FoDA L4

Bayes' Rule

MLEs and Log-likelihoods

Sep 1, 2022

Student Names and Personal Pronouns

Class rosters are provided to the instructor with the students legal name as well as Preferred first name (if previously entered by you in the Student Profile section of your CIS account, which managed can be managed at any time). While CIS refers to this as merely a preference, I will honor you by referring to you with the name and pronoun that feels best for you in class or on assignments. Please advise me of any name or pronoun changes so I can help create a learning environment in which you, your name, and your pronoun are respected. If you need any assistance or support, please reach out to the LGBT Resource Center. https://lgbt.utah.edu/campus/faculty_resources.php

my pronouns: he/him/his

Prof Phillips

text $\xrightarrow{ML}$ [ ]
vector

Review

$Pr(P=blue \wedge S=green)$

pants \ shirt	S = green	S = red	S = blue
P = blue	0.3	0.1	0.2
P = white	0.05	0.2	0.15
	0.35	0.3	0.35

$P_S(\cdot)$

marginal

$$Pr(P=blue \mid S=red) = \frac{Pr(P=blue \wedge S=red)}{Pr(S=red)} = \frac{0.1}{0.3}$$

f_X pdf

$$\int_{w \in \Omega} f_X(w)\, dw = 1$$

$$f_X(w) \in [0, \infty)$$

R.V.s X, Y

$$f_{X,Y}(x, y) \xrightarrow{\text{marginal}} f_X(x) = \int_{y \in \Omega_Y} f(x, y)\, dy$$

$$f_{X|Y}(x, Y = y_1) = \frac{f_{X,Y}(x, y_1)}{f_Y(y_1)}$$

$$f_Y(y_1) = \int_{x \in \Omega_X} f(x, y_1)\, dx$$

Bayes' Rule

Two events M, D

M: model

D: data

can compute

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

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

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

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

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

$$Pr(D) = P(D=1)$$

	$M=1$	$M=0$	
$D=1$	0.25	0.5	0.75
$D=0$	0.2	0.05	
	0.45		

$$Pr(D \mid M) = \frac{Pr(D \cap M)}{P(M)} = \frac{0.25}{0.45}$$

$$P(M \mid D) = \frac{\overset{\frac{0.25}{0.45}}{Pr(D \mid M)} \cdot \overset{0.45}{P(M)}}{\underset{0.75}{Pr(D)}} = \frac{0.25}{0.75} = \frac{1}{3}$$

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

Cracked Windshield

event: W windshield cracked.

3 factories: A, B, C

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

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

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

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

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

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

$M \in \Omega_M$ ← space of model $\qquad$ maximizes $P(M \mid D)$ → M^* $\qquad$ MAP maximum a posteriori

$$M^* = \underset{M \in \Omega_M}{\arg\max} \Pr(M \mid D) = \underset{M \in \Omega_M}{\arg\max} \frac{\Pr(D \mid M) \cdot P(M)}{\cancel{\Pr(D)}}$$

$$= \underset{M \in \Omega_M}{\arg\max} \underbrace{\Pr(D \mid M) \cdot \Pr(M)}_{\text{posterior}}$$

prior

likelihood

Maximum Likelihood Estimator

→ MLE assume $\Pr(M)$ constant

$$M^* = \underset{M \in \Omega_M}{\arg\max} \Pr(M \mid D) = \underset{M \in \Omega_M}{\arg\max} \Pr(D \mid M)$$

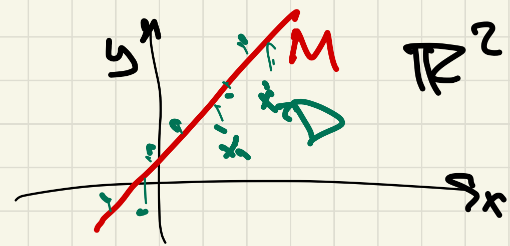
$D = \{x_1, x_2 \ldots x_n\}$ independent observations

M: explaining structure in data

1. $D \subset \mathbb{R}^1$ Model: M = value of average height (Normally distributed)

2. linear regression

M line in $\mathbb{R}^2$

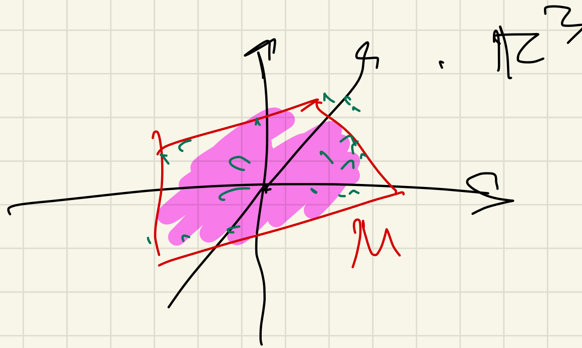


3. clustering $D \in \mathbb{R}^d$

M set of points

4. PCA

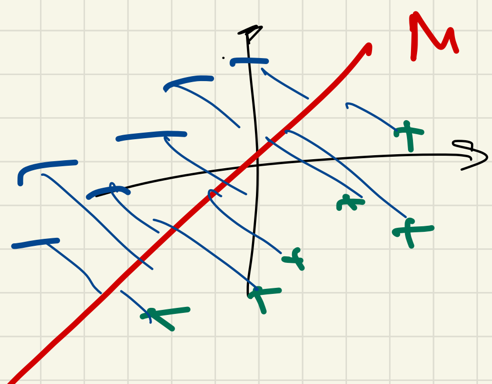
$D \in \mathbb{R}^d$ $\quad M \in F_k$ $\quad$ k-dimensional subspace



dimensionality reduction

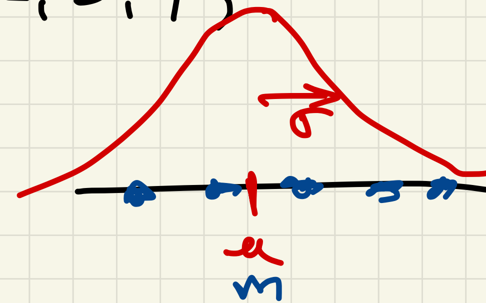
5. (linear) classification

$D \in \mathbb{R}^{d \times \{-1,+1\}}$ $\quad x_i \in \mathbb{R}^d$ $\quad y_i \in \{-1,+1\}$



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

independent



Modeling

$x_i \sim \mathcal{N}(\mu, \sigma)$

model $M = m$

$m \in \mathbb{R} = \Omega_M$

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

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

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

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

$$m^* = \underset{m \in \mathbb{R}}{\arg\max} \; P_o(D \mid m) = \underset{m \in \mathbb{R}}{\arg\max} \; \log\left(P_r(D \mid M)\right)$$

log-likelihood

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

$$= \underset{m}{\arg\max} \; \log\left(\prod_{x_i} g(x_i)\right) = \underset{m}{\arg\max} \sum_{x_i} \underset{\ln}{\log}\left(g(x_i)\right),$$

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

$$= \text{average}(\{x_1 \ldots x_n\})$$