IIT DELHI

07/08/2025 [AIL7024] (Recap) MLE (Non-ML)

$\mathcal{D} = \{x_i\}$ $\quad x_i \sim \mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$

$\theta = \{\mu, \sigma^2\} \in \Theta$ $\quad \underset{\text{actual}}{\hat{\theta}} = \{\hat{\mu}, \hat{\sigma}^2\}$

Goal: To find $\hat{\theta}$.

$$\theta_{MLE} = \underset{\theta \in \{\mu, \sigma^2\}}{\operatorname{argmax}} \prod_{i=1}^{N} \mathcal{N}(x_i; \mu, \sigma^2)$$

$$= \operatorname{argmax} \frac{-1}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu)^2 - \frac{N}{2} \log \sigma^2 + C$$

$$\frac{\partial \theta_{MLE}}{\partial \mu} = 0 \Rightarrow \mu_{MLE} = \sum x_i / N$$

$$\sigma^2_{MLE} = \frac{\sum (x_i - \mu)^2}{N}$$

Max Likelihood $\equiv$ Min Cross Entropy

— α —

MLE (ML-Setup)

$\mathcal{D} = \{x_i, y_i\}$

$$y|x \sim \mathcal{N}(\hat{w}^T x + \hat{b}, \sigma^2)$$

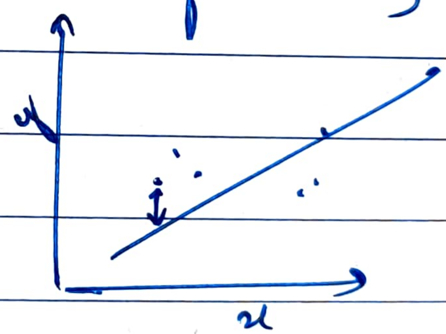
$$y = \hat{w}^T x + \hat{b} + \epsilon$$ gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$

$$\underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{i=1}^{N} \log \mathcal{N}(y_i ; w^T x_i + b, \sigma^2)$$

$$= \underset{\theta}{\operatorname{argmax}} \sum \frac{-1}{2\sigma^2} (y_i - (w^T x_i + b))^2 - \frac{N}{2} \log \sigma^2$$

$$= \underset{\theta}{\operatorname{argmax}} \frac{-1}{2\sigma^2} \underbrace{\sum (y_i - (w^T x_i + b))^2}_{\text{Loss fn of least squares (linear regression)}} - \frac{N}{2} \log \sigma^2$$

minimizing sq loss
$\equiv$ maximizing log likelihood.



— α —

MLE → Least Square Linear Regression

$$\underset{w}{\operatorname{argmax}} \frac{-1}{2\sigma^2} \sum_{i=1}^{n} (w^T x_i - y_i)^2$$

m: # features
n: # datapts.

$x = \{x_1, x_2 \, .. \, x_n\}$

$w^T x$

$w^T x = w_1 x_1 + w_2 x_2 + \dots + w_0 b$

$$w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{pmatrix} \leftarrow w_0$$

IIT DELHI

$$\underset{w}{\arg\max} \; -\Sigma (w^T x_i - y_i)^2$$

$$= \underset{w}{\arg\min} \; \underbrace{\Sigma (w^T x_i - y_i)^2}$$

↳ This looks like an l_2-norm

square : $\|\vec{v}\|_2^2 = v_1^2 + v_2^2 + \dots + v_m^2 = \|V\|_2^2$

$= \Sigma_i v_i^2 = V^T V$

$(w^T x_i - y_i)$ is the ith entry of ________ ? ⇒ $XW - Y$

Defining some matrices,

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}_{n \times 1}$$

$$X = \begin{pmatrix} \text{———} \\ \text{———} \\ \text{———} \\ \text{———} \end{pmatrix}_{n \times m}$$ ⇒ Design Matrix

$$W = \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix}_{m \times 1}$$

IIT DELHI

$W|_{m\times 1}$ $\quad Y|_{n\times 1}$

$X|_{n\times m}$

Now,

$$\underset{W}{\arg\min} \|XW - Y\|_2^2$$

$$\underset{W}{\arg\min} \|XW - Y\|_2^2 = (XW-Y)^T(XW-Y)$$

$$= (W^TX^T - Y^T)(XW-Y)$$

$$= W^TX^TXW - \underbrace{W^TX^TY - Y^TXW}_{\text{(Scalars)}} + Y^TY$$

$$= W^TX^TXW - 2Y^TXW + Y^TY$$

Gradient

$$\nabla_W = 2X^TXW - 2X^TY = 0 \quad \text{(H.W.)}$$

$$X^TXW = X^TY$$

** $\boxed{W = (X^TX)^{-1}X^TY}$

↳ if X^TX is invertible

Closed form solution for Linear Regression

$X^TX|_{m\times m}$

→ full rank

→ independent columns/features

feature selection

IIT DELHI

$\nabla_w^2 = 2X^TX \leftarrow$ Hessian (H.W.)

Check!

↳ Claim: positive semi definite

why? — (Homework!)

$Y|X \overset{i.i.d.}{\sim} \mathcal{N}(\ )$

① if the distn is Non Gaussian?

② if the ~~same~~ distributions are non-identical?

③ Sampling has dependencies?

①

Assume Laplace Distribution.

$$\text{Laplace} \sim \frac{1}{2\sigma} \exp\left\{\frac{-|y-\mu|}{\sigma}\right\}$$

$$\underset{w}{\operatorname{argmax}} \sum_{i=1}^{n} \log \text{laplace}(y_i; w^T x_i, \sigma)$$

$$\sum \frac{-1}{\sigma} |y_i - w^T x_i| + C$$

$$\frac{-1}{\sigma} \underbrace{\sum |y_i - w^T x_i|}_{} + C$$

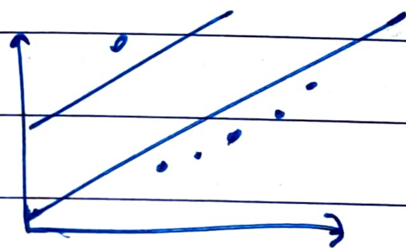
↳ Least Absolute Linear Regression.

problems?
- no closed form solution.

↓

Iterative Optimization!

why do we need this form then?
↳ useful in certain cases.
↳ in presence of outliers.



Discussion on Outliers.

heavy tailed → higher prob for outliers

gaussian ↳ tend to give lesser prob to outliers

Gaussian will give low prob to outlier.

(2) Non-identical

$$y_i | x_i \sim \mathcal{N}(w^T x_i, \sigma^2)$$

in our case,

$$y_i | x_i \sim \mathcal{N}_i(w^T x_i, \sigma_i^2)$$

Doing MLE,

$$\arg\max_{W} \sum \log \mathcal{N}(y_i; W^T x_i, \sigma_i^2)$$

$$\sum \frac{1}{2\sigma_i^2} \underbrace{(W^T x_i - y_i)^2}_{\hookrightarrow\, i^{th} \text{ entry of } (XW-Y)^T \Sigma (XW-Y)} + C$$

claim: $\exists$ closed form solution.

$$(XW-Y)^T \Sigma (XW-Y)$$

↳ Diagonal Matrix

H.W.!

$$\Sigma = \begin{pmatrix} 1/\sigma_1^2 & & 0 \\ & 1/\sigma_2^2 & \\ 0 & & \ddots \\ & & & 1/\sigma_n^2 \end{pmatrix}$$

3 Gaussian Distn w/ Dependent Samples

↳ Homework!

Next Lecture : MAP.