Maximum Likelihood Estimation (MLE)

$\mathcal{X}$: training data set

θ : parameters

$$\theta^*_{MLE} = \arg\max_{\theta} p(\mathcal{X} | \theta)$$

$$p(\mathcal{X} | \theta) = \prod_{i=1}^{N} p(x_i | \theta)$$

Maximum a Posteriori Estimation (MAP)

$$\theta^*_{MAP} = \arg\max_{\theta} p(\theta | \mathcal{X})$$

$$= \arg\max_{\theta} \frac{p(\mathcal{X} | \theta)\, p(\theta)}{p(\mathcal{X})}$$

Parametric Regression :

$$y = \underbrace{f(x)}_{\text{underlying process}} + \underbrace{\epsilon}_{\text{noise}}$$

$$x_{N+1} \rightarrow \hat{y}_{N+1} = ?$$

$$f(x_{N+1}) \rightarrow \hat{y}_{N+1} \leftarrow g(x_{N+1}|\theta)$$

$\Rightarrow$ approximate $f(x)$ with $g(x|\theta)$

$E[x]=\gamma$

$E[x+c]=\gamma+c$

$VAR[x]=\alpha^2$

$VAR[x+c]=\alpha^2$

Assumptions :

I. $p(\epsilon) \sim N(\epsilon; 0, \sigma^2)$

II. $p(y|x) \sim \underbrace{N(y; g(x|\theta), \sigma^2)}$

$$y|x \Rightarrow f(x) + \epsilon$$

$$y|x \Rightarrow \underbrace{g(x|\theta)}_{\text{constant}} + \underbrace{\epsilon}_{\text{R.V.}}$$

$$E[y|x] = E[\underbrace{g(x|\theta)}+\epsilon]$$

$$= \underbrace{E[g(x|\theta)]} + \underbrace{E[\epsilon]}$$

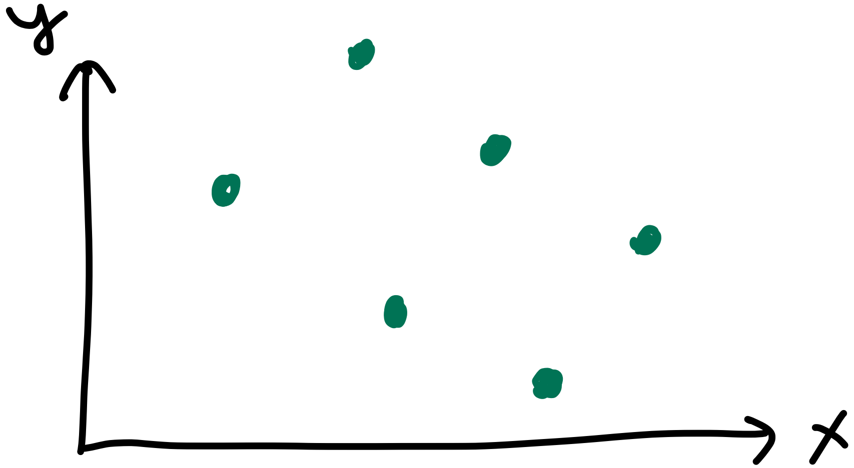
$$= g(x|\theta) + 0$$

$$VAR[y|x] = VAR[g(x|\theta)+\epsilon]$$

$$= VAR[\epsilon] = \sigma^2$$

$$x \sim N(x; 0, 9)$$

$$x+5 \sim N(x; 5, 9)$$



$$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N} \qquad x_i \in \mathbb{R}^1 \quad y_i \in \mathbb{R}^1$$

$$(x_i, y_i) \sim p(x_i, y_i)$$

↳ i.i.d.

$p(x,y)$
$= p(x|y)p(y)$
$p(x,y)$
$= p(y|x)p(x)$

$$p(x_1, y_1, x_2, y_2, \dots, x_N, y_N) = \prod_{i=1}^{N} p(x_i, y_i)$$

$$\mathcal{L}(\theta|\mathcal{X}) = \prod_{i=1}^{N} p(x_i, y_i)$$

$$= \prod_{i=1}^{N} \left[p(y_i|x_i)\, p(x_i) \right]$$

x_{N+1}

log likelihood $\Rightarrow \log \mathcal{L}(\theta|\mathcal{X}) = \log\left[\prod_{i=1}^{N} \left[p(y_i|x_i)\, p(x_i)\right]\right]$

$$= \sum_{i=1}^{N} \left[\log p(y_i|x_i) + \log p(x_i)\right]$$

constant

maximize $\sum_{i=1}^{N} \log \underbrace{p(y_i|x_i)}$ $\Rightarrow$ maximize $\sum_{i=1}^{N} \log\left[N(\overset{R.V.}{y_i}; \overset{\mu}{\boxed{g(x_i|\theta)}}, \overset{\sigma^2}{\sigma^2})\right]$

$$\frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \Rightarrow N(x;\mu,\sigma^2)$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] \Rightarrow N(y;\mu,\sigma^2)$$

R.V. $\quad \mu$

maximize $\displaystyle\sum_{i=1}^{N} \log\left[\frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(y_i - g(x_i|\theta))^2}{2\sigma^2}\right]\right]$

σ^2

constant

maximize $\displaystyle\sum_{i=1}^{N} -\frac{(y_i - g(x_i|\theta))^2}{2\sigma^2}$

constant

$g(x_i|\theta) \to$ 2nd order polynomial

$\Downarrow$

$w_0 + w_1 . x_i + w_2 . x_i^2 = \hat{y}_i$

$\theta = \{w_0, w_1, w_2\}$

minimize $\displaystyle\sum_{i=1}^{N} (y_i - \underbrace{g(x_i|\theta)}_{\hat{y}_i})^2$

$\Downarrow$

minimize $\displaystyle\sum_{i=1}^{N} (\underbrace{y_i - \hat{y}_i}_{e_i})^2 = \sum_{i=1}^{N} e_i^2$

$g(x_i|\theta) \to$ linear

$\Downarrow$

$w_0 + w_1 . x_i = \hat{y}_i$

$\theta = \{w_0, w_1\}$

minimize $\sum_{i=1}^{N}\left[y_i - g(x_i|\theta)\right]^2$

$g(x_i|\theta) = w_0 + w_1 x_i$

$\theta^* = \{w_0^*, w_1^*\}$

$$\text{Error}[\theta|\mathcal{X}] = \sum_{i=1}^{N}\left[y_i - (w_0 + w_1 . x_i)\right]^2$$

$$\frac{\partial \text{Error}}{\partial w_0} = \sum_{i=1}^{N} \cancel{2} . \left[y_i - (w_0 + w_1 x_i)\right]^1 . (\cancel{-}1) = 0$$

$$\frac{\partial \text{Error}}{\partial w_1} = \sum_{i=1}^{N} \cancel{2} \left[y_i - (w_0 + w_1 . x_i)\right]^1 (\cancel{-} x_i) = 0$$

I $\quad \sum_{i=1}^{N} y_i = \sum_{i=1}^{N} w_0 + \sum_{i=1}^{N} w_1 . x_i$

$\sum_{i=1}^{N} (y_i x_i) = \sum_{i=1}^{N} w_0 . x_i + \sum_{i=1}^{N} w_1 . x_i^2$

$$\begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} y_i \cdot x_i \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} 1 & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

$$A^{-1} b = \underbrace{A^{-1} A}_{} \theta$$

$$A^{-1} \cdot b = \theta^*$$

$$\det(A) = 2x_1^2 + 2x_2^2 - x_1^2 - 2x_1x_2 - x_2^2$$

$$= (x_1 - x_2)^2$$

$$\neq 0 \to \text{invertible}$$

when $N=2$

$$\begin{bmatrix} 2 & x_1+x_2 \\ x_1+x_2 & x_1^2+x_2^2 \end{bmatrix}$$

when $N=1$

$$A = \begin{bmatrix} 1 & x_1 \\ x_1 & x_1^2 \end{bmatrix}$$

$$\det(A) = x_1^2 - x_1 \cdot x_1 = 0$$

$\downarrow$

not invertible.

Exercise: When $N \geq 2$, A matrix is invertible. Prove this statement.

Polynomial Regression

$$g(x_i \mid w_0, w_1, \ldots, w_K) = \underbrace{w_0 + w_1 x_i + \ldots\ldots + w_K x_i^K}_{K^{th} \text{ order polynomial}}$$

!!! to many exam questions in the previous years

$K=1 \Rightarrow$ linear regression

$K=0 \Rightarrow$ constant line

$$N(w_0) = \left(\sum_{i=1}^{N} y_i\right) \Rightarrow w_0^* = \frac{\sum_{i=1}^{N} y_i}{N} = \bar{y}$$

$$\Rightarrow \underbrace{\begin{bmatrix} N & \sum_{i=1}^{N} x_i & \cdots & \sum_{i=1}^{N} x_i^K \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 & \cdots & \sum_{i=1}^{N} x_i^{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{N} x_i^K & \sum_{i=1}^{N} x_i^{K+1} & \cdots & \sum_{i=1}^{N} x_i^{2K} \end{bmatrix}}_{\substack{(K+1)\times(K+1) \\ D^T.D \Leftarrow A}} \underbrace{\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_K \end{bmatrix}}_{\substack{(K+1)\times 1 \\ \theta}} = \underbrace{\begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} (y_i x_i) \\ \vdots \\ \sum_{i=1}^{N} (y_i . x_i^K) \end{bmatrix}}_{\substack{(K+1)\times 1 \\ b}}$$

$$\underline{\underline{\theta^* = A^{-1}.b}}$$

$$D = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^K \\ 1 & x_2 & x_2^2 & \cdots & x_2^K \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_N & x_N^2 & \cdots & x_N^K \end{bmatrix}$$

Exercise: Show that

$$D^T . D = A$$

if $N < K+1$, $D^T.D$ is not invertible

if $N \geq K+1$, $D^T.D$ is invertible

$$\underbrace{\begin{array}{cccc} x_1^0 & x_2^0 & & x_N^0 \\ \left[\begin{matrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \\ \vdots & \vdots & & \vdots \\ x_1^K & x_2^K & \cdots & x_N^K \end{matrix}\right. & & & \left.\right] \end{array}}_{(K+1)\times(N)}^{D^T} \overbrace{\underbrace{\begin{bmatrix} 1 & x_1 & \cdots & x_1^K \\ 1 & x_2 & \cdots & x_2^K \\ \vdots & \vdots & & \vdots \\ 1 & x_N & \cdots & x_N^K \end{bmatrix}}_{N\times(K+1)}}^{D} = \overbrace{\underbrace{\begin{bmatrix} N & \sum_{i=1}^{N} x_i & \\ \sum_{i=1}^{N} x_i & & \\ & & \end{bmatrix}}_{(K+1)\times(K+1)}}^{A}$$

$N = ~~2~~\ 5$

$K = 3$

$$[\quad]_{4\times \cancel{2}\,5} [\quad]_{\cancel{2}\,5\times 4} = [\quad]_{4\times 4}$$

$(N=2) \Rightarrow$ rank-deficient