

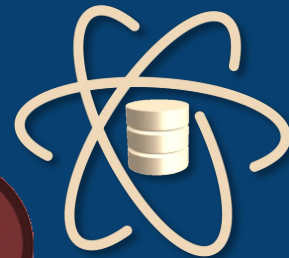
A hand is shown in the foreground, reaching out towards a complex, futuristic digital interface. The interface features a large circular gauge with multiple concentric rings, some of which are illuminated with green and blue light. Inside the gauge, there are several interlocking gears. The background is dark and filled with various digital elements, including lines, dots, and abstract shapes, suggesting a high-tech or artificial intelligence environment.

Bayesians

Cláudia Antunes

Instituto Superior Técnico – Universidade de Lisboa

BAYESIAN CLASSIFIER



Probabilistic
Reasoning

Dealing with
Uncertainty

MAP CLASSIFIER



$$\mathbf{x} \rightarrow \mathbf{y}$$

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

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

MAP CLASSIFIER



$$\mathbf{x} \rightarrow \mathbf{y}$$

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

prior probability
for each class

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

MAP CLASSIFIER



$$\mathbf{x} \rightarrow \mathbf{y}$$

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

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

Likelihood of \mathbf{x}
belonging to each class

MAP CLASSIFIER



$$\mathbf{x} \rightarrow \mathbf{y}$$

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

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

probability of \mathbf{x}
being observed

MAP CLASSIFIER



$$\mathbf{x} \rightarrow \mathbf{y}$$

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

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

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

BAYESIAN CLASSIFIERS



Records represented as **tuples of d values**

Training algorithm

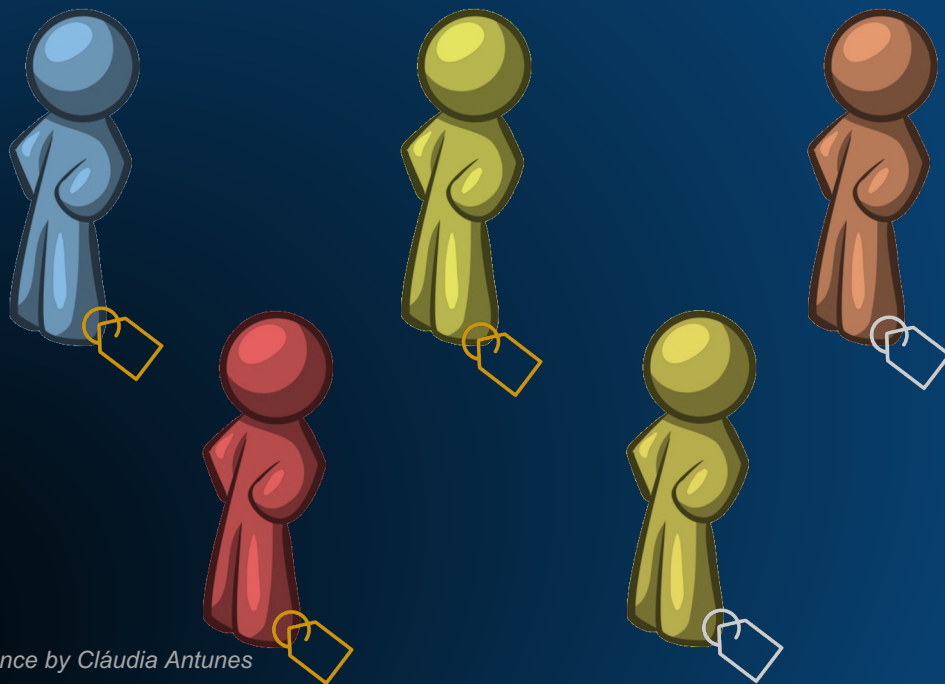
Compute **prior probabilities** for each class

Classification algorithm

For each **Z** to be classified

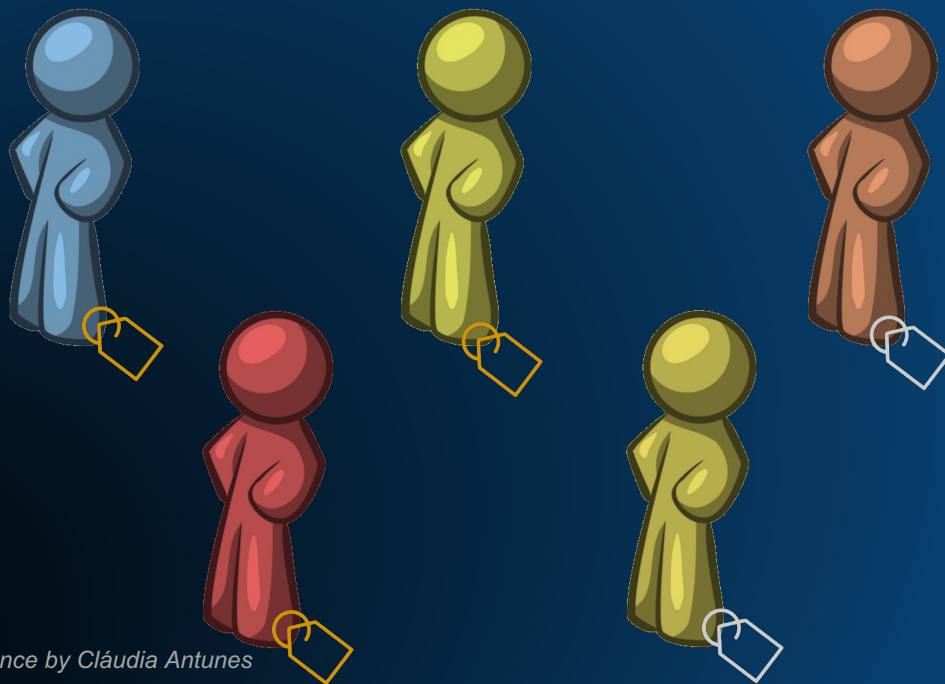
- Estimate **likelihood** for **z** given each class
- Classify **Z** in the **most probable class**

ESTIMATION OF PRIOR PROBABILITIES



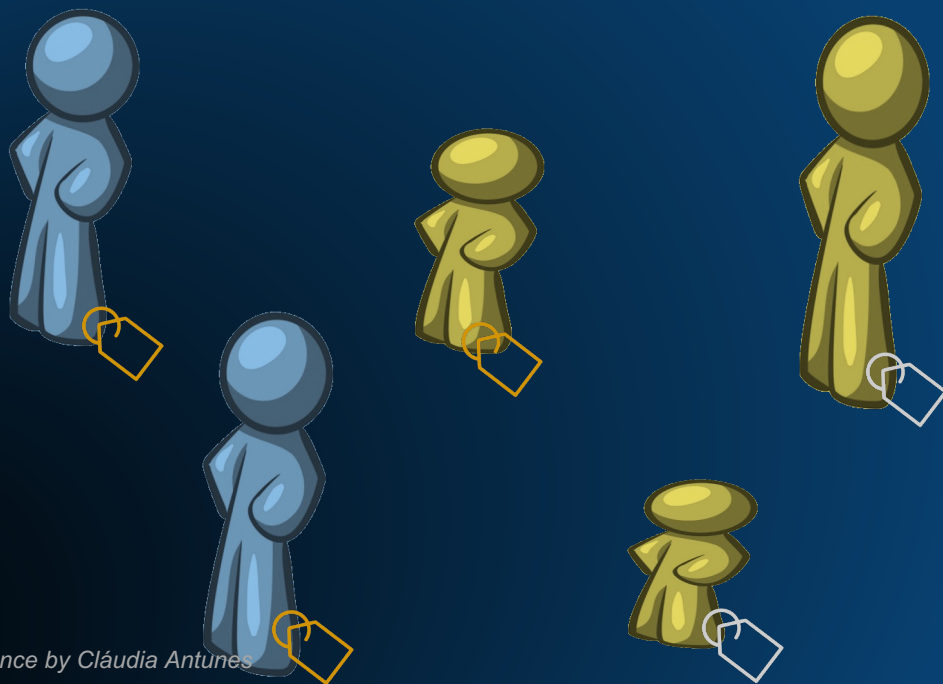
$$\hat{P}(c_i) = \frac{n_i}{n}$$

ESTIMATION OF LIKELIHOOD



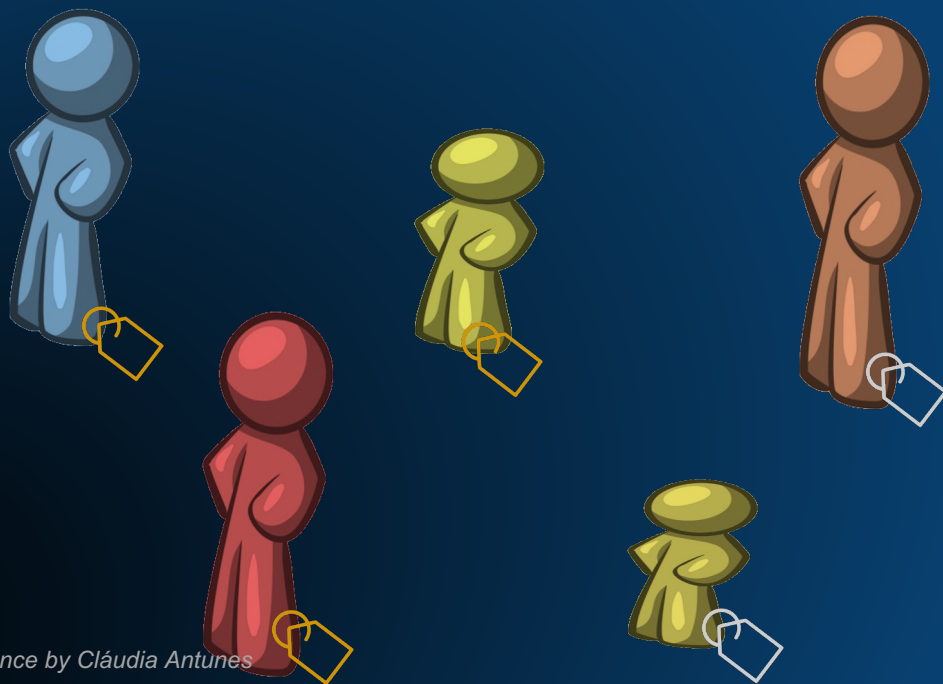
$$\hat{P}(x|c_i) = \frac{n_{x|i}}{n_i}$$

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



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

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

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



Naïve Bayes

NAÏVE BAYES ASSUMPTION



“All variables describing the data
are conditionally independent”

V_1, \dots, V_d *are independent*



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

NAÏVE BAYES ALGORITHM



$$\mathbf{x} \rightarrow \mathbf{y}$$

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

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

$$\hat{\mathbf{y}} = \arg \max_{c_i} \left\{ P(c_i) \times \prod_{j=1}^d P(x_j | c_i) \right\}$$

NAÏVE BAYES ALGORITHM



Records represented as **tuples of d values**

Training algorithm

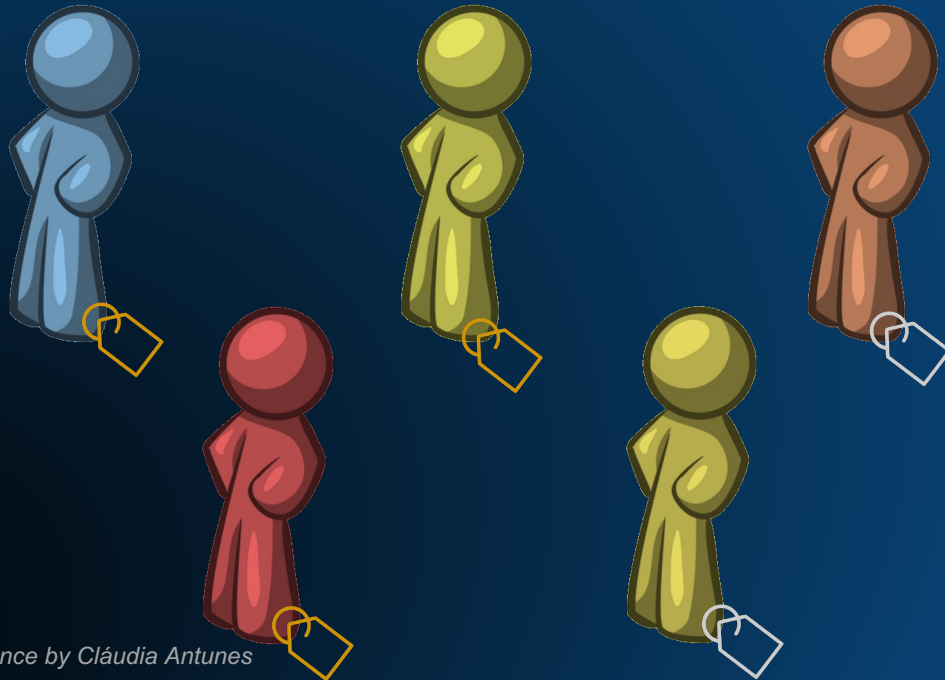
Compute **prior probabilities** for each class

Classification algorithm

For each **Z** to be classified

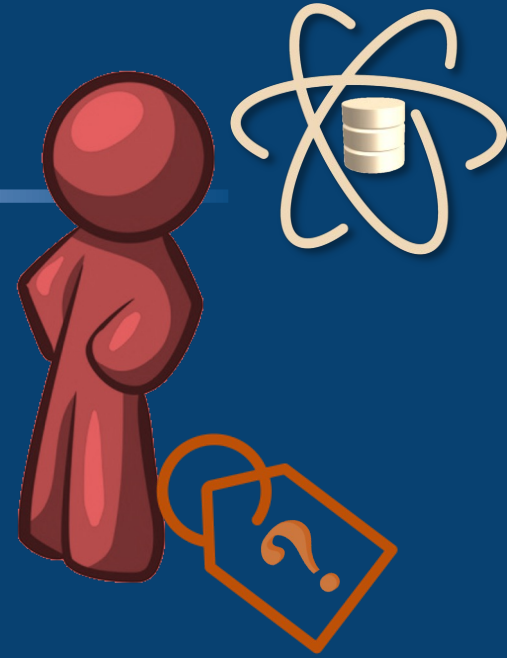
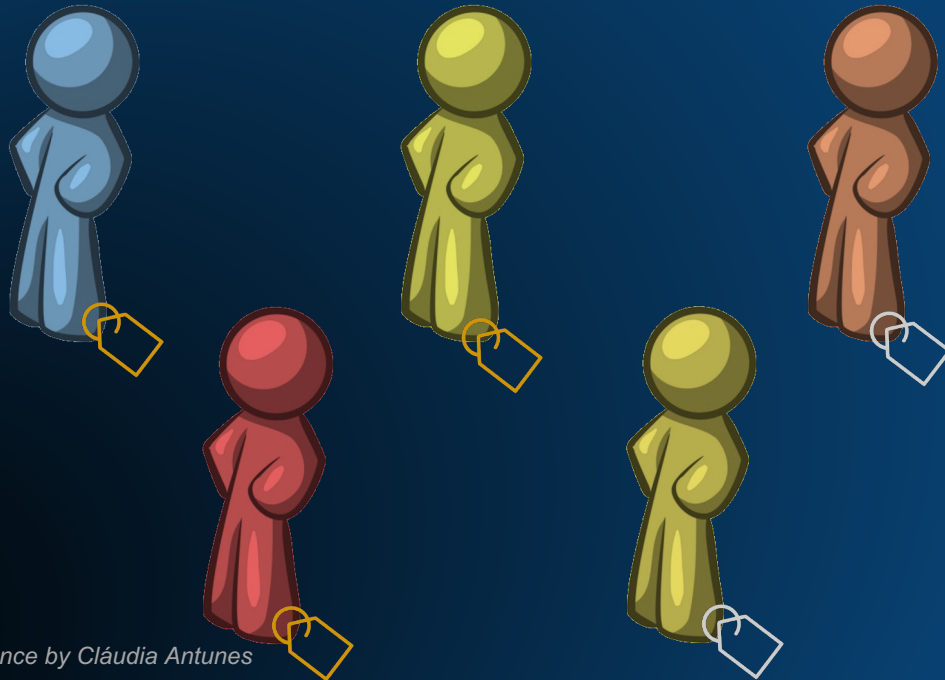
- Estimate **likelihood** for **Z** for all its **d dimensions** given each **class**
- Classify **Z** in the **most probable class**

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



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

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

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



JOINT PROBABILITIES



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

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

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

EXAMPLE



70cm



80cm



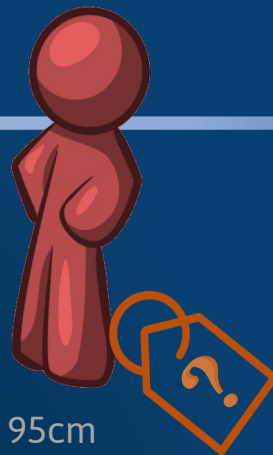
60cm



40cm



90cm

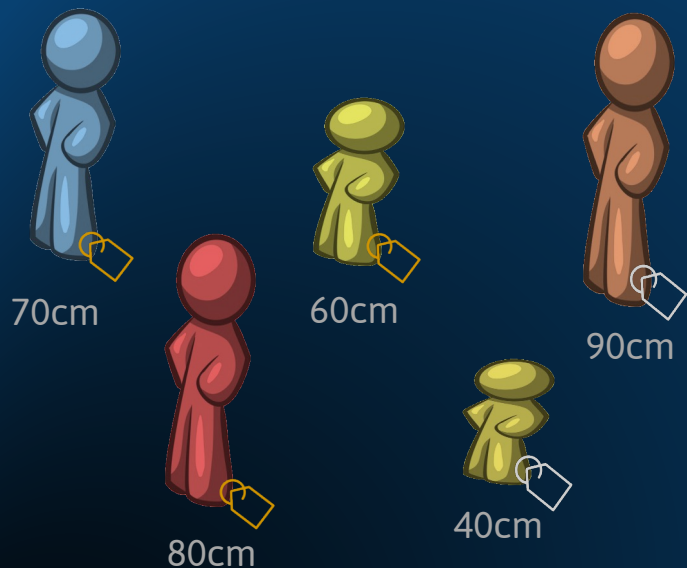


95cm

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

$$\hat{y} = \arg \max_{c_i} \{P(c_i) \times P(\text{red}|c_i) \times P(95|c_i)\}$$

EXAMPLE



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

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



95cm

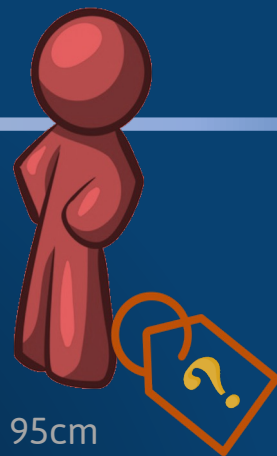
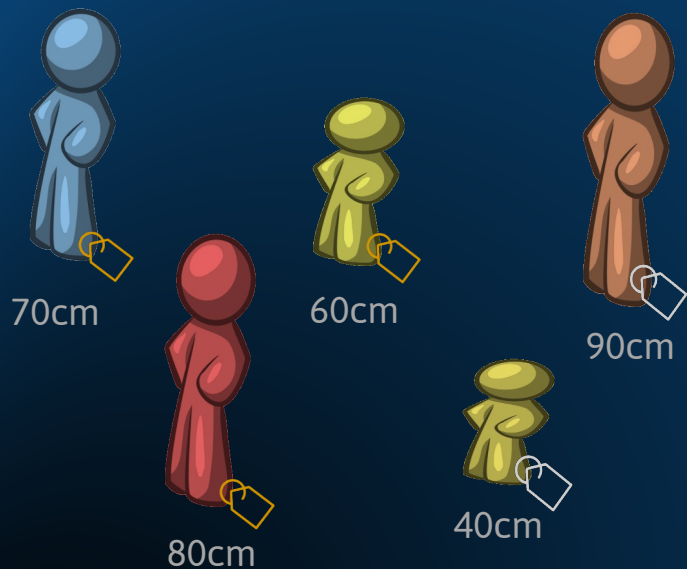
$$\hat{P}(gold) = \frac{3}{5}$$

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

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

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

EXAMPLE



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

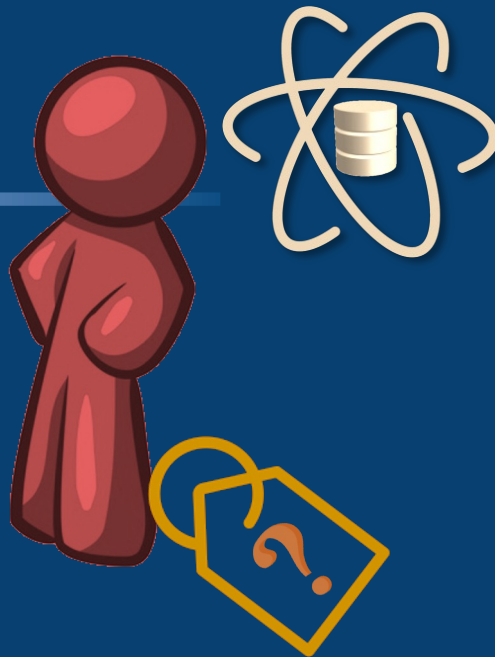
$$\hat{P}(\text{red} \wedge 95 | \text{silver}) =$$

$$\begin{aligned} &= \hat{P}(\text{red} | \text{silver}) \times \hat{P}(95 | \text{silver}) \\ &= 0 \times f_{\text{silver}}(95) \\ &= 0 \end{aligned}$$

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

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

EXAMPLE



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

$$\hat{y} = \arg \max_{c_i} \{P(c_i) \times P(\text{red} | c_i) \times P(95 | c_i)\}$$

$$\hat{y} = \arg \max_{\{\text{gold}, \text{silver}\}} \left\{ \begin{array}{l} P(\text{gold}) \times P(\text{red} | \text{gold}) \times P(95 | \text{gold}), \\ P(\text{silver}) \times P(\text{red} | \text{silver}) \times P(95 | \text{silver}) \end{array} \right\}$$

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





*Thank
you!*

