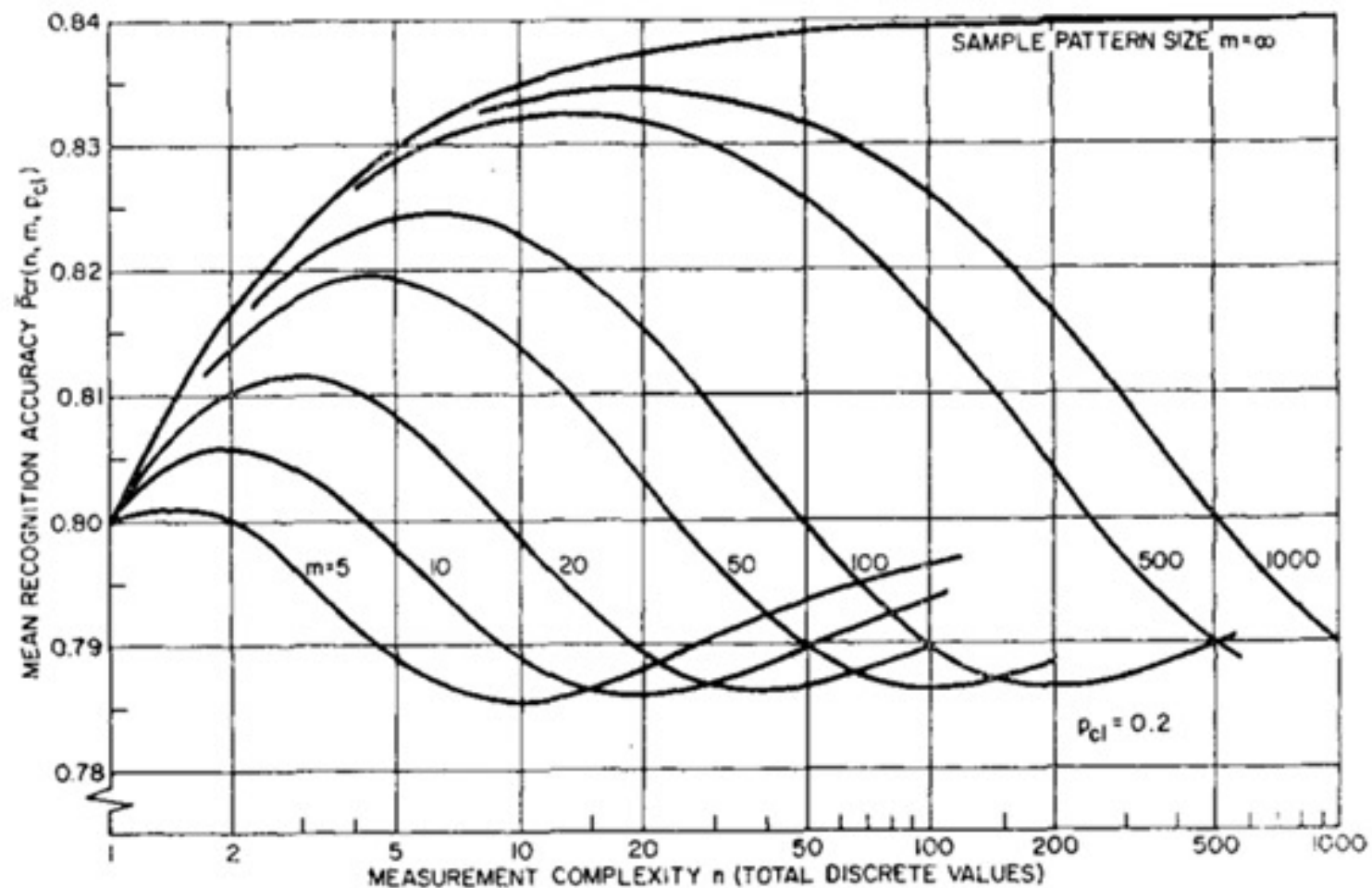
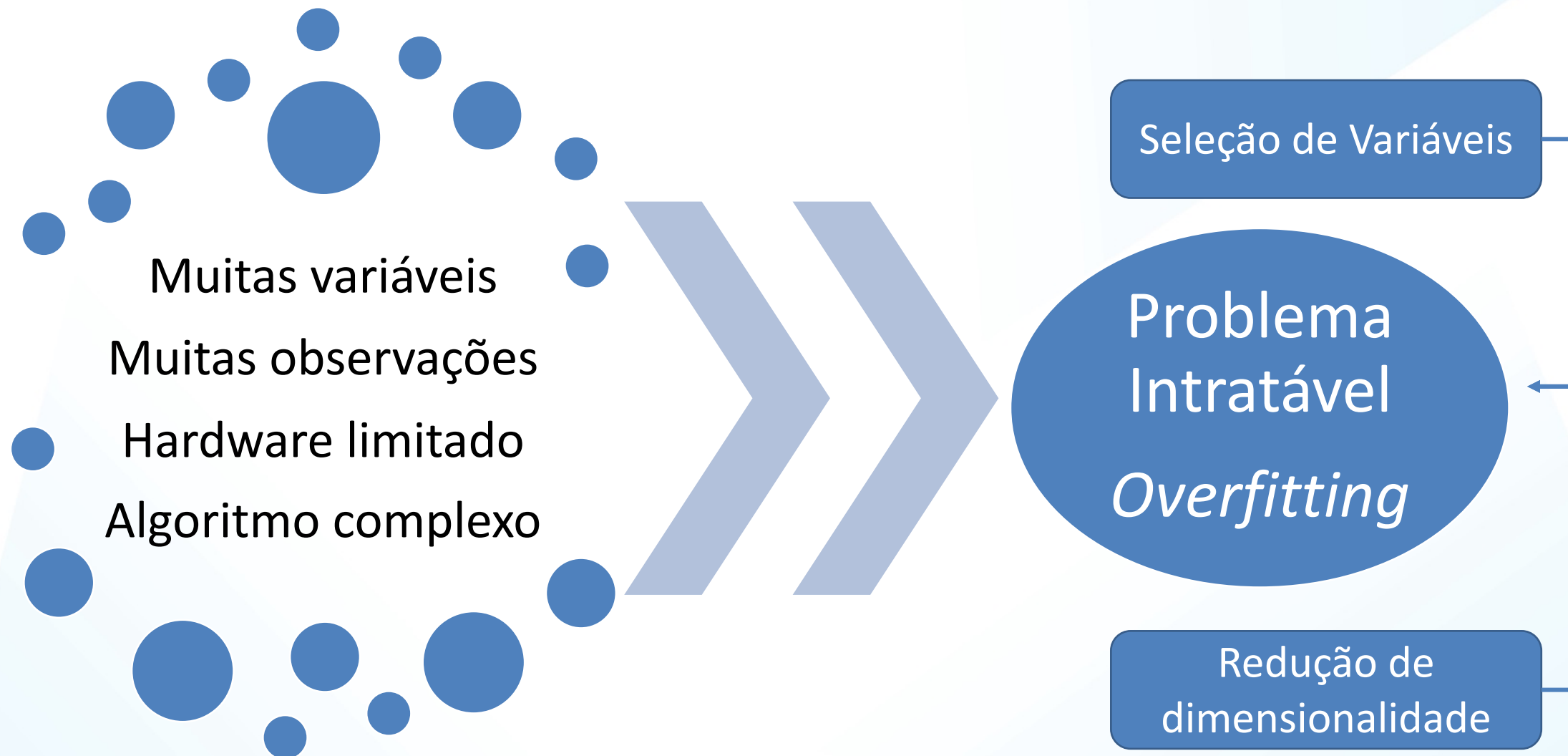


Fig. 4. Finite data set accuracy ($p_{cl} = \frac{1}{5}$).

Redução de dimensionalidade



Redução de dimensionalidade

Reduzir 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 clássicos



escola
britânica de
artes criativas
& tecnologia

Profissão: Cientista de Dados

Análise de Componentes Principais



Reconhecimento de imagens





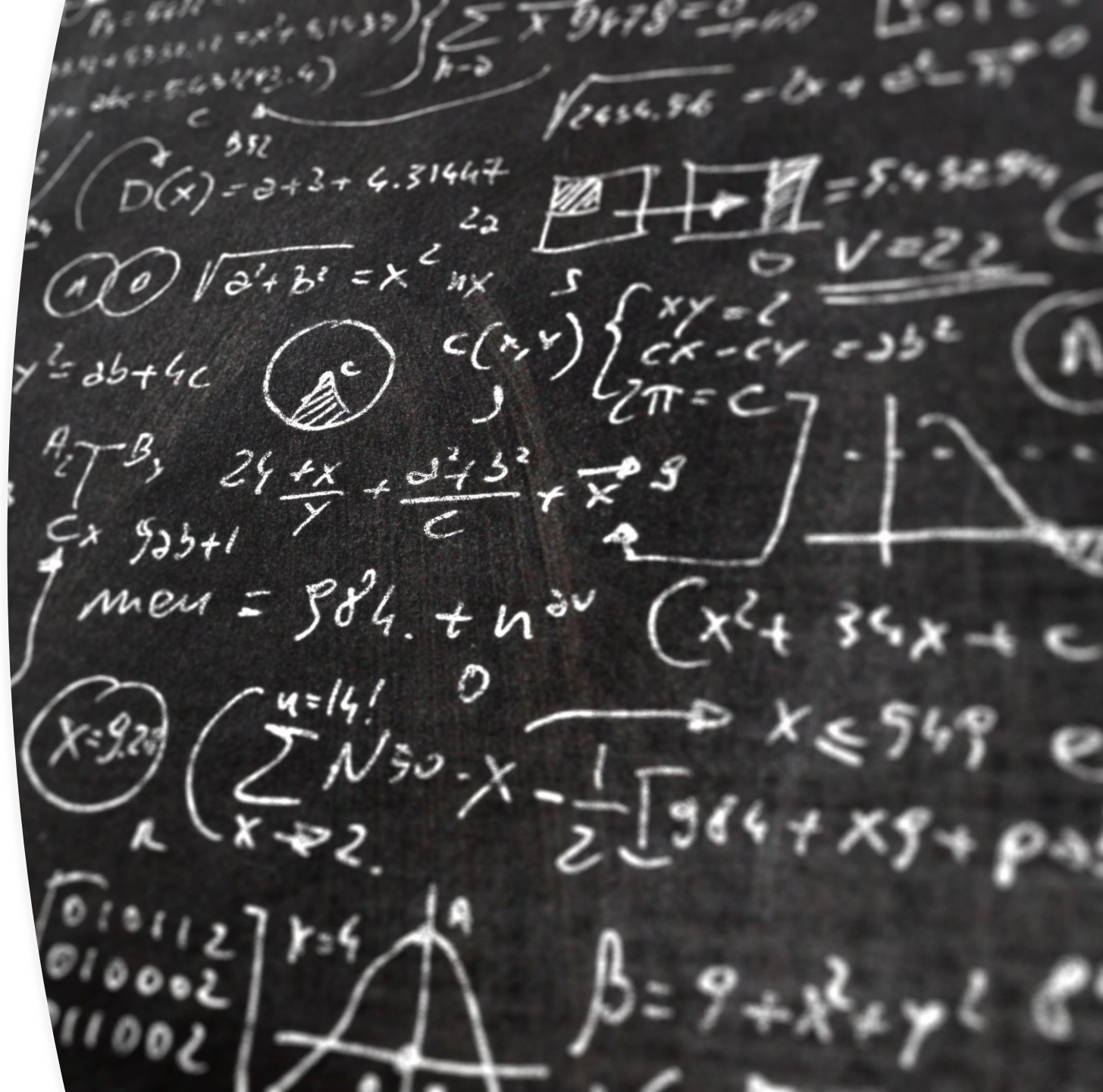
Dados de
acelerômetro e
giroscópio

Processamento de linguagem natural



Situações

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



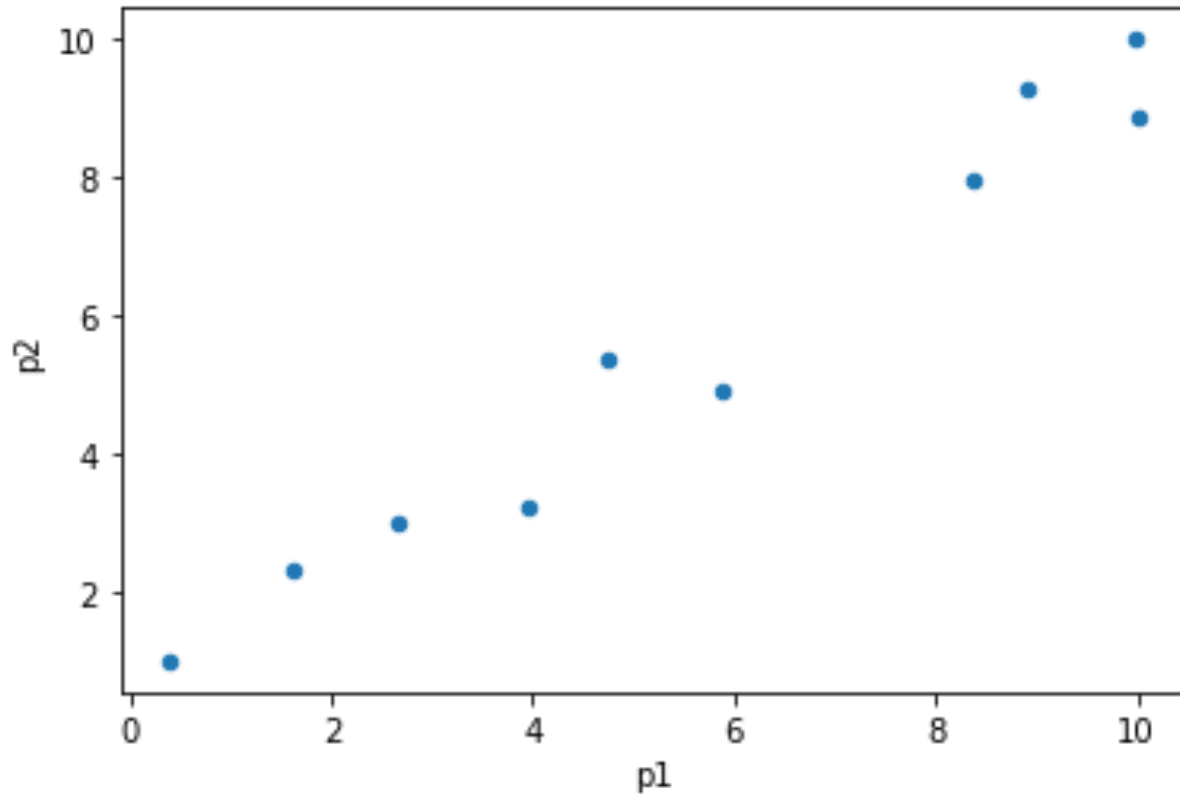
Objetivos

- 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

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

$$R = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix}$$