FoDA

L4 : Bayes' Rule

Choose S, P

	S=green	S=red	S=blue
P=blue	0.3	0.1	0.2
P=white	0.05	0.2	0.15
f_S	0.35	0.3	0.35

$$f_{P|S}(\cdot \mid S=\text{red}) = \Pr(P=\text{blue} \mid S=\text{red}) = \frac{0.1}{0.3} = \frac{1}{3}$$

$$\Pr(P=\text{white} \mid S=\text{red}) = \frac{0.2}{0.3} = \frac{2}{3}$$

Gaussian Distribution

$f = G_d : \mathbb{R}^d \to \mathbb{R}$

$d = 2$

$$G_d(v) = \frac{1}{\sigma^d \sqrt{(2\pi)^d}} \exp\left(-\frac{\|v-\mu\|^2}{2\sigma^2}\right)$$

$\left(G_d = N_{\text{normal}}\right)$

std. dev

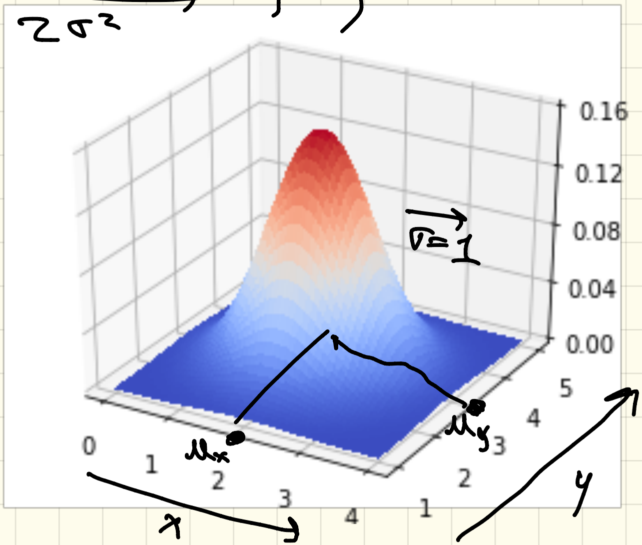
mean

$$G_2(v) = \frac{1}{2\pi\sigma^2} \exp\left(\frac{-(v_x-\mu_x)^2-(v_y-\mu_y)^2}{2\sigma^2}\right)$$

$v = (v_x, v_y)$

$\mu = (2,3)$

$\sigma = 1$



Bayes' Rule

Two R.V. M, D

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

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

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

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

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

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

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

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

$$= \frac{0.25}{0.25+0.5} = \frac{1}{3}$$

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

$$= \frac{\left(\frac{0.25}{\cancel{0.2+0.25}}\right)\left(\cancel{0.2+0.25}\right)}{0.25+0.5}$$

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

Cracked Windshield

event $\underline{W}$ $\Leftarrow$ windshield on my car cracked

R.V. $F \in \{A, B, C\}$
factory

my region

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

$Pr(W|A) = 0.01$
$Pr(W|B) = 0.10$
$Pr(W|C) = 0.02$

$$Pr(F|W) = \frac{P(W|F) \cdot Pr(F)}{Pr(W)}$$

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

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

$$P(C|W) = \frac{(0.02)(0.2)}{Pr(W)} = \frac{0.004}{Pr(W)}$$

$$1 = \frac{1}{Pr(W)}(0.005 + 0.03 + 0.004)$$

$$Pr(W) = 0.039$$

$$\Pr(\underset{\text{model}}{M} \mid \underset{\text{data}}{D})$$

maximum a posteriori (MAP)

$M \in \Omega_M$ ← space of models

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

$$= \underset{M \in \Omega_M}{\operatorname{argmax}} \underbrace{\Pr(D \mid M)}_{\substack{\text{likelihood of model} \\ L(M)}} \cdot \Pr(M)$$

$$S = \{a, b, c\}$$

f	f(x)
a	7
b	10
c	3

max

argmax

$$\max_{x \in S} f(x) = 10$$

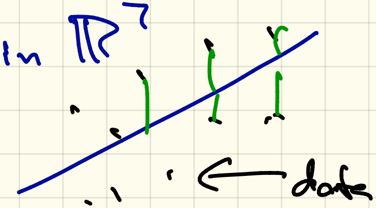
$$x^* = \underset{x \in S}{\operatorname{argmax}} f(x) = b$$

Examples Models / Data

model: simple pattern, summary of data

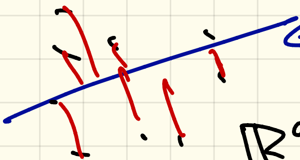
data: points in $\mathbb{R}^d$, values in $\mathbb{R}$

- M single point
- linear regression — M = line in $\mathbb{R}^2$



- clustering — M = set of k points

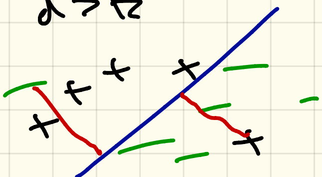
- PCA



← k-dim subspace

$\mathbb{R}^d$ $d > k$

- linear classification



Log - likelihood

assume $Pr(M) = Pr(M')$ for all M, M'

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

likelihood $L(M)$

$$= \underset{M \in \Omega_M}{\operatorname{argmax}} \; \log\left(Pr(D \mid M)\right)$$

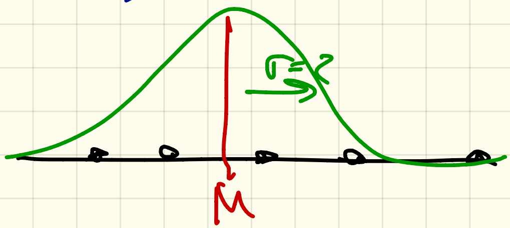
maximum likelihood estimate (MLE)

$$\log_{b_1}(x) = \frac{\log_{b_2}(x)}{\log_{b_2}(b_1)}$$

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

data $D = \{1, 3, 12, 5, 9\}$

assume $\sim$ independently $\mathcal{N}(\mu, \sigma = 2)$ model



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

$$= \Pr(D = x \mid \mu)$$

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

$$\ln(\Pr(D \mid \mu)) = \ln\left(\prod_{x \in D} \left(\frac{1}{\sqrt{8\pi}} \cdot \exp\left(-\frac{1}{8}(\mu - x)^2\right)\right)\right)$$

$$= \sum_{x \in D} \left(-\frac{1}{8}(\mu - x)^2\right) + \cancel{|D| \ln\left(\frac{1}{\sqrt{8\pi}}\right)}$$ in argmax

$$\mu^* = \arg\min_{\mu \in \mathbb{R}} \sum_{x \in D} (\mu - x)^2 = \frac{1}{|D|} \sum_{x \in D} x = \text{mean}(D)$$

product because independence noise

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

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

$$= \sum_{x \in D} \left(-\frac{1}{8}(M-x)^2\right) + \cancel{|D| \ln\left(\frac{1}{\sqrt{8\pi}}\right)}$$ in argmax

$$M^* = \arg\min_{M \in \mathbb{R}} \sum_{x \in D} (M-x)^2 = \frac{1}{|D|} \sum_{x \in D} x = \text{mean}(D)$$

why $\displaystyle \operatorname*{argmin}_{x \in \mathbb{R}} \sum_{x \in D} (x-M)^2 = \frac{1}{|D|} \sum_{x \in D} x$?

$n = |D|$

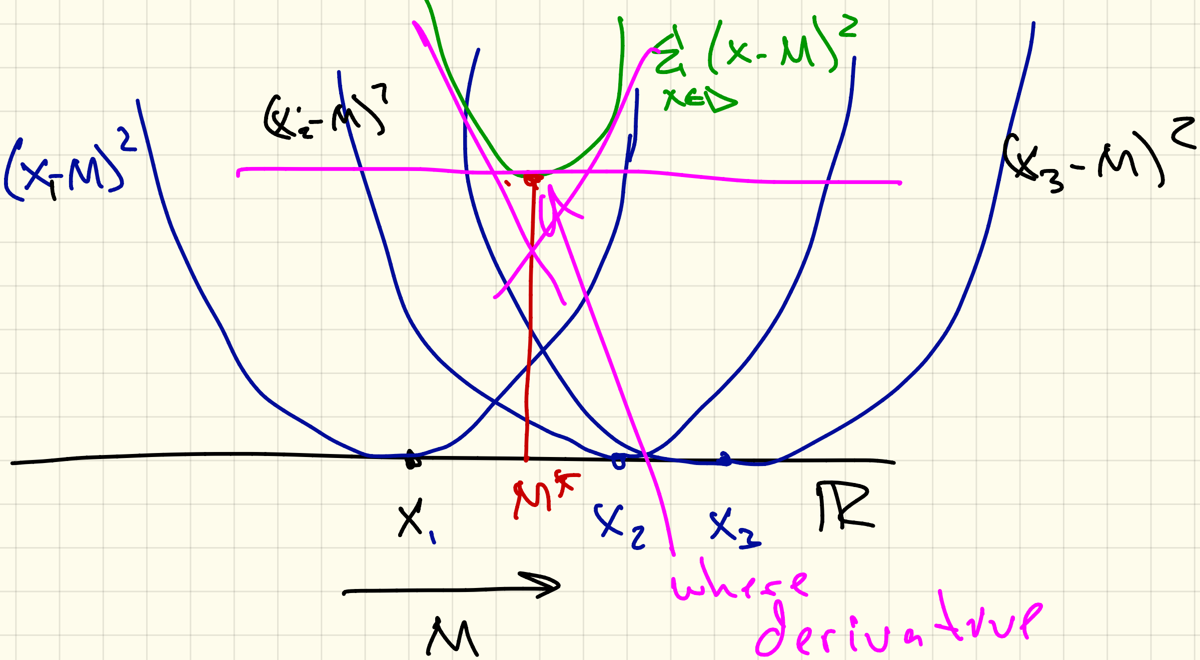
$$S_D(M) = \sum_{x \in D} (x-M)^2 = \sum_{x \in D} M^2 - 2xM + x^2 = nM^2 - \Big(2\sum_{x \in D} x\Big)M + \sum_{x \in D} x^2$$

$$\frac{\partial S_D(M)}{\partial M} = 2nM - 2\sum_{x \in D} x = 0$$

set to 0

$$M(2n) = 2\sum_{x \in D} x$$

$$M = \frac{1}{n}\sum_{x \in D} x = \text{mean}(D)$$



where derivative $\frac{\partial S_x}{\partial} = 0$

$$\frac{\partial^2}{\partial M^2} S_x(M^*) = 2n > 0$$

$\hookrightarrow$ minimum