Maximum Likelihood Estimation (MLE)

x : training data set θ : parameters

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

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

Maximum a Posteriori Estimation (MAP)

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

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

$$= \arg\max_{\theta} \frac{p(x|\theta)\underbrace{p(\theta)}_{\text{our prior belief about } \theta}}{p(x)} \rightarrow \text{independent of } \theta$$

Parametric Regression:

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

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

$$\boxed{f(x_{N+1}) \rightarrow \hat{y}_{N+1}}$$

Learning problem: 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 $p(\epsilon) \sim N(\epsilon; 0, \sigma^2)$

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

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

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

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

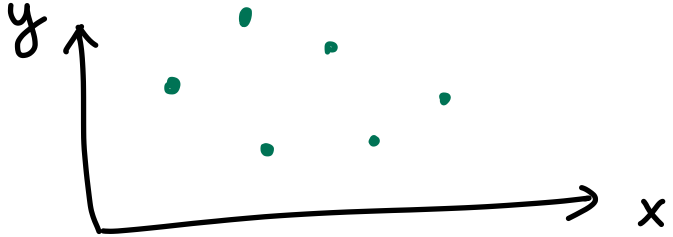
$$= g(x|\theta) + E[\epsilon] = g(x|\theta)$$

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

$$= 0 + VAR[\epsilon] = \sigma^2$$

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

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



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

$$\boxed{\begin{array}{l} \checkmark\ p(x,y) = p(y|x)\,p(x) \\ p(x,y) = p(x|y)\,p(y) \end{array}}$$

$$\underbrace{(x_i, y_i)}_{i.i.d} \sim p(x_i, y_i)$$

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

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

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

$$= \sum_{i=1}^{N} \left[\log\Big[\underbrace{p(y_i|x_i)}_{N(y_i;\, g(x_i|\theta),\, \sigma^2)}\Big] + \cancel{\log\left[p(x_i)\right]}\right] \quad \text{constant}$$

$$\stackrel{+}{=} \sum_{i=1}^{N} \left[\log\Big[N\big(\underset{\substack{\uparrow \\ \text{R.V.}}}{y_i};\, \underbrace{g(x_i|\theta)}_{\text{constant}},\, \underbrace{\sigma^2}_{\text{constant}}\big)\Big]\right]$$

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

$$N(x;\mu,\sigma^2)$$

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

cancel each other

$$\text{maximize} \sum_{i=1}^{N}\left[-\frac{[y_i-g(x_i|\theta)]^2}{\cancel{2\sigma^2}}\right]$$

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

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

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

$$w_3.\sin(x_i)+w_2\cos(x_i)+w_1.x_i^2+w_0$$

linear regression

$$\sum_{i=1}^{N}\left[y_i-[w_1 x_i+w_0]\right]^2$$

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

$$\sum_{i=1}\left[y_i-[w_2x_i^2+w_1.x_i+w_0]\right]^2$$

2nd order polynomial regression

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

minimize $\sum_{i=1}^{N} [y_i - g(x_i|\theta)]^2$ $\qquad g(x_i|\theta) = w_0 + w_1 . x_i$

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

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

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

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

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

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

$$\underbrace{\begin{bmatrix} \sum_{i=1}^{N} 1 & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 \end{bmatrix}}_{A}{}_{2\times 2} \underbrace{\begin{bmatrix} w_0 \\ w_1 \end{bmatrix}}_{\theta}{}_{2\times 1} = \underbrace{\begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} y_i x_i \end{bmatrix}}_{b}{}_{2\times 1}$$

N

Prove that A matrix is invertible when $N \geq 2$.

$$\begin{bmatrix} 2 & 3 \\ 4 & -5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ 12 \end{bmatrix}$$

$2x + 3y = 7$
$4x - 5y = 12$

$Ax = b$
$x^* = A^{-1} . b$

$N=1$

$$A = \begin{bmatrix} 1 & x_1 \\ x_1 & x_1^2 \end{bmatrix} \qquad \det(A) = x_1^2 - x_1 . x_1 = 0$$

$N=2$

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

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

Polynomiel Regression:

k^{th} order polynomiel

$$g(x_i \mid \theta) = \underbrace{w_0}_{w_0 . x_i^0} + \underbrace{w_1 . x_i}_{w_1 . x_i^1} + w_2 x_i^2 + \dots + w_k x_i^k$$

when $k=1 \Rightarrow$ lineer regression

when $k=0$

when $k=0$

$$w_0^* = \frac{\sum_{i=1}^{N} y_i}{N} = \bar{y}$$

$$\begin{bmatrix} N & \sum_{i=1}^{N} x_i & \dots & \sum_{i=1}^{N} x_i^k \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 & \dots & \sum_{i=1}^{N} x_i^{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{N} x_i^k & \sum_{i=1}^{N} x_i^{(k+1)} & \dots & \sum_{i=1}^{N} x_i^{2k} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i y_i \\ \sum_{i=1}^{N} x_i^2 y_i \\ \vdots \\ \sum_{i=1}^{N} x_i^k y_i \end{bmatrix}$$

$(k+1)\times(k+1)$ $\qquad (k+1)\times 1$ $\qquad (k+1)\times 1$