Date :

Maximum Likelihood Estimation – Linear Regression

⇒ Least Squared Error is optimal loss function for Linear Regression

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{m} (\hat{y}^i - y^i)^2$$

How is data generated? → IID gaussian noise (identically & independentally distributed) normal distribution

$$y^i = \theta^T x^i + \epsilon^i$$

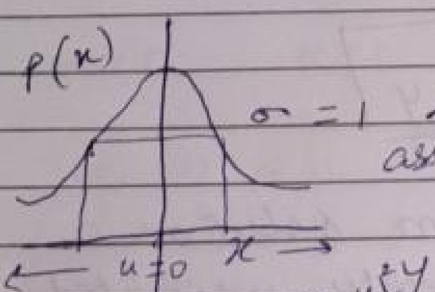
y ↑ target, ϵ ↑ error

$$= \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n + \epsilon^{(i)}$$

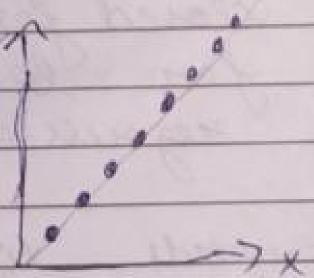
$$\epsilon^i \sim N(0, \sigma^2)$$

normal distribution, mean, variance

$\sigma = 1$ ↗ assumption

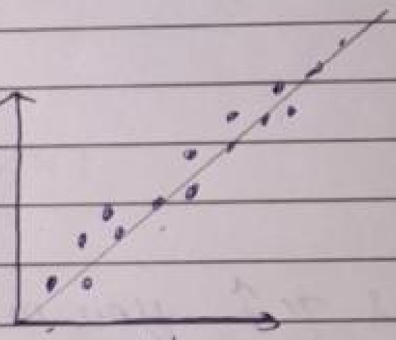


$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-u}{\sigma}\right)^2}$$



ideal

$\theta_1 x_1 + \theta_0 = y$



with noise

$y = \theta_1 x_1 + \theta_0 +$ error

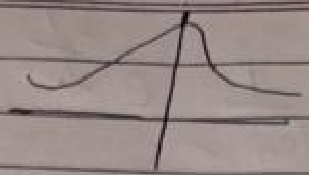
which is like the real data we usually collect

Page No.

Date :

std normal distribution

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$



$$\epsilon^i \sim N(0, \sigma^2) \quad \text{— (1)}$$

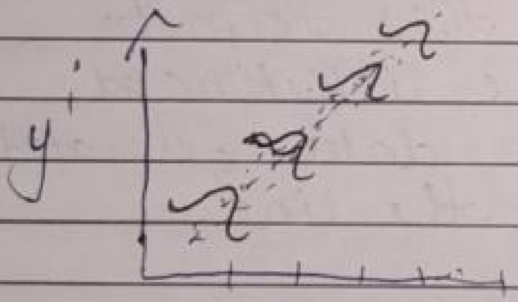
$$P(\epsilon^i) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{(\epsilon^i - 0)^2}{\sigma^2}}$$

$$y^i = \underbrace{\theta^T x^i}_{\text{constant}} + \epsilon^i$$

coming from normal distribution $N(0, \sigma^2)$

$$y^i = N(\theta^T x_i, \sigma^2)$$

$$P(y^i / x^i; \theta) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{(y^i - \theta^T x^i)^2}{\sigma^2}}$$



std normal distribution at each x^i

maximum likelihood estimate (probabilistic approach)

θ that maximize the prob of generating & data is generated through the line x

$$\theta_* = \arg\max_{\theta} P(\text{data}, \theta)$$

$$D = y^1 y^2 \ldots y^m \mid x^1 x^2 \ldots x^m$$

$$P(y^1 y^2 \ldots y^m \mid x^1 x^2 \ldots x^m, \theta) = \prod_{i=1}^{m} P(y^i / x^i; \theta)$$

Page No.
Date :

$$= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-1}{2\sigma^2}(y^i - \theta^T x^i)^2}$$

$P(\vec{y} | x; \theta) = L(\theta) \rightarrow$ likelihood

$$\underset{\theta}{\operatorname{argmax}}\ L(\theta)$$

Take log likelihood

maximize

$$\log L(\theta) = \sum_{i=1}^{m} \underbrace{\log \frac{1}{\sqrt{2\pi}\,\sigma}}_{\text{const}} - \underbrace{\left(\frac{1}{2} \sum_{i=1}^{m} \frac{(y^i - \theta^T x^i)^2}{\sigma^2}\right)}_{\text{minimize}}$$

$$= \operatorname{argmin} \frac{1}{2} \sum_{i=1}^{m} \frac{(y^i - \theta^T x^i)^2}{\sigma^2}$$

isn't this $J(\theta)$ loss fnx

minimize this term to maximize likelihood of finding data generated through the line