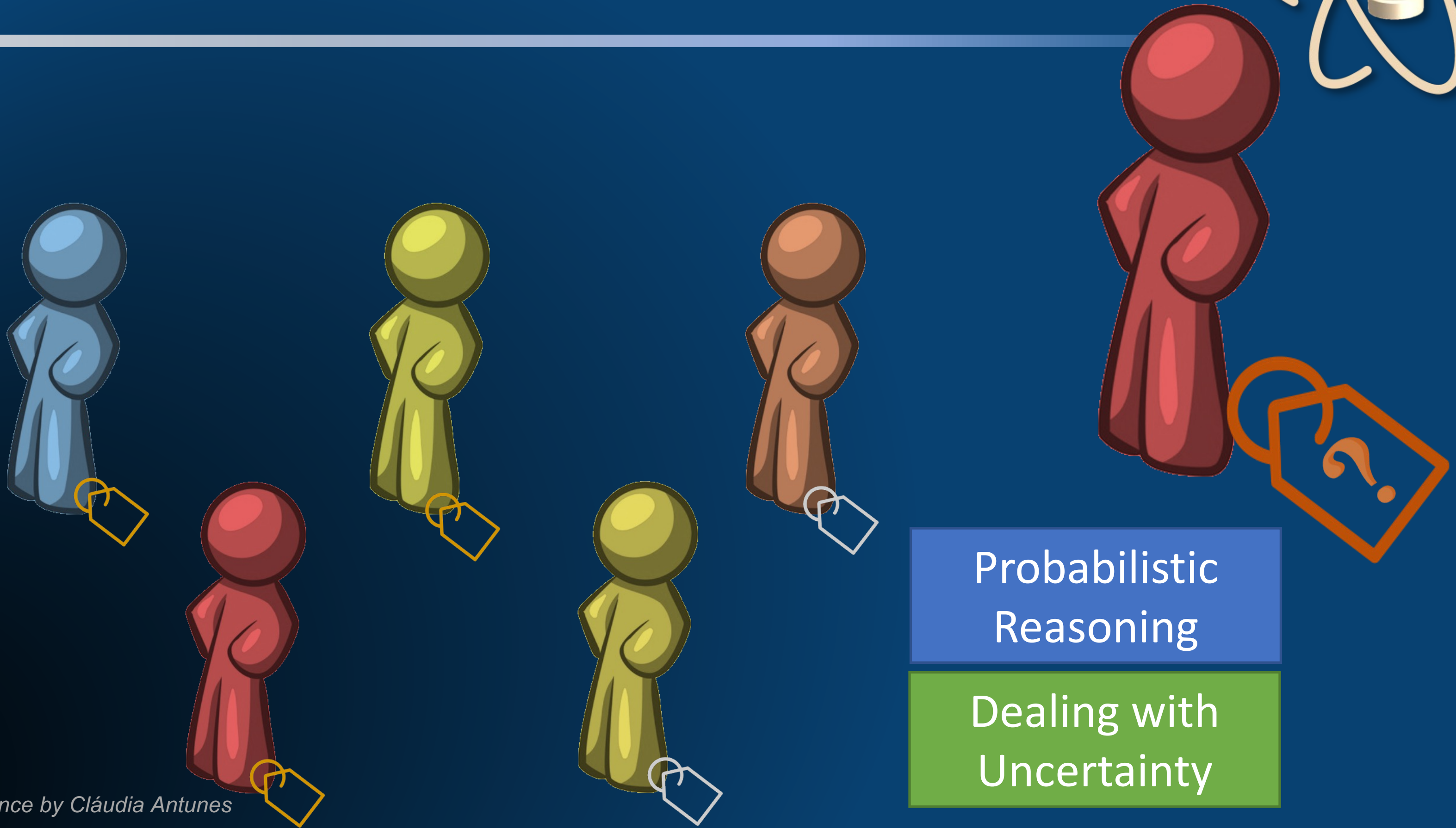
Bayesians

Cláudia Antunes

Instituto Superior Técnico – Universidade de Lisboa

Bayesian classifier



Probabilistic Reasoning

Dealing with Uncertainty

MAP CLASSIFIER

$$\boldsymbol{x} \rightarrow \boldsymbol{y}$$

$$\hat{y} = \arg\max_{c_i} \{P(c_i|\boldsymbol{x})\}$$

$$= \arg\max_{c_i} \left\{ \frac{P(c_i) \times P(\boldsymbol{x}|c_i)}{P(\boldsymbol{x})} \right\}$$

MAP CLASSIFIER

$$\boldsymbol{x} \rightarrow \boldsymbol{y}$$

$$\hat{y} = \underset{c_i}{\arg\max}\{P(c_i|\boldsymbol{x})\}$$

prior probability
for each class

$$= \underset{c_i}{\arg\max}\left\{\frac{P(c_i) \times P(\boldsymbol{x}|c_i)}{P(\boldsymbol{x})}\right\}$$

MAP CLASSIFIER

$$\boldsymbol{x} \rightarrow \boldsymbol{y}$$

$$\hat{y} = \arg\max_{c_i}\{P(c_i|\boldsymbol{x})\}$$

$$= \arg\max_{c_i}\left\{\frac{P(c_i)\times P(\boldsymbol{x}|c_i)}{P(\boldsymbol{x})}\right\}$$

Likelihood of x belonging to each class

MAP CLASSIFIER

$$\boldsymbol{x} \rightarrow \boldsymbol{y}$$

$$\hat{y} = \underset{c_i}{\arg\max}\{P(c_i|\boldsymbol{x})\}$$

$$= \underset{c_i}{\arg\max}\left\{\frac{P(c_i) \times P(\boldsymbol{x}|c_i)}{P(\boldsymbol{x})}\right\}$$

probability of x
being observed

MAP CLASSIFIER

$$\boldsymbol{x} \rightarrow \boldsymbol{y}$$

$$\hat{y} = \arg\max_{c_i}\{P(c_i|\boldsymbol{x})\}$$

$$= \arg\max_{c_i}\left\{\frac{P(c_i)\times P(\boldsymbol{x}|c_i)}{P(\boldsymbol{x})}\right\}$$

$$\hat{\boldsymbol{y}} = \arg\max_{c_i}\{P(c_i)\times P(x|c_i)\}$$

Bayesian Classifiers

Records represented as **tuples of d values**

Training algorithm

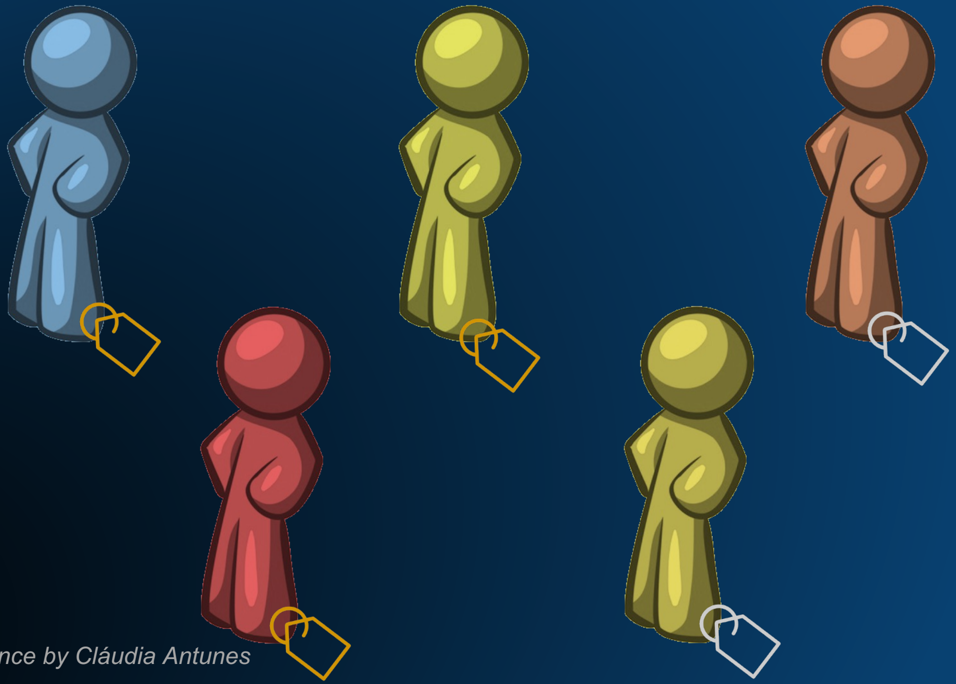
Compute prior probabilities for each class

Classification algorithm

For each Z to be classified

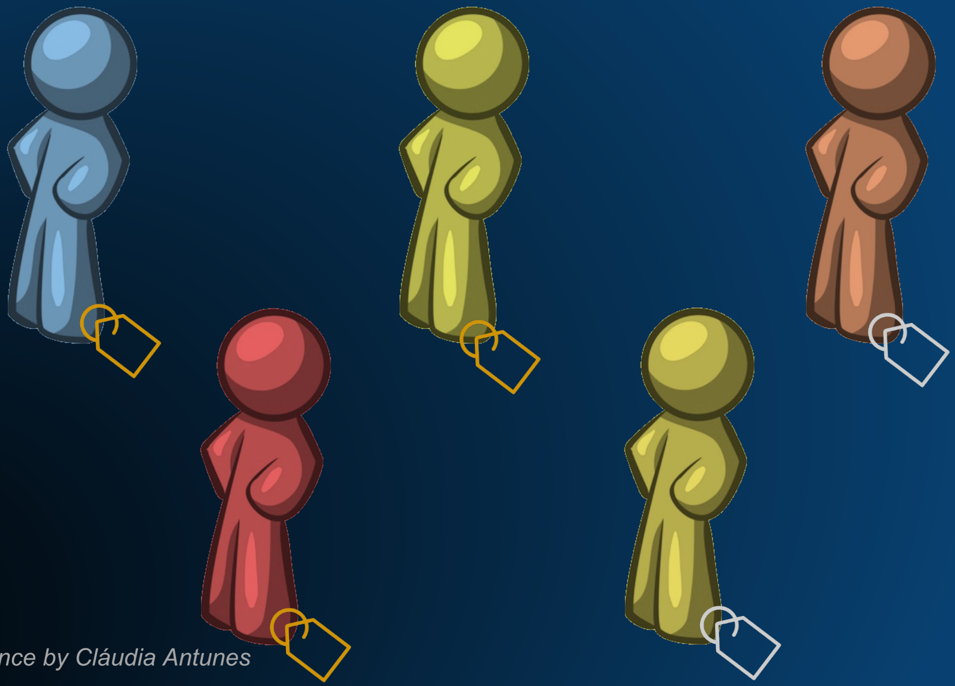
- Estimate likelihood for z given each class
- Classify Z in the **most probable class**

Estimation of prior probabilities



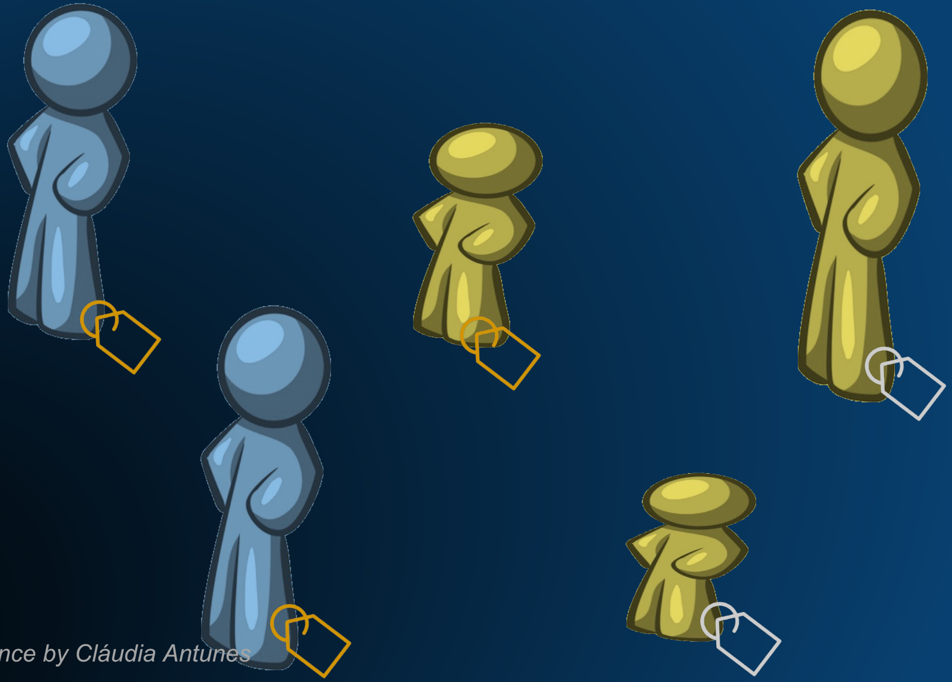
$$\hat{P}(c_i) = \frac{n_i}{n}$$

Estimation of likelihood



$$\hat{P}(x|c_i) = \frac{n_{x|i}}{n_i}$$

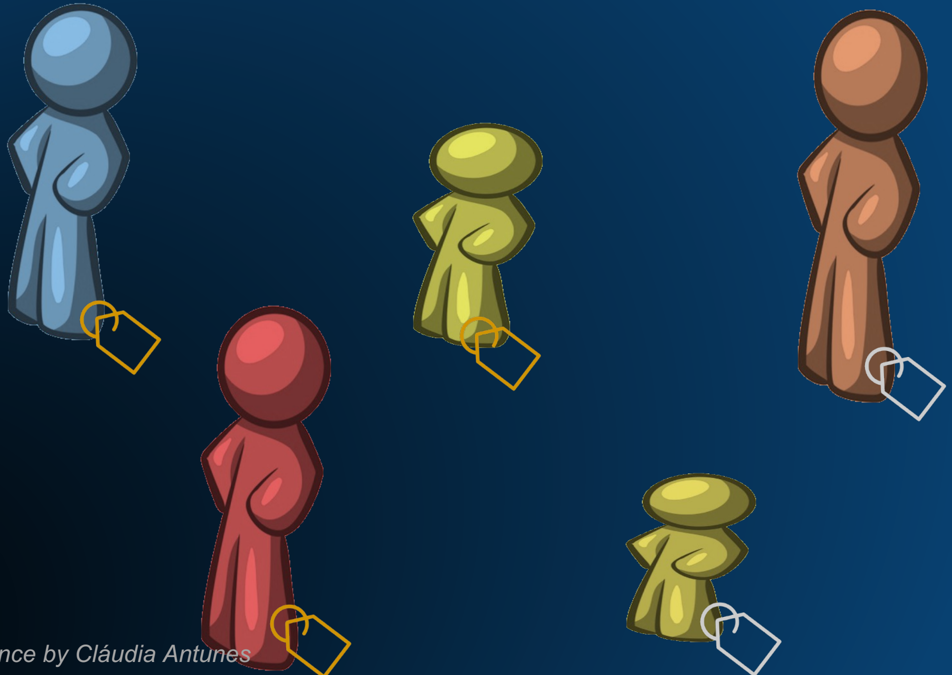
Probabilities estimation



$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$$

$$\hat{P}(x|c_i) = f_i(x|\mu_i, \sigma_i)$$

Joint probabilities



$$\vec{X} \sim \mathcal{N}(\vec{\mu}, \Sigma^2)$$

$$\hat{P}(\vec{x}|c_i) = f_i(\vec{x}|\vec{\mu}_i, \Sigma_i)$$

$$\hat{P}(x_1 \wedge \cdots x_d|c_i) = f_i(x_1, \ldots, x_d|\vec{\mu}_i, \Sigma_i)$$

Naïve Bayes

Naïve Bayes assumption

"All variables describing the data are conditionally independent"

$$V_1, \dots, V_d \; are \; independent$$

$$\leftrightarrow$$

$$P(V_1 = v_1 \land \cdots \land V_d = v_d) = \prod_{j=1}^{d} P(V_j = v_j)$$

NAÏVE BAYES ALGORITHM

$$\boldsymbol{x} \rightarrow \boldsymbol{y}$$

$$\hat{y} = \arg\max_{c_i}\{P(c_i|\boldsymbol{x})\}$$

$$\hat{\boldsymbol{y}} = \arg\max_{\boldsymbol{c_i}}\{\boldsymbol{P(c_i)\times P(x|c_i)}\}$$

$$\hat{\boldsymbol{y}} = \boldsymbol{arg\,max}_{\boldsymbol{c_i}}\left\{\boldsymbol{P(c_i)\times\prod_{j=1}^{d}P(x_j|c_i)}\right\}$$

NAÏVE BAYES ALGORITHM

Records represented as **tuples of d values**

Training algorithm

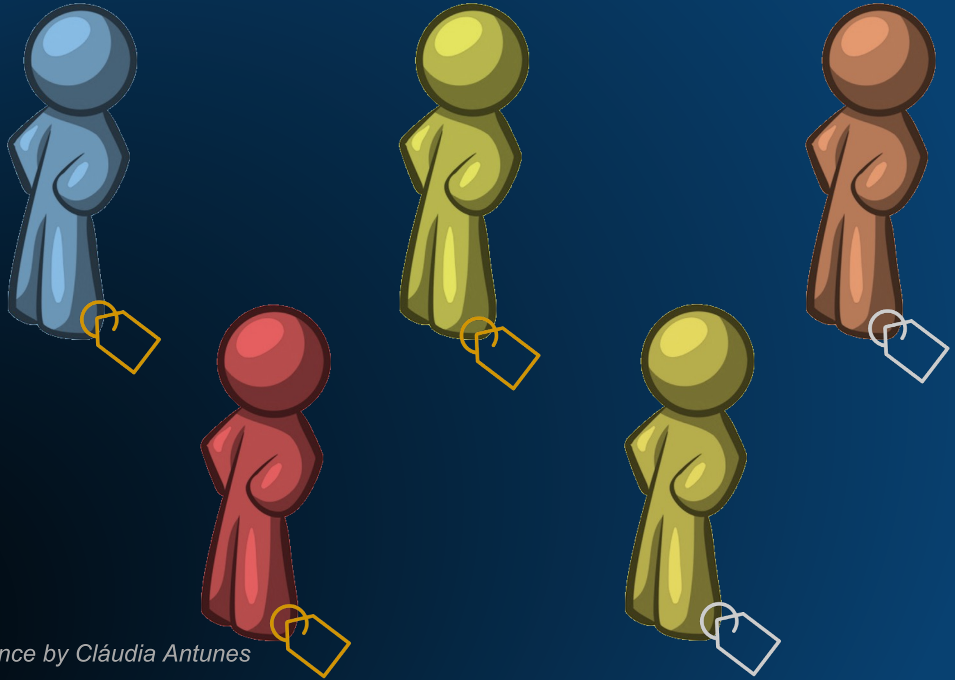
Compute prior probabilities for each class

Classification algorithm

For each Z to be classified

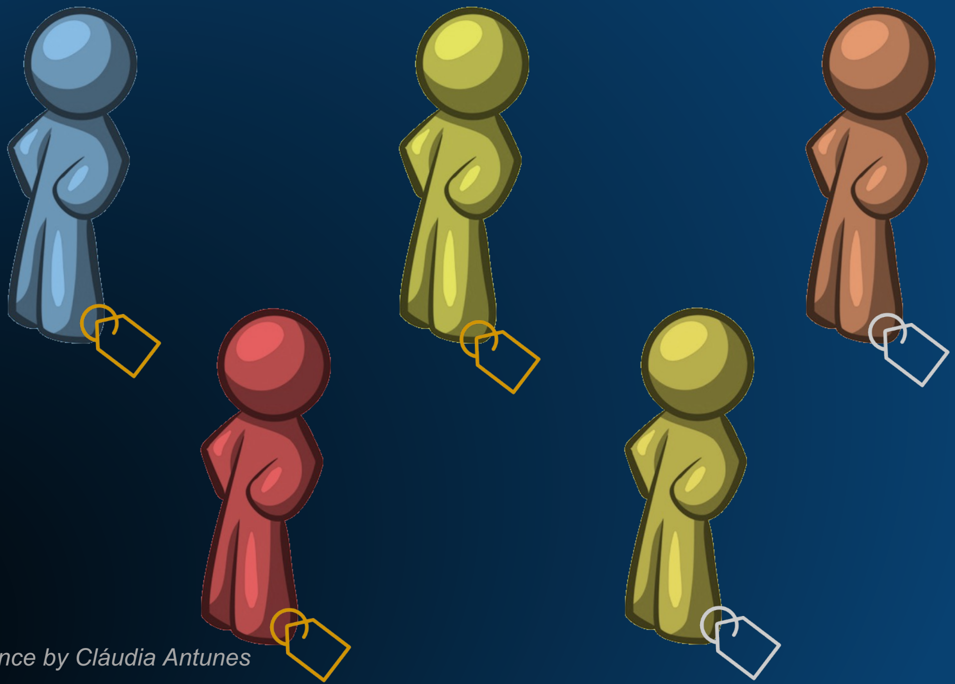
- Estimate likelihood for Z for all its **d** dimensions given each class
- Classify Z in the **most probable class**

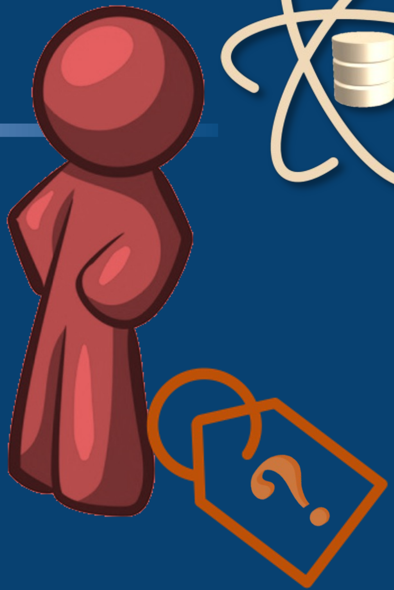
Estimation of prior probabilities



$$\hat{P}(c_i) = \frac{n_i}{n}$$

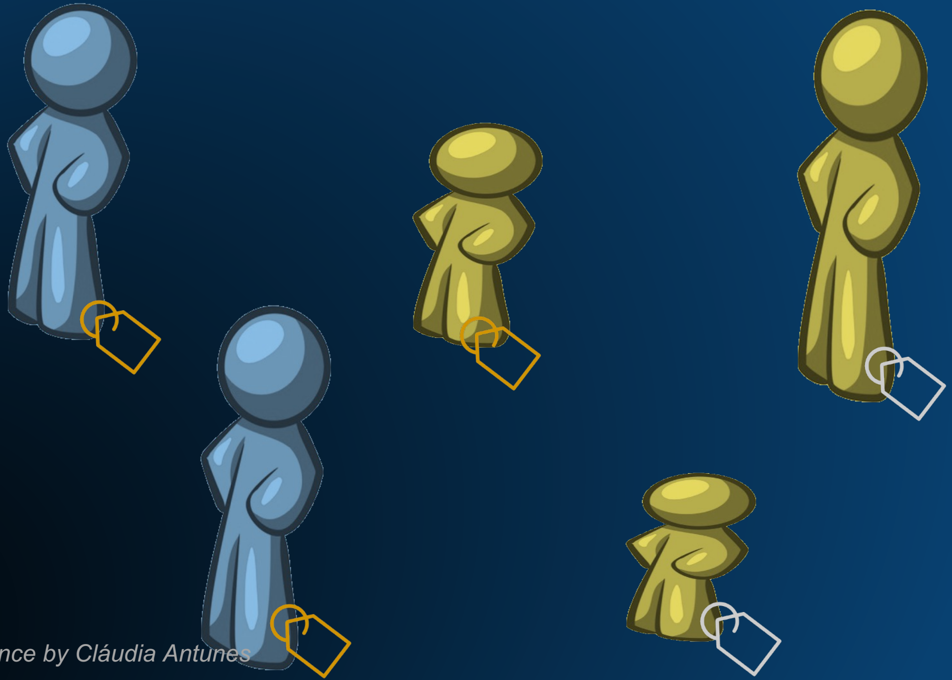
Estimation of likelihood

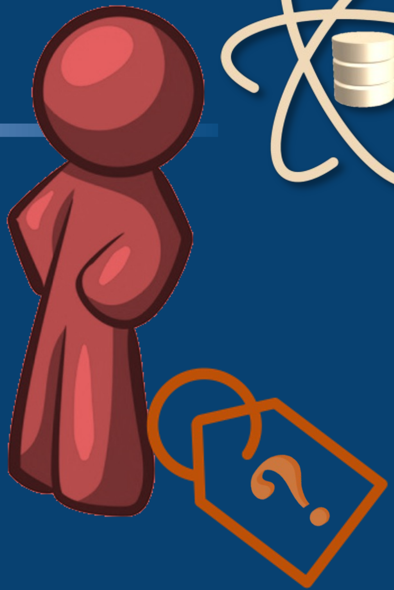




$$\hat{P}(x|c_i) = \frac{n_{x|i}}{n_i}$$

Probabilities estimation

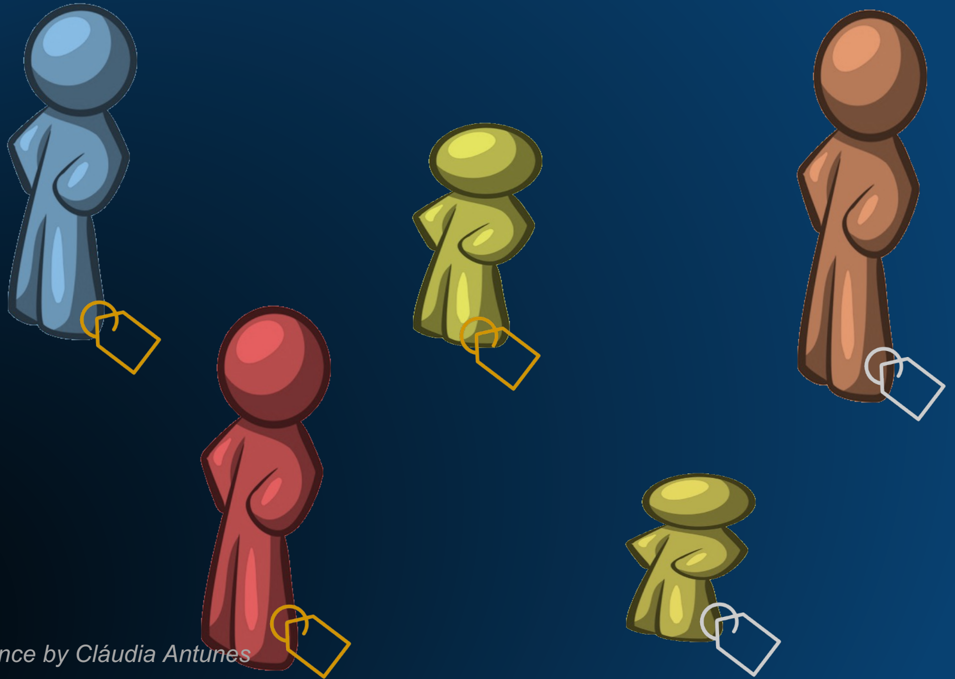


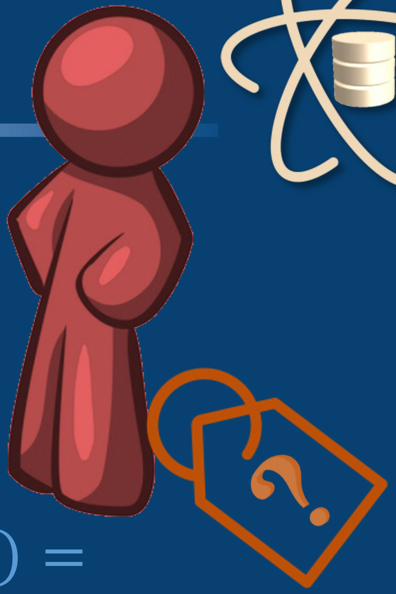


$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$$

$$\hat{P}(x|c_i) = f_i(x|\mu_i, \sigma_i)$$

Joint probabilities



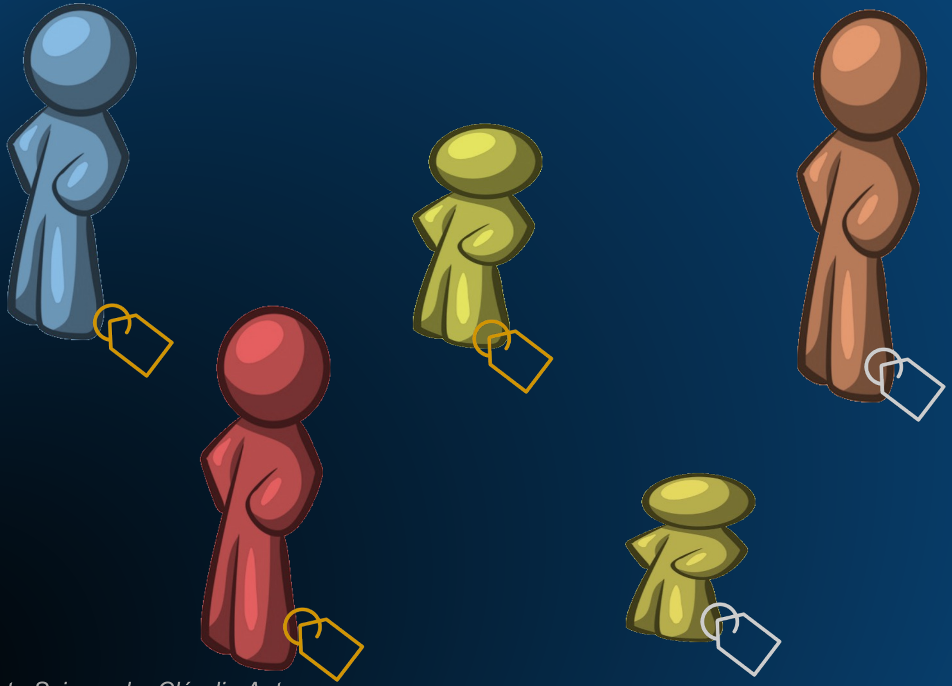


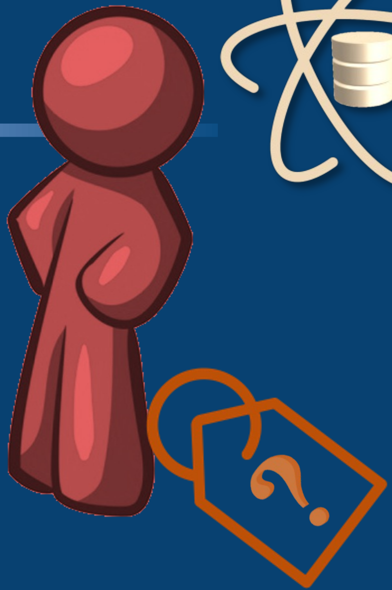
$$\widehat{P}(\vec{x}|c_i) =$$

$$\widehat{P}(x_1 \wedge \cdots \wedge x_d | c_i)$$

$$= \prod_{j=1}^{d} P(x_j|c_i)$$

JOINT PROBABILITIES



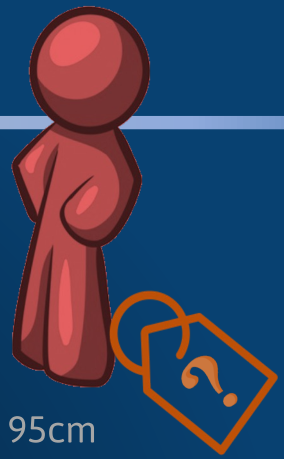


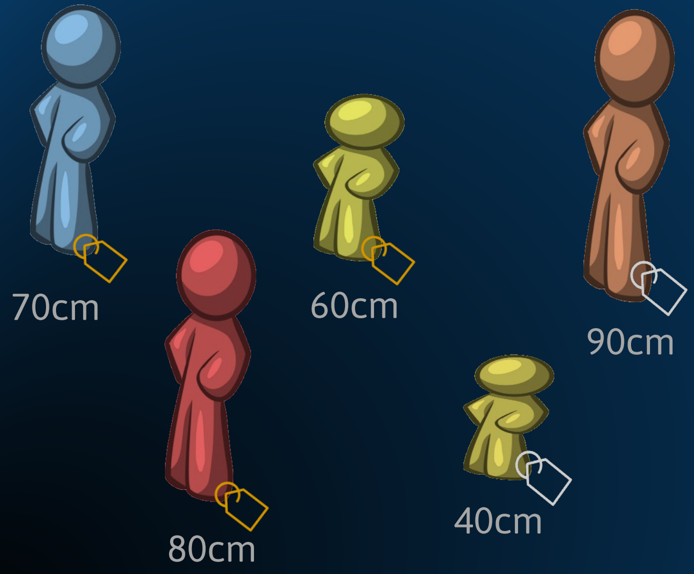
$$\vec{X} \sim \mathcal{N}(\vec{\mu}, \Sigma^2)$$

$$\hat{P}(\vec{x}|c_i) = f_i(\vec{x}|\vec{\mu}_i, \Sigma_i)$$

$$\hat{P}(x_1 \wedge \cdots x_d|c_i) = f_i(x_1, \ldots, x_d|\vec{\mu}_i, \Sigma_i)$$

Example

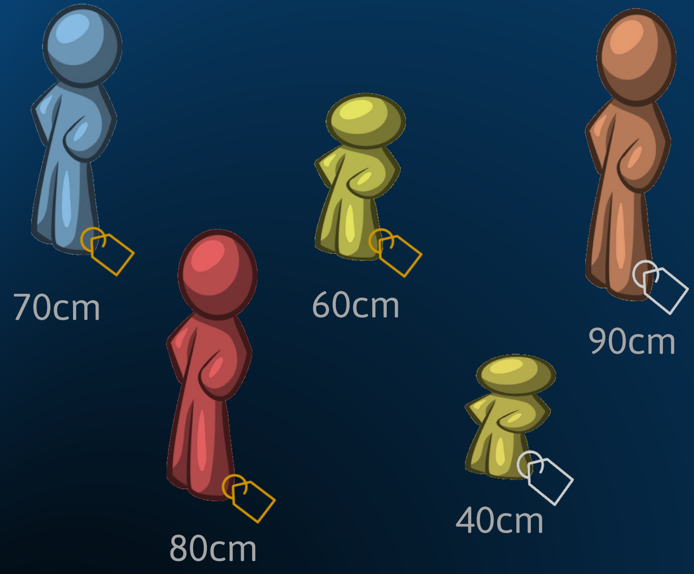


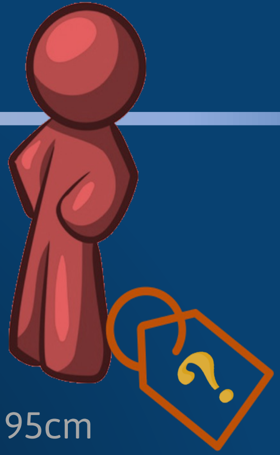


$$\hat{y} = \arg\max_{c_i}\{P(c_i|\boldsymbol{x})\}$$

$$\hat{y} = \boldsymbol{arg\,max}_{c_i}\{\boldsymbol{P(c_i)\times P(red|c_i)\times P(95|c_i)}\}$$

Example





$$\hat{P}(gold) = \frac{3}{5}$$

$$\hat{P}(red \wedge 95|gold) =$$

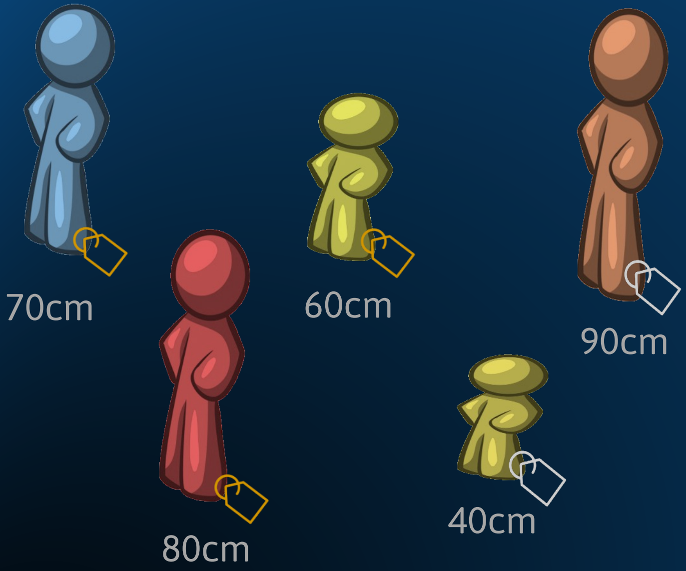
$$= \hat{P}(red|gold) \times \hat{P}(95|gold)$$

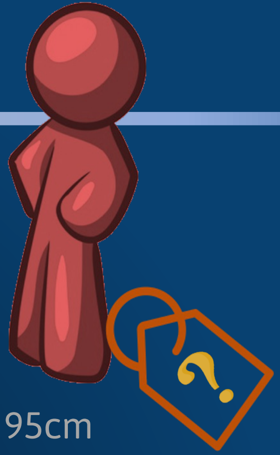
$$= \frac{1}{3} \times f_{gold}(95)$$

$$gold \sim \mathcal{N}(\mu = 70cm, \sigma^2 = 10^2)$$

$$f_{gold}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Example





$$\hat{P}(silver) = \frac{2}{5}$$

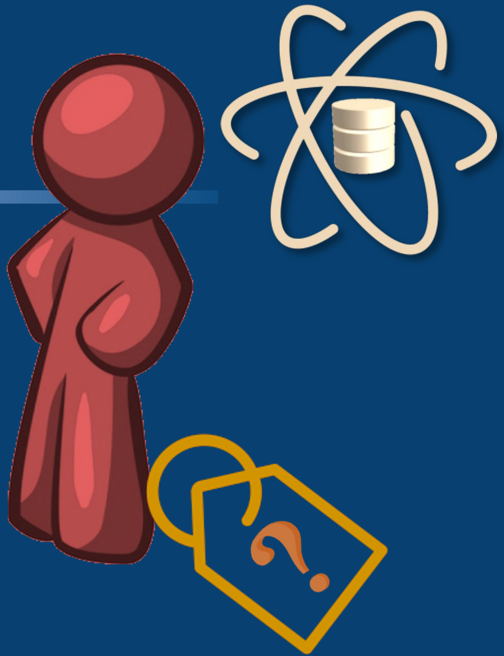
$$\hat{P}(red \wedge 95 | silver) =$$

$$= \hat{P}(red|silver) \times \hat{P}(95|silver)$$
$$= 0 \times f_{silver}(95)$$
$$= 0$$

$$silver \sim \mathcal{N}(\mu = 65cm, \sigma^2 = 35^2)$$

$$f_{silver}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

EXAMPLE



$$\hat{y} = \underset{c_i}{\arg\max}\{P(c_i|\boldsymbol{x})\}$$

$$\hat{\boldsymbol{y}} = \underset{c_i}{\boldsymbol{arg\,max}}\{\boldsymbol{P(c_i)\times P(red|c_i)\times P(95|c_i)}\}$$

$$\hat{\boldsymbol{y}} = \underset{\{gold,silver\}}{\boldsymbol{arg\,max}}\left\{\begin{matrix}\boldsymbol{P(gold)\times P(red|gold)\times P(95|gold),}\\ \boldsymbol{P(silver)\times P(red|silver)\times P(95|silver)}\end{matrix}\right\}$$

$$\hat{\boldsymbol{y}} = \underset{\{gold,silver\}}{\boldsymbol{arg\,max}}\left\{\boldsymbol{\frac{3}{5}\times\frac{1}{3}\times P(95|gold),\frac{2}{5}\times 0\times P(95|silver)}\right\}$$

Thank you!

claudia.antunes@tecnico.ulisboa.pt