Maximum Likelihood Estimation (MLE)

$\mathcal{X}$: training data set

θ: parameters

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

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

Maximum A Posteriori Estimation (MAP)

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

$$= \arg\max_{\theta} \frac{P(\mathcal{X} \mid \theta)\,\overbrace{P(\theta)}^{\text{prior distribution}}}{P(\mathcal{X})}$$

Parametric Regression:

$y = f(x) + \epsilon \rightarrow$ noise

y → observations

$f(x)$ → underlying process

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

$p(y|x) \sim N(g(x|\theta), \sigma^2)$

Learning problem:

approximate $\boxed{f(\cdot)}$ with $\boxed{g(\cdot|\theta)}$

θ → parameters

y_1	x_1	$\hat{y}_1 = g(x_1|\theta)$
y_2	x_2	$\hat{y}_2 = g(x_2|\theta)$
$\vdots$		$\vdots$
y_N	x_N	$\hat{y}_N = g(x_N|\theta)$

$$\boxed{\begin{aligned} &E[X] = 5 \\ &E[X+3] = 8 = E[X]+3 \end{aligned}}$$

$$\boxed{\begin{aligned} &VAR[X+c] \\ &= VAR[X] + \underbrace{VAR[c]}_{0} \end{aligned}}$$

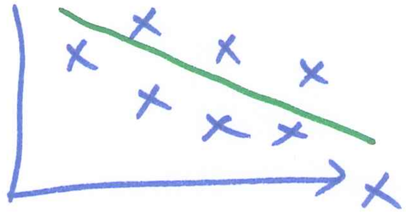
$$y = \overbrace{f(x)}^{g(x|\theta)} + \epsilon$$

$$y = \underbrace{g(x|\theta)}_{\text{constant}} + \underbrace{\epsilon}_{\text{random variable}}$$

$$E[y|x] = E[g(x|\theta) + \epsilon] = g(x|\theta) + \underbrace{E[\epsilon]}_{0} = g(x|\theta)$$

$$VAR[y|x] = VAR[g(x|\theta) + \epsilon] = \underbrace{VAR[g(x|\theta)]}_{0} + VAR[\epsilon] = \sigma^2$$

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



$$\boxed{\begin{aligned} f(a) &= 2a+5 \\ f(b) &= 2b+5 \end{aligned}}$$

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

↳ i.i.d.

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

$$\text{Log likelihood}(\theta \mid \mathcal{X}) = \log \prod_{i=1}^{N} p(x_i, y_i)$$

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

$$= \underbrace{\sum_{i=1}^{N} \log p(y_i|x_i)}_{\text{function of } \theta} + \underbrace{\cancel{\sum_{i=1}^{N} \log p(x_i)}}_{\text{constant}}$$

$$\log N(\overbrace{g(x_i|\theta)}^{\mu}, \overbrace{\sigma^2}^{\sigma^2})$$

$$\boxed{\text{maximize} \quad \sum_{i=1}^{N} \log p(y_i|x_i)}$$

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

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

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

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

maximize $L(\theta|\mathcal{X}) \Rightarrow$ maximize $\sum_{i=1}^{n} -[y_i - g(x_i|\theta)]^2$

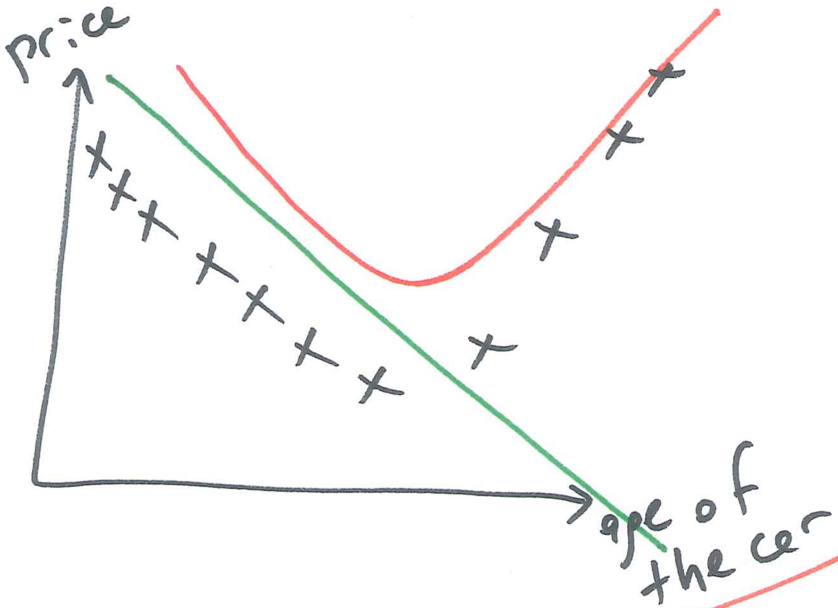
$$\Downarrow$$

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

$$\boxed{\text{minimize } \sum_{i=1}^{N}(y_i - \underbrace{\hat{y}_i}_{w x_i + w_0})^2}$$

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

$w x_i + w_0 \Rightarrow$



find w & w_0

$w_2 x_i^2 + w_1 . x_i + w_0$

$$\Downarrow$$

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

$$ERROR[\theta|\mathcal{X}] = \sum_{i=1}^{N}\left(y_i - \overbrace{[w_1 x_i + w_0]}^{\hat{y}_i}\right)^2$$

$$\frac{\partial \, \mathrm{ERROR}[\theta \mid \mathcal{X}]}{\partial w_0} \Rightarrow \sum_{i=1}^{N} y_i = N \boxed{w_0} + \boxed{w_1} \sum_{i=1}^{N} x_i$$

$$\frac{\partial \, \mathrm{ERROR}[\theta \mid \mathcal{X}]}{\partial w_1} = \sum_{i=1}^{N} x_i y_i = \boxed{w_0} \sum_{i=1}^{N} x_i + \boxed{w_1} \sum_{i=1}^{N} (x_i)^2$$

$$Ax = b$$

$$x = A^{-1} \cdot b$$

Exercise #2

Show that is invertible

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

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

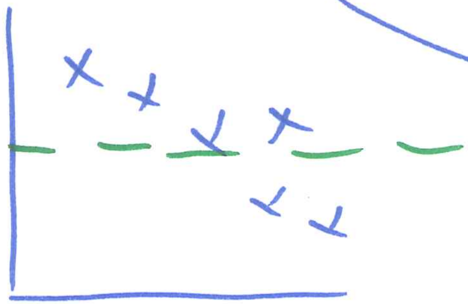
Polynomial Regression:

$$g(x_i \mid w_k, w_{k-1}, \ldots, w_2, w_1, w_0) \Rightarrow w_k x_i^k + w_{k-1} x_i^{k-1} + \ldots + w_2 x_i^2 + w_1 x_i + w_0$$

$$\underbrace{\begin{bmatrix} \overset{\sum x_i^0}{N} & \sum x_i^1 & \sum x_i^2 & \cdots & \sum x_i^k \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \cdots & \sum x_i^{k+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_i^k & \sum x_i^{(k+1)} & \sum x_i^{(k+2)} & \cdots & \sum x_i^{(2k)} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_k \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} \sum x_i^0 y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \\ \vdots \\ \sum x_i^k \cdot y_i \end{bmatrix}}_{b}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

$$\Rightarrow x = A^{-1} \cdot b$$



$$\begin{bmatrix} \sum x_i^0 \end{bmatrix} \begin{bmatrix} w_0 \end{bmatrix} = \begin{bmatrix} \sum x_i^0 y_i \end{bmatrix}$$

$$w_0 = \frac{\sum y_i}{\underbrace{\sum x_i^0}_{N}} = \frac{\sum y_i}{N}$$

$$D = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^k \\ 1 & x_2 & x_2^2 & \cdots & x_2^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & x_N^2 & \cdots & x_N^k \end{bmatrix}_{N \times (k+1)}$$

$$\boxed{A = \underbrace{D^T}_{(k+1) \times N} \cdot \underbrace{D}_{N \times (k+1)}}$$

$$\boxed{b = D^T \cdot y}$$

$A \cdot x = b \Rightarrow (D^T \cdot D) \cdot x = D^T \cdot y$

$$\boxed{x^* = (D^T \cdot D)^{-1} \cdot D^T \cdot y}$$

it is invertible if

$N \geq (k+1)$