FoDA L4

Bayes' Rule

Bayes' Rule

two events M, D

model data

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

likelihood prior

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

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

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

	$M=1$	$M=0$	
$D=1$	0.25	0.5	0.75
$D=0$	0.2	0.05	0.25
	0.45	0.55	

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

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

$$= \frac{(0.25)\cancel{(0.45)}}{\cancel{(0.45)}(0.75)} = \frac{0.25}{0.75} = \frac{1}{3}$$

Cracked Windshield Examp.

event W (windshield cracked)

factory: $\{A, B, C\} \in M$ models

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

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

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

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

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

$\Pr(M \mid D) \rightarrow$ Maximum a posteriori (MAP)

$\in$ \in model $M \in \Omega_M$

prior

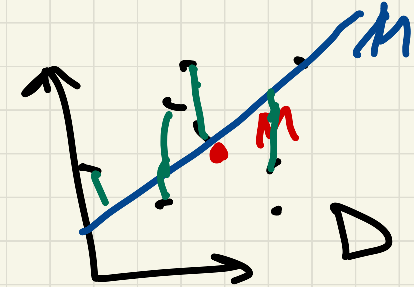
$$M^* = \underset{M \in \Omega_M}{\operatorname{argmax}} \Pr(M \mid D) = \underset{M \in \Omega_M}{\operatorname{argmax}} \frac{\Pr(D \mid M) \cdot P(M)}{\Pr(D)}$$

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

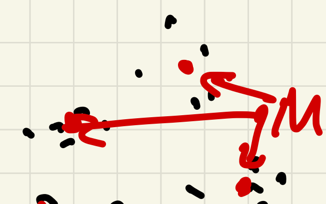
Likelihood $L(M) = \Pr(D \mid M)$

What is model? and data?

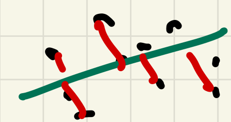
- model point in $\mathbb{R}^d$ (+ noise)
  data set iid points in $\mathbb{R}^d$
- <u>linear regression</u>
  model M a line in $\mathbb{R}^2$



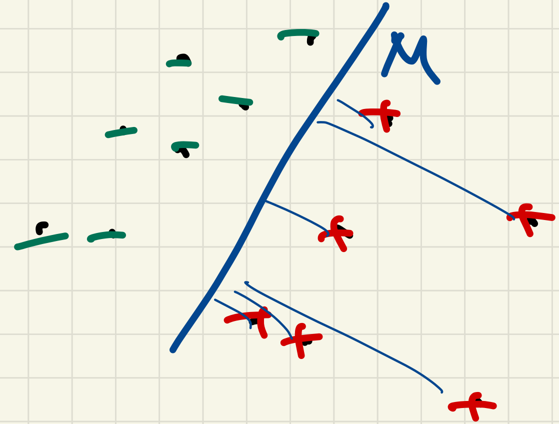
- <u>clustering</u>
  model M a set of k points



- <u>dim reduction</u>
  model M k-dimensional hyperplane



classification



Log-likelihood

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

assume uniform prior

$\boxed{\Pr(M) \text{ const.}}$

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

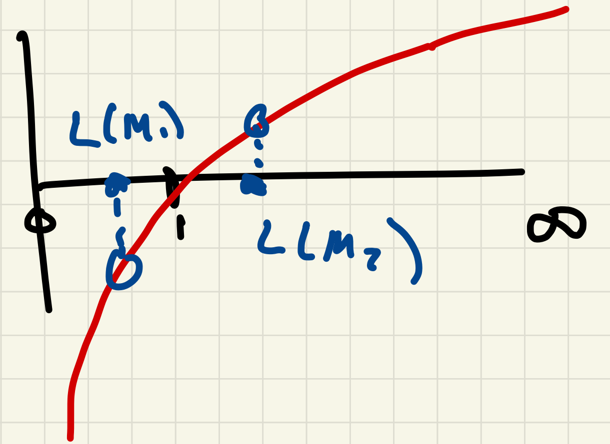
$$= \underset{M}{\arg\max} \; \Pr(D \mid M) = \underset{M}{\arg\max} \; L(M)$$

Maximum Likelihood Estimate (MLE)

$$M^* = \arg\max_M L(M)$$

$$= \arg\max_M \log(L(M))$$

$$L(M) = \Pr(D|M)$$



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

independence

likelihood $L(M) = \underbrace{\Pr(x_1)}_{p_1} \cdot \underbrace{\Pr(x_2)}_{p_2} \cdots \underbrace{\Pr(x_n)}_{p_n}$

$$\log(L(M)) = \log\left(\prod_{i=1}^{n} p_i\right) = \sum_{i=1}^{n} \log(p_i)$$

product

sum

Data set of points in $\mathbb{R}^1$: $\{\overset{x_1}{1}, \overset{x_2}{3}, \overset{x_3}{12}, \overset{x_4}{5}, \overset{x_5}{9}\} = D$

independent observations

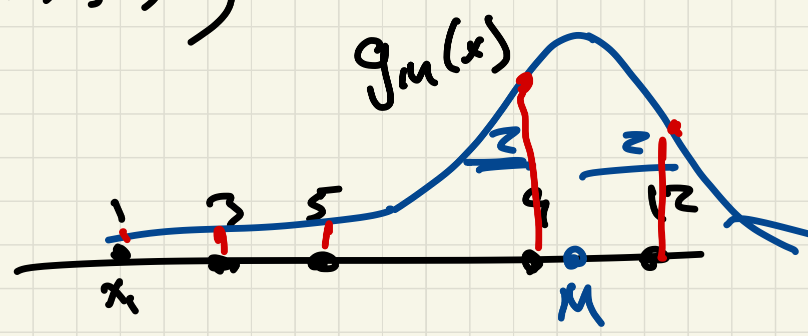
Model $M \in \mathbb{R}$ (+ Normal noise

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

$P(D \mid M)$

$$L(M) = \prod_{x \in D} g_M(x)$$

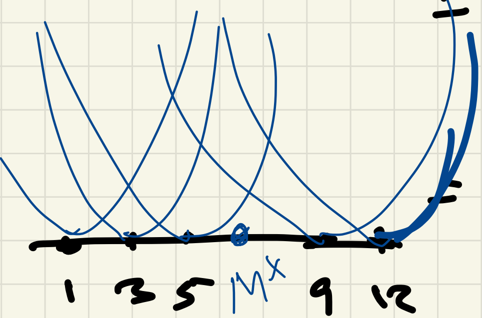
$$= \prod_{x \in D} \left(\frac{1}{\sqrt{8\pi}} \cdot \exp\left(-\frac{1}{8}(M-x)^2\right)\right)$$



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

$$M^* = \frac{1}{|D|}\sum_{x\in D} x$$

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



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

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

$$M^* = \underset{M\in\mathbb{R}}{\operatorname{argmax}}\ \ln(L(M)) = \underset{M\in\mathbb{R}}{\operatorname{argmax}} -\cancel{\tfrac{1}{8}}\sum_{x\in D}(M-x)^2$$

$$= \underset{M\in\mathbb{R}}{\operatorname{argmin}} \sum_{x\in D}(M-x)^2 = \text{average}(D)$$