Linear Regression

Bayesian Inference
&
Maximum likelihood Estimate

$\vec{\Theta}$: Θ_1 , Θ_2

slope — Intercept

$\vec{\Theta}$: $\underbrace{\Theta_1, \ldots, \Theta_{n-1}}_{\text{coefficients}}, \Theta_n$

Intercept

$\vec{\Theta}$
- OLS : 2.1
- GDA : 2.2
- MLE : 2.3

prior

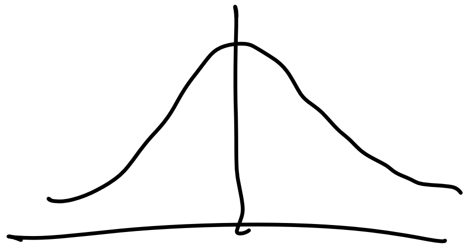
$$p(\vec{\Theta} \mid \vec{x}) = \frac{p(\vec{x} \mid \vec{\Theta}) \times p(\vec{\Theta})}{p(\vec{x})}$$

likelihood — evidence

Baye's theorem

Θ : ✗ point estimate

$\vec{\Theta}$: $p(\vec{\Theta} \mid \vec{y}, \vec{x})$ =



Distribution

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

$$p(\vec{x} \mid \vec{\theta}) = \mathcal{L}(\vec{\theta}; \vec{x})$$

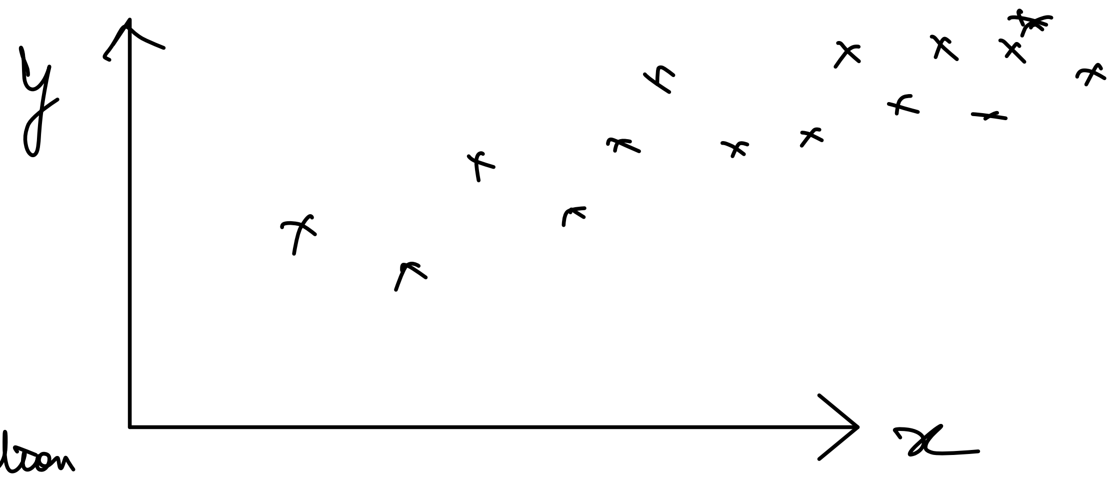
1. Distribution → parameters

Linear Regression:

$\vec{\theta} : (\theta_1, \theta_2)$

$(\underbrace{\theta_1, \dots, \theta_{n-1}}_{\text{coeff}}, \underbrace{\theta_n}_{\text{Intercept}})$

$$\vec{\hat{Y}} = \vec{\theta}[X]$$



Assumption

① y : normal distribution

$$\left(\mu = \vec{\theta} X, \quad \sigma_y^2 = \sigma_\varepsilon^2 \right)$$

$$Y = \vec{\theta} X + \varepsilon_e$$

$$\varepsilon_e : \mathcal{N}(0, \sigma^2)$$

$$E(Y) = E(\vec{\theta} X) + E(\varepsilon_e)$$

$$E(Y) = E\underbrace{(\vec{\theta} X)} + 0$$

$$E(Y) = \vec{\theta} X$$

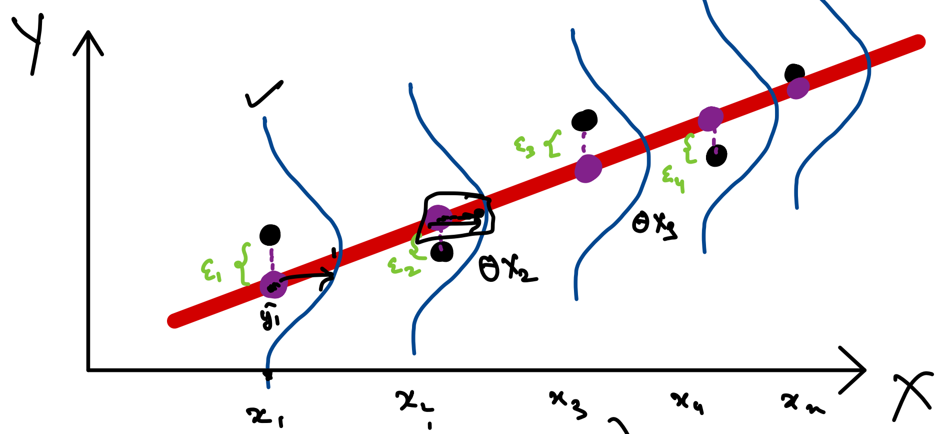
$$\boxed{\mu_Y = \vec{\theta} X}$$

$$Y = \theta X + \varepsilon_e$$

$$Var(Y) = \underbrace{Var(\theta X)} + Var(\varepsilon_e)$$

$$Var(Y) = 0 + \sigma_\varepsilon^2$$

$\mu_Y = \theta X$



at x_1 $\boxed{\mu_Y = \theta X_1}$

$$p(\vec{y} \mid \vec{\theta}) = \mathcal{L}(\vec{\theta}, \sigma^2 \mid \vec{X})$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} e^{-\frac{(y_i - \vec{\theta}x)^2}{2\sigma^2}}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\right)^n e^{-\sum_{i=1}^{n} \frac{(y_i - \vec{\theta}x)^2}{2\sigma^2}}$$

$$\ln(L(\vec{y} \mid \vec{\theta})) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{(Y - \theta X)^T (Y - \theta X)}{2\sigma^2}$$

$$\frac{\partial LL(LL(\vec{\theta}))}{\partial \vec{\theta}} = 0$$

$$\frac{1}{2\sigma^2}\left(0 - 2X^T Y + 2X^T X\vec{\theta}\right) = 0$$

$$\Rightarrow \boxed{\vec{\theta} = (X^T X)^{-1} X^T Y}$$

$$X = \underbrace{\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & \vdots & & & \\ 1 & \vdots & & & \\ 1 & \vdots & & & \\ \cdot & x_{n1} & \cdots & \cdots & x_{nk} \end{bmatrix}}_{(1+k)\text{ parameters}} \; \Big\downarrow n \text{ samples}$$

$\boxed{\text{OLS}}$: Vector

AIC
BIC

model
complexity with accuracy

n ; $LL(\vec{\theta})$