L4: Bayes' Rule

1.14.2026

FoDA

Jeff M. Phillips

Bayes' Rule

Events M , D

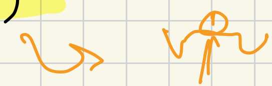
model , data

can compute

$$\Pr(M \mid D) = \frac{\Pr(D \mid M) \cdot \Pr(M)}{\Pr(D)}$$

provided or choice.

prob. model given data



$$P(M \mid D) = \frac{\Pr(M \cap D)}{\Pr(D)} \Rightarrow \Pr(M \cap D) = \Pr(M \mid D) \cdot P(D)$$

$$= \Pr(D \cap M)$$

$$= P(D \mid M) \cdot P(M)$$

Solve for $P(M \mid D) \cdot \Pr(D) = \Pr(D \mid M) \cdot P(M)$

$$\Pr(M \mid D) = \frac{\Pr(D \mid M) \cdot \Pr(M)}{\Pr(D)}$$

Continuous Analog

$$f_{M|D}(m|d) = \frac{f_{D|M}(d|m) \cdot f_M(m)}{f_D(d)}$$

$D = \{0, 1\}$

$M = \{0, 1\}$

$Pr(M) = 0.45$

	$M = 0$	$M = 1$
$D = 1$	0.5	0.25
$D = 0$	0.05	0.2

$Pr(D=1) = 0.75$

$$Pr(M \mid D) = \frac{Pr(M \cap D)}{Pr(D)} = \frac{0.25}{0.75} = \frac{1}{3}$$

$$Pr(M \mid D) = \frac{Pr(D \mid M) \cdot P(M)}{Pr(D)}$$

$= \frac{1}{3}$

$$= \frac{\left(\frac{0.25}{\cancel{0.45}}\right) \cdot (\cancel{0.45})}{0.75}$$

$$= \frac{0.25}{0.75} = \frac{1}{3}$$

Cracked Windshield Example

Event : W windshield cracked.

3 factories : A, B, C

prior $Pr(A) = 0.5$ $\quad Pr(B) = 0.3$ $\quad Pr(C) = 0.2$

$Pr(W \mid A) = 0.01$ $\quad Pr(W \mid B) = 0.1$ $\quad Pr(W \mid C) = 0.02$

$$Pr(A \mid W) = \frac{Pr(W \mid A) \cdot Pr(A)}{Pr(W)} = \frac{(0.01)(0.5)}{Pr(W)} = \frac{0.005}{Pr(W)}$$

$$Pr(B \mid W) = \frac{Pr(W \mid B) \cdot Pr(B)}{Pr(W)} = \frac{(0.1)(0.3)}{Pr(W)} = \frac{0.03}{Pr(W)}$$

$$Pr(C \mid W) = \qquad\qquad = \frac{0.004}{Pr(W)}$$

M = model $\qquad D$ = data $\qquad \Pr(M \mid D)$

$M \in \Omega_M$ $\qquad$ maximize $\Pr(M \mid D)$

sample space models

$\hookrightarrow M^*$ = maximum a priori

posterior

MAP

$$M^* = \underset{M \in \Omega_M}{\operatorname{argmax}} \Pr(M \mid D) = \underset{M \in \Omega_M}{\operatorname{argmax}} \frac{\Pr(D \mid M) \cdot \Pr(M)}{\Pr(D)}$$

Lets assume $\Pr(M)$ constant

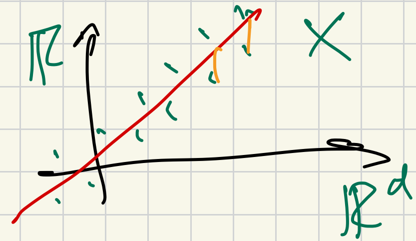
Maximum Likelihood Estimate

MLE $\qquad M^* = \underset{M \in \Omega_M}{\operatorname{argmax}} \Pr(D \mid M) = \underset{M \in \Omega_M}{\operatorname{argmax}} \Pr(M \mid D)$

$D = \{x_1, x_2, \dots, x_n\} = X$ <u>independent</u> observations

M = explanation of structure of data.

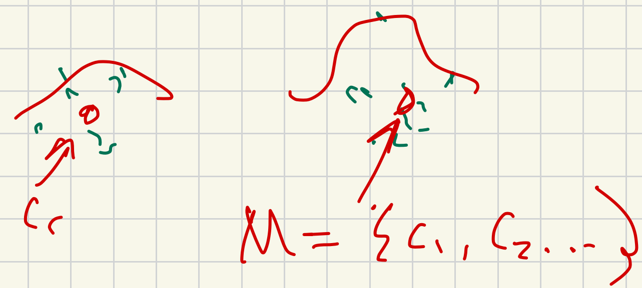
① $D \subset \mathbb{R}^1$ Model = $M \in \mathbb{R}^1$ + Normal Noise.



M = line + Normal Noise

② linear regression

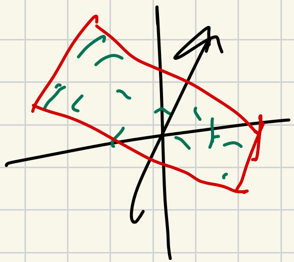
③ clustering $D \subset \mathbb{R}^d$



④ Dimensionality Reduction

$D \subset \mathbb{R}^d$

model $M \subset F_k$



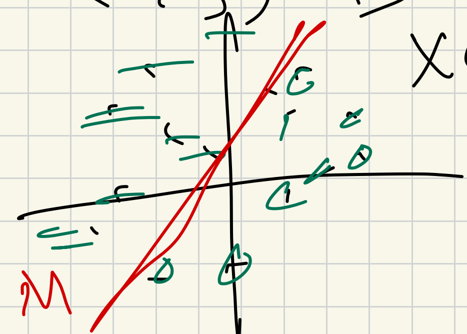
$k < d$

F space k-dim subspaces

⑤ Linear Classification

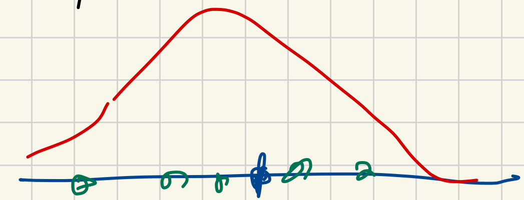
$(X, y) = D$

$X \subset \mathbb{R}^d$



Data $D \subset \mathbb{R}^1 = \{x_1, x_2, \dots x_n\} = \{1, 3, 12, 5, 9\}$

independent



$$D \sim N_{\mu,\sigma}$$

μ parameter

Normal

model

$$\boxed{N_{\mu,2}(x) = g(x) = \frac{1}{\sqrt{8\pi}} \exp\left(-\frac{1}{8}(\mu - x)^2\right)}$$

"Pr" $(x_i = D \mid M = m)$

$$Pr(D \mid m) = \prod_{x_i \in D} g(x_i) = \prod_{x_i \in D} \left(\frac{1}{\sqrt{8\pi}} \exp\left(-\frac{1}{8}(m - x_i)^2\right) \right)$$

$$\Pr(D \mid m) = \prod_{x_i \in D} \underbrace{\left(\frac{1}{\sqrt{8\pi}} \cdot \exp\left(-\frac{1}{8} (m - x_i)^2 \right) \right)}_{g(x_i)}$$

$$m^* = \underset{m \in \mathbb{R}}{\operatorname{argmax}} \Pr(D \mid m) = \underset{m \in \mathbb{R}}{\operatorname{argmax}} \overbrace{\log\left(\Pr(D \mid m) \right)}^{\text{log-likelihood}}$$

$$\ln(\exp(z)) = z$$
$$\log(\exp(z)) = z \cdot c$$

$$\log(a \cdot b) = \log(a) + \log(b)$$

$$= \underset{m}{\operatorname{argmax}} \log\left(\prod_{x_i \in D} g(x) \right) = \sum_{x_i \in D} \log(g(x_i))$$

$$= \underset{m}{\operatorname{argmax}} \sum_{x_i \in D} \left(-\frac{1}{8} (m - x_i)^2 \right) \cancel{+ \sum_{x_i \in D} \log\left(\frac{1}{\sqrt{8\pi}} \right)}$$

no dependence on m

$$= \underset{m}{\operatorname{argmin}} \sum_{x_i \in D} (m - x_i)^2 = \text{average}\{x_1, x_2, \ldots x_n\}$$