CS3244 Lecture 2

Wednesday, 11 May 2022 9:20 AM

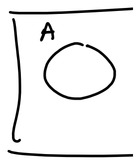
Problem (w data) $\Rightarrow$ $y = w^T x$ (model) $\Rightarrow$ $\min_w f(w)$ (objective function) $\Rightarrow$ $\nabla f(w) = 0$ (optimization) $\Rightarrow$ Model !! ☺

Problem (w data) $\Rightarrow$ $p(D|\theta)$ [w] (model) $\Rightarrow$ $p(\theta|D)$ [w, data] (inference) $\Rightarrow$ solve $\Rightarrow$ Model

---

Probability Space

$\underbrace{\mathbb{P}(\{⚁, ⚃, ⚅\})}_{\in E} = \frac{1}{2}$ $\mathbb{P}(⚀) = 1$



$(\Omega, E, \mathbb{P})$ $\mathbb{P}: E \mapsto [0, 1]$

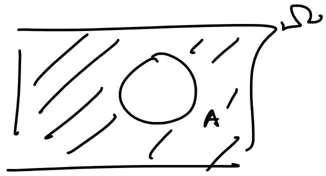
Ω: e $\{⚀, ⚁, ⚂, \dots, ⚅\}$

E: Set of ~~sets~~ subsets of Ω (technically wrong)

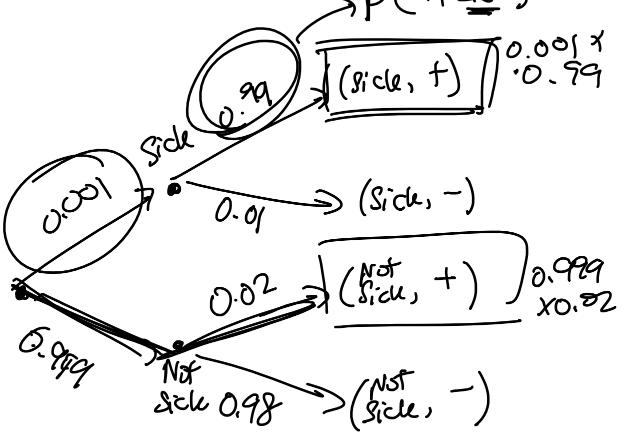
① $\mathbb{P}(\Omega) = 1$
② $\mathbb{P}(\emptyset) = 0$
③ $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}$

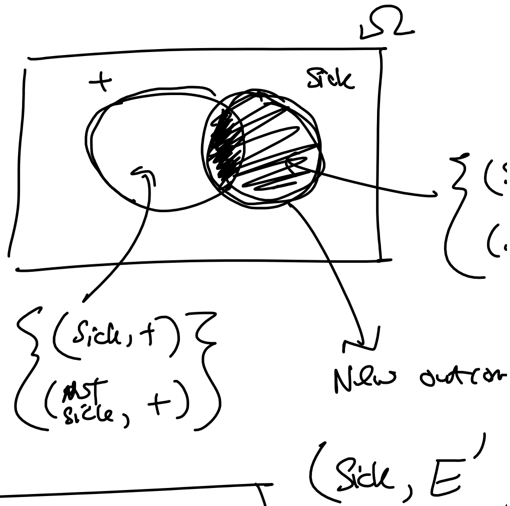
E satisfies:

① if $A \in E$, $B \in E$, then $A \cup B \in E$
② " " " " $A \cap B \in E$
③ if $A \in E$, then $A^c \in E$ ($\Omega - A$)
④ $\emptyset \in E$, $\Omega \in E$



$p(+|\text{Sick})$





$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$\boxed{P(+|\text{Sick}) = \frac{P(\text{Sick}, +)}{P(\text{Sick})}}$$

Product Rule

$$P(\text{Sick}|+) = \frac{P(\text{Sick}, +)}{P(+)} = \frac{0.001 \times 0.99}{0.001 \times 0.99 + 0.999 \times 0.02} = \#$$

$P(+)$ $0.001 \times 0.2 \ldots$

Sum rule.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \implies P(A \cap B) = P(B)P(A|B)$$

Product Rule.

$$P(A_1 \cap A_2 \cap \ldots, A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2, A_1)\ldots P(A_n|A_1 \ldots A_{n-1})$$

$$= P(A_{13})P(A_{17}|A_{13})P(A_{21}|A_{13}, A_{17}) -$$

Independence

def $P(A \cap B) = P(A)P(B)$

$$P(A \cap B) = P(A)P(B|A) = P(A)P(B)$$

if A and B disjoint

$\implies$ A and B are dependent.

Random Variable

if X is a random variable (r.v.), then:

$$X : \Omega \to \mathbb{R}$$

Example: $\Omega = \{$⚀, ⚁, ..., ⚅$\}$

$X(\omega) = \begin{cases} 1 & \text{if } \omega = \text{⚀} \\ 2 & \text{if } \omega = \text{⚁} \\ \vdots & \\ 6 & \text{if } \omega = \text{⚅} \end{cases}$

$1 + 6 =$

$X_1(\omega_1) + X_2(\omega_2) =$

$[\omega_1 + \omega_2]$

⚀ + ⚅

Indicator r.v.

$$\mathbb{1}_L(\omega) = \begin{cases} 0 & \text{if } L \text{ is false} \\ 1 & \text{if } L \text{ is true.} \end{cases}$$

is even?

$$\mathbb{1}_{\text{is even}}(\omega) = \begin{cases} 0 & \text{if } \omega \in \{\text{⚀} \\ 1 & \text{if } \omega \in \{\text{⚁} \end{cases}$$

Ω

if $\mathbb{1}$

$(\{0,1\}, E, \mathbb{P}')$ $\quad \mathbb{P}(X=1)$, $\mathbb{P}(X=0)$

$E = \{\emptyset, \{0,1\}, \{0\}, \{1\}\}$.

<u>Distributions</u>

<u>Discrete</u>

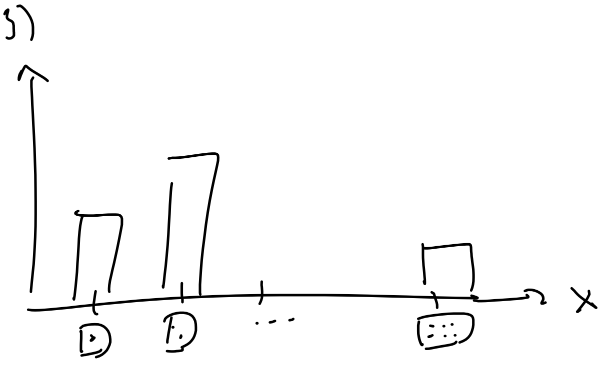
$(\Omega, E, \mathbb{P})$

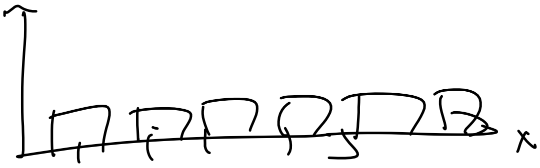
$\Omega = \{⚀, \dots, ⚅\}$

$\mathbb{P}(\{⚀\})$

$\vdots$

$\mathbb{P}(\{⚅\})$





Discrete PMF

Uniform $p(x) = \frac{1}{|\Omega|}$

Poisson $p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$

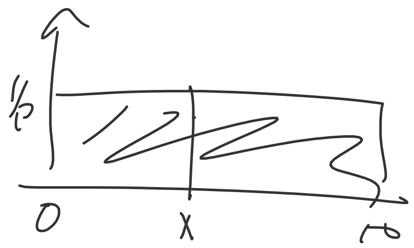
$\{⚀, ⚁, ⚄\} \Rightarrow \sum_{x \in \{⚀, ⚁, ⚄\}}$

<u>continuous distribution.</u>

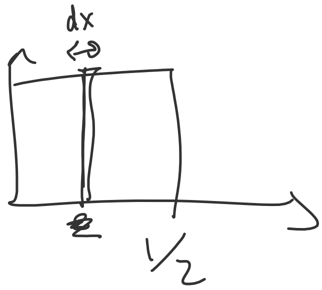
<u>$pdf(x) > 1$</u> ?

$pdf(\frac{1}{4}) = 2$

YES

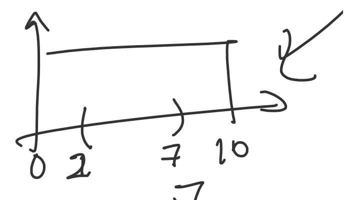


0

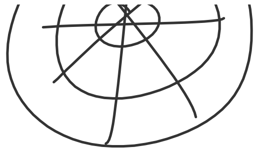


$\mathbb{P}(C) = \int_C pdf(x)\, dx$

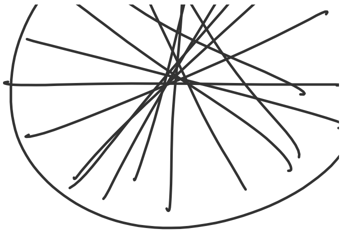
$\mathbb{P}([2,7]) = \int_{[2,7]} \frac{1}{10}\, dx$



$$|\mathbb{P}(\{S\}) = 0|$$

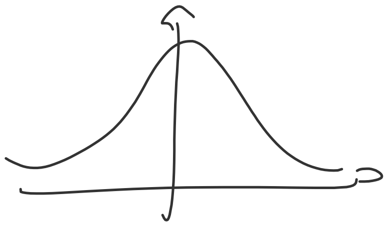


w.p zero.

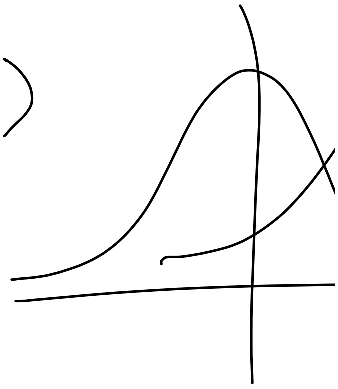


Gaussian distribution

$$pdf_N(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$



if X, Y follows a gaussian distribution:



then ① $X + Y$ is gaussian $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

② aX is gaussian $N(a\mu_1, a^2\sigma_1^2)$

③ $X + a$ is gaussian $N(\mu_1 + a, \sigma_1^2)$

Learning parameter for gaussian distribution

Given $N(\mu_*, 1)$ $\mathcal{D} = \{x_1, x_2, \ldots, x_N\}$ sampled from a

don't know.

Guess: $\frac{1}{N}\sum_{n=1}^{N} x_n \approx \mu_*$

$= \hat{\mu}$

we need justification...

$$\boxed{\underset{\mu}{\text{maximize}}\ p(\mathcal{D}|\mu)}$$

Maximum likelihood Estimation. (MLE)

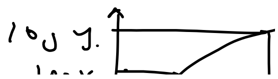
$$p(\mathcal{D}|\mu) = p(x_1|\mu)\,p(x_2|\mu) \cdots p(x_N|\mu)$$

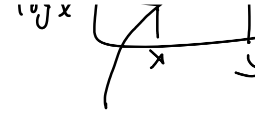
$$= \prod_{n=1}^{N} p(x_n|\mu)$$

$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(x_n-\mu)^2}{2}\right)$$

$\forall n: \; p(x_n|\mu) = \frac{1}{\sqrt{2\pi}} \exp$

Detour





log.
$$\Longrightarrow$$

$$\log p(D|\mu) = \log \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(x_n-\mu)^2}{2}\right)$$

$$= \sum_{n=1}^{N} \underbrace{\log \frac{1}{\sqrt{2\pi}}}_{\text{ind. from } \mu} + \log \exp\left(\frac{-(x_n-\mu)^2}{2}\right)$$

$$\Longrightarrow \sum_{n=1}^{N} \cancel{\log \exp}\left(\frac{-(x_n-\mu)^2}{2}\right)$$

$$= \sum_{n=1}^{N} \frac{-(x_n-\mu)^2}{2}$$

max $-(\cdot)$

min $(\cdot)$

$$\xrightarrow{\min} \sum_{n=1}^{N} \frac{(x_n-\mu)^2}{2}$$

$$\xrightarrow{\frac{d}{d\mu}} \sum_{n=1}^{N} \frac{-\cancel{2}(x_n-\mu)}{\cancel{2}} = \sum_{n=1}^{N} -(x_n-\mu) \overset{\text{set}}{=} 0$$

$$\underbrace{\sum_{n=1}^{N} \mu}_{= N\mu} = \sum_{n=1}^{N} x_n \Longrightarrow \boxed{\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n}$$

if $x < y$

then $\log x < \log y$.

if $f(\mu_1) < f(\mu_2)$

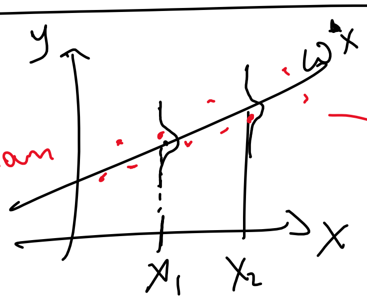
then $\log f(\mu_1) < \log f(\mu_2)$.

Linear regression REVISITED

given x true w_* ← unknown

$\rightarrow w_*^T x + \epsilon$

$\epsilon \sim N(0, 1)$



$\rightarrow$ D :

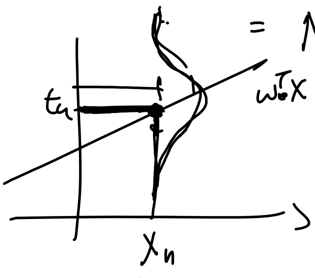
$$\log p(D|w)$$

$$= \log \prod_{n=1}^{N} p(t_n|x_n, w)$$

$$= \log \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(t_n - w^T x_n)^2}{2}\right)$$

$$= \sum_{n=1}^{N} \underbrace{\log \frac{1}{\sqrt{2\pi}}}_{\text{ind. from } w} + \frac{-(t_n - w^T x_n)^2}{2}$$

maximize.



$p(t_n|x_n, w)$

$= N(w_*^T x_n, 1)$

$t_n = w_*^T x_n + \epsilon$

$t_n \sim N(w_*^T x_n + \epsilon$

N(0,1)

$$\Rightarrow -\sum_{n=1} \frac{(t_n - w^T x_n)}{2} \leftarrow$$

$$\Rightarrow \sum_{n=1}^{N} \frac{(t_n - w^T x_n)^2}{2} \leftarrow \text{minimize}$$

$$\leftarrow E(w)$$

$t_n = w x_n + \epsilon$

$\sim N(w^T x_n, 1)$

Maximum a posteriori estimator

$p(D|w)$

$$p(w|D) = \frac{p(D|w)\,p(w)}{p(D)}$$

$p(D)$: Don't care

Linear Regression.

$$\log p(w|D) = \log \frac{p(D|w)\,p(w)}{p(D)}$$

$$= \log p(D|w) + \log p(w) - \underbrace{\log p(D)}_{\text{ind. } w}$$

$$\Rightarrow \log p(D|w) + \log p(w).$$

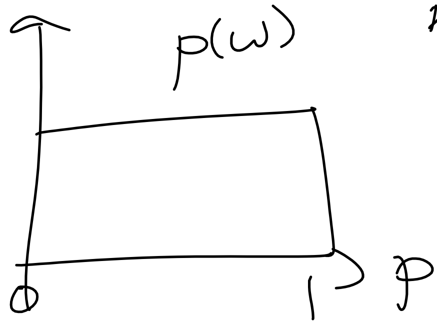
$$= \sum_{n=1}^{N} \frac{-(t_n - w^T x_n)^2}{2} + \log p(w).$$

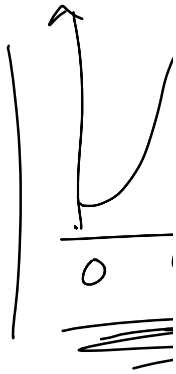
$$= \sum_{n=1}^{N} \frac{-(t_n - w^T x_n)^2}{2} + \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-w^2}{2\sigma^2}\right)\right)$$

$$= \sum_{n=1}^{N} \cdots + \underbrace{\log \frac{1}{\sqrt{2\pi\sigma^2}}}_{\text{ind. } w} + \frac{-w^2}{2\sigma^2}$$

$$\Rightarrow -\sum_{n=1}^{N} \frac{(t_n - w^T x_n)^2}{2} - \frac{w^2}{2\sigma^2} \leftarrow \max$$

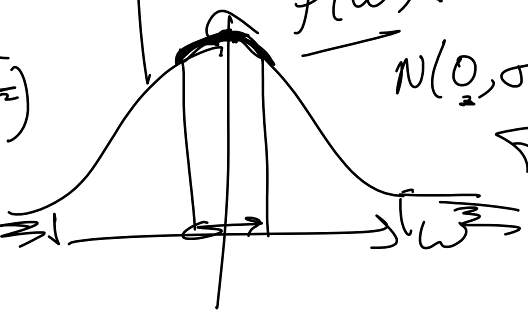
$$\Rightarrow \sum_{n=1}^{N} \frac{(t_n - w^T x_n)^2}{2} + \frac{1}{2} w^2 \leftarrow \min$$





$p(w) \leftarrow$

$N(0, \sigma^2)$



$(\Omega, E, \mathbb{P})$ $\quad \mathbb{P}(\Omega)$

has to have $\Omega, \emptyset$