

Bayes Rule

$P(\theta|D)$
↓
posterior

$$= \frac{\overset{\text{Likelihood}}{P(D|\theta)} \cdot \overset{\text{Prior}}{P(\theta)}}{P(D)}$$

COIN TOSS

$$D = \{H, H, T, \dots\}$$

Problem: $P(H) = ?$

$$P(H) = \theta$$

$$P(T) = 1 - \theta$$

Bernoulli Distributionⁿ

$$P(D_1 = H) = \theta$$

$$P(D_1 = T) = 1 - \theta$$

$$P(D)$$

$$P(D_1 = H) = \theta$$

$$P(D_1 = T) = 1 - \theta$$

$$P(D_1 = x) = \theta^x (1 - \theta)^{1-x}$$

$$P(D_1 = x) = \begin{cases} \theta; & x=1 \text{ (H)} \\ 1-\theta; & x=0 \text{ (T)} \end{cases}$$

$$P(D|\theta) = ? = F(\theta)$$

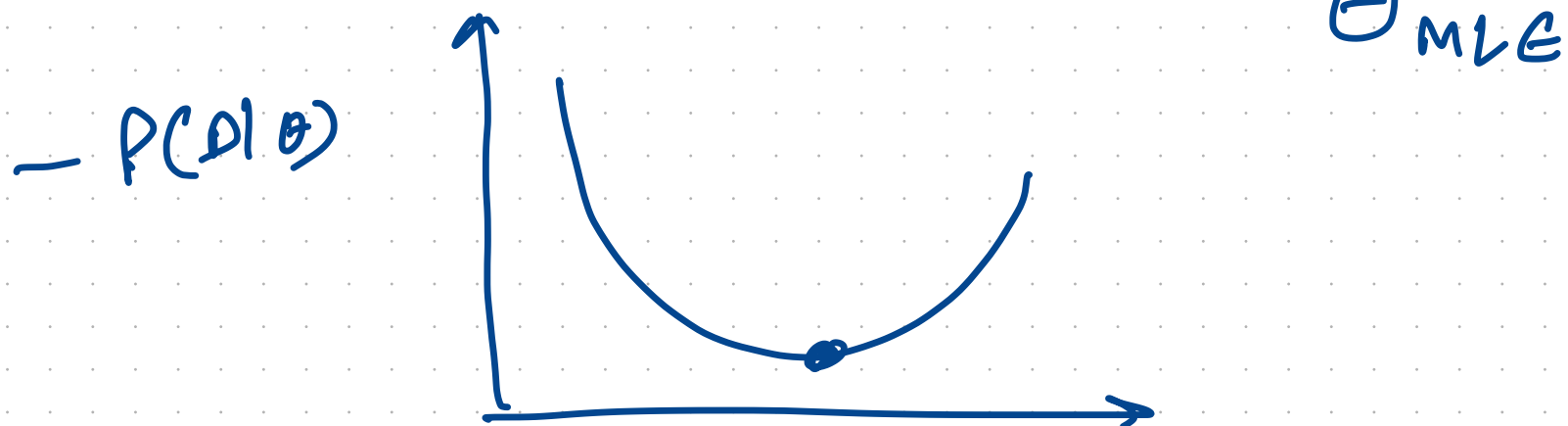
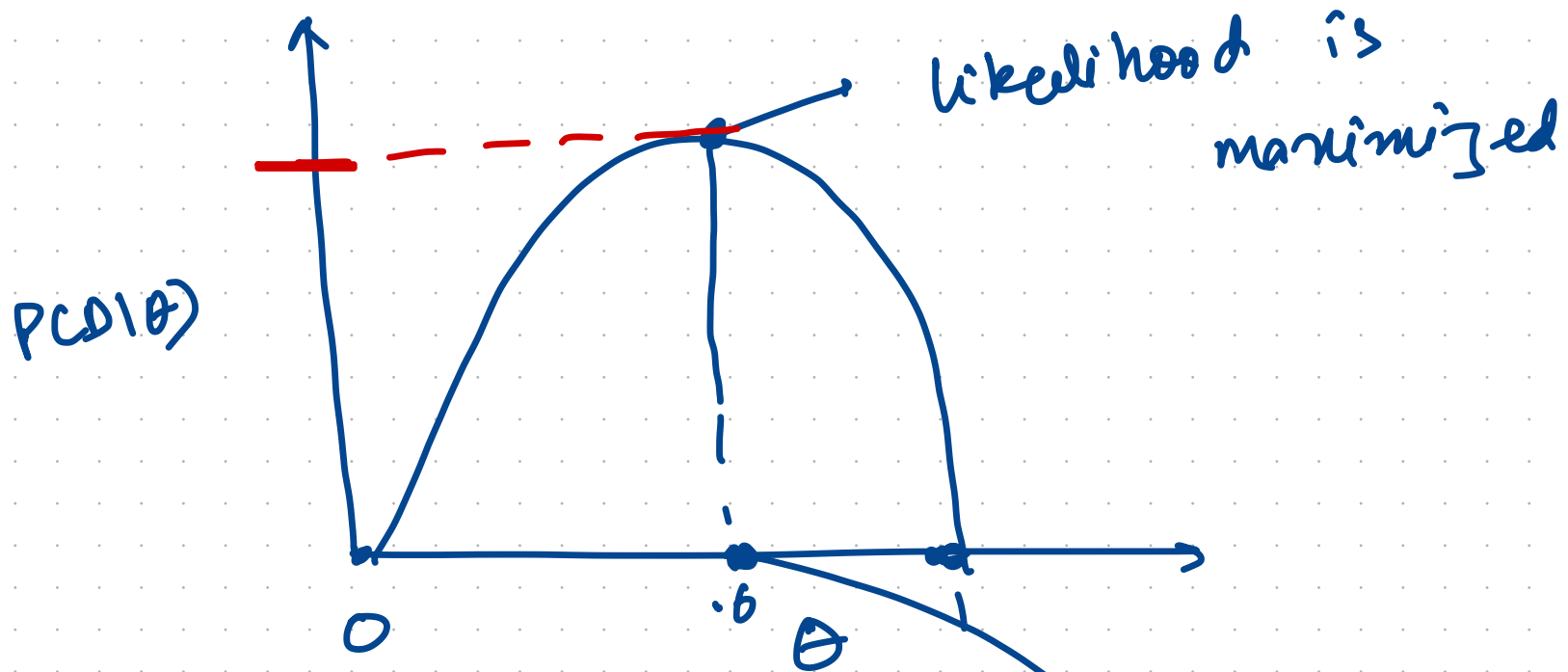
$$D = \{H, H, T, T, H\}$$

$$P(D|\theta) = P(D_1|\theta) \cdot P(D_2|\theta) \dots P(D_5|\theta)$$

$$= \theta \cdot \theta \cdot (1-\theta) \cdot (1-\theta) \cdot \theta$$

$$= \theta^3 (1-\theta)^2$$

$$P(D|\theta) = \theta^{n_H} (1-\theta)^{N - n_H}$$



$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} P(D|\theta) =$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmin}} (-P(D|\theta))$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmin}} -\log P(D|\theta)$$

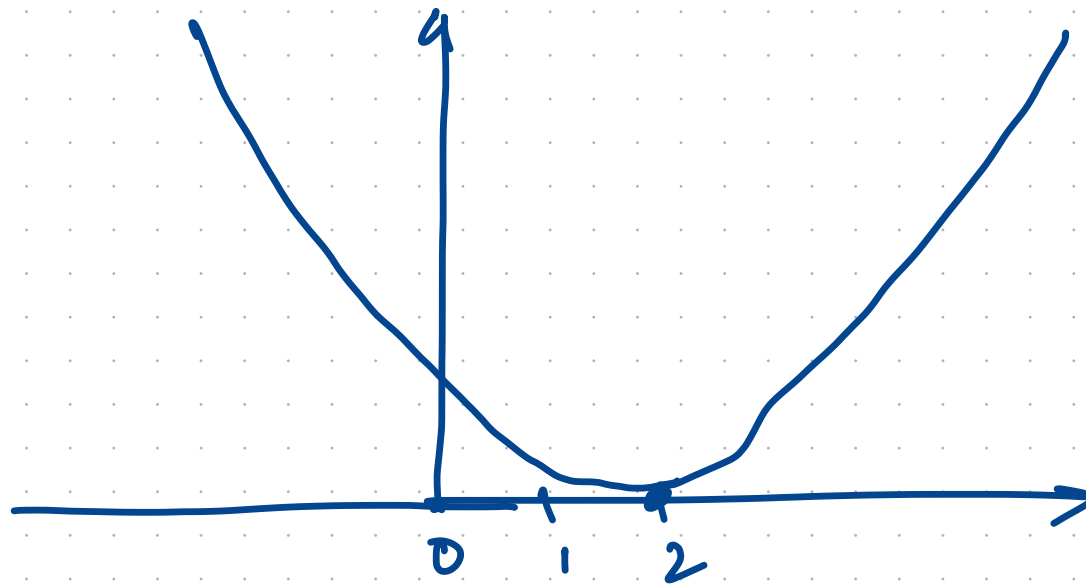
$$= \underset{\theta}{\operatorname{argmin}} \text{NLL}(\theta)$$

(log is
monotonically
increasing
)

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmin}} J(\theta) \rightarrow \text{COST FUNCTION}$$

You're minimizing

$$f(\theta) = (\theta - 2)^2$$



$$\min f(\theta) = ? \quad = 0$$

$$\arg\min f(\theta) = ? \quad = 2$$

$$D = \{n, \dots, n\}$$
$$= 10n, 0T$$

$$\hat{\theta}_{MLE} = \frac{nH}{N} = \frac{10}{10} = 1$$

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

Likelihood
 $= f_1(\theta)$

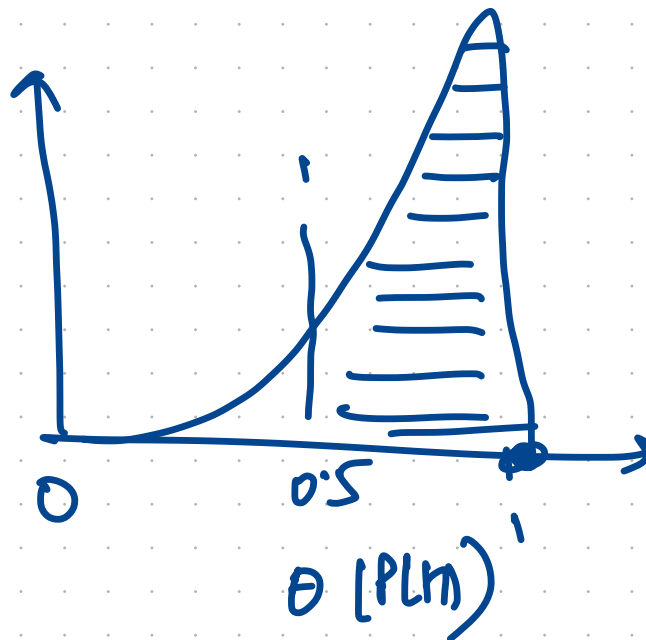
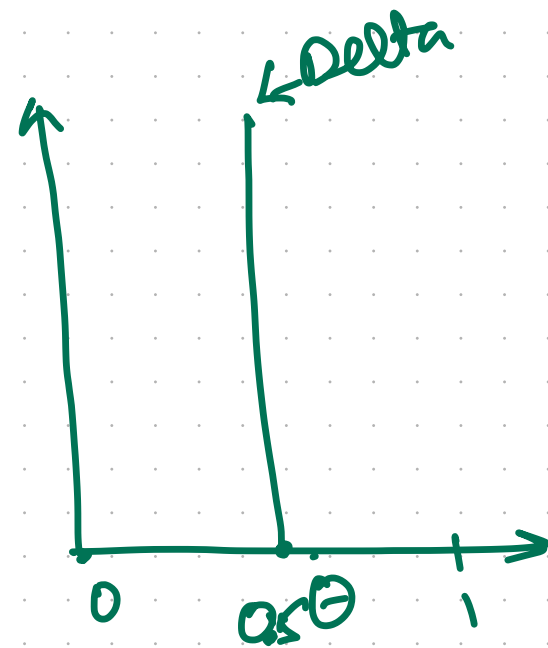
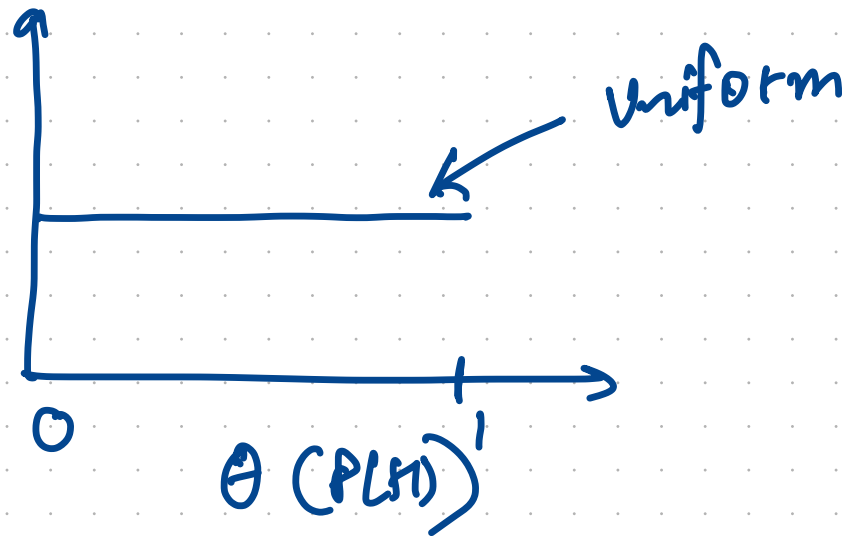
Posterior
 $f_2(\theta)$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} P(D|\theta)$$

$$\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}} P(\theta|D)$$

Maximum A-posteriori

$$\theta = PLH)$$



$$J(\theta) = \text{NLL}(\theta) + \lambda f(\theta)$$

$$f(\theta) = \theta^T \theta$$

$$\text{OR } \|\theta\|$$

Penalty to
reduce
overfitting

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_M \end{bmatrix}$$

$$\min J(\theta) + \lambda(\theta_0^2 + \theta_1^2 + \dots + \theta_M^2)$$

$$\min. J(\theta)$$

$$\text{s.t. } \theta_0^2 + \theta_1^2 + \dots + \theta_M^2 \leq \delta^2$$

$$\theta = \begin{bmatrix} \theta_0 \\ \vdots \\ \theta_m \end{bmatrix} \quad m \times 1$$

$$\theta^T = [\theta_0 \quad \dots \quad \theta_m] \quad 1 \times m$$

$$\boxed{\theta^T \theta = \sum_{i=0}^m \theta_i^2}$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmin}} \operatorname{NLL}(\theta)$$

$$= \underset{\theta}{\operatorname{argmin}} -\log \operatorname{PLD}(\theta)$$

$$p(\theta) \sim \mathcal{N}(\mu, \sigma)$$

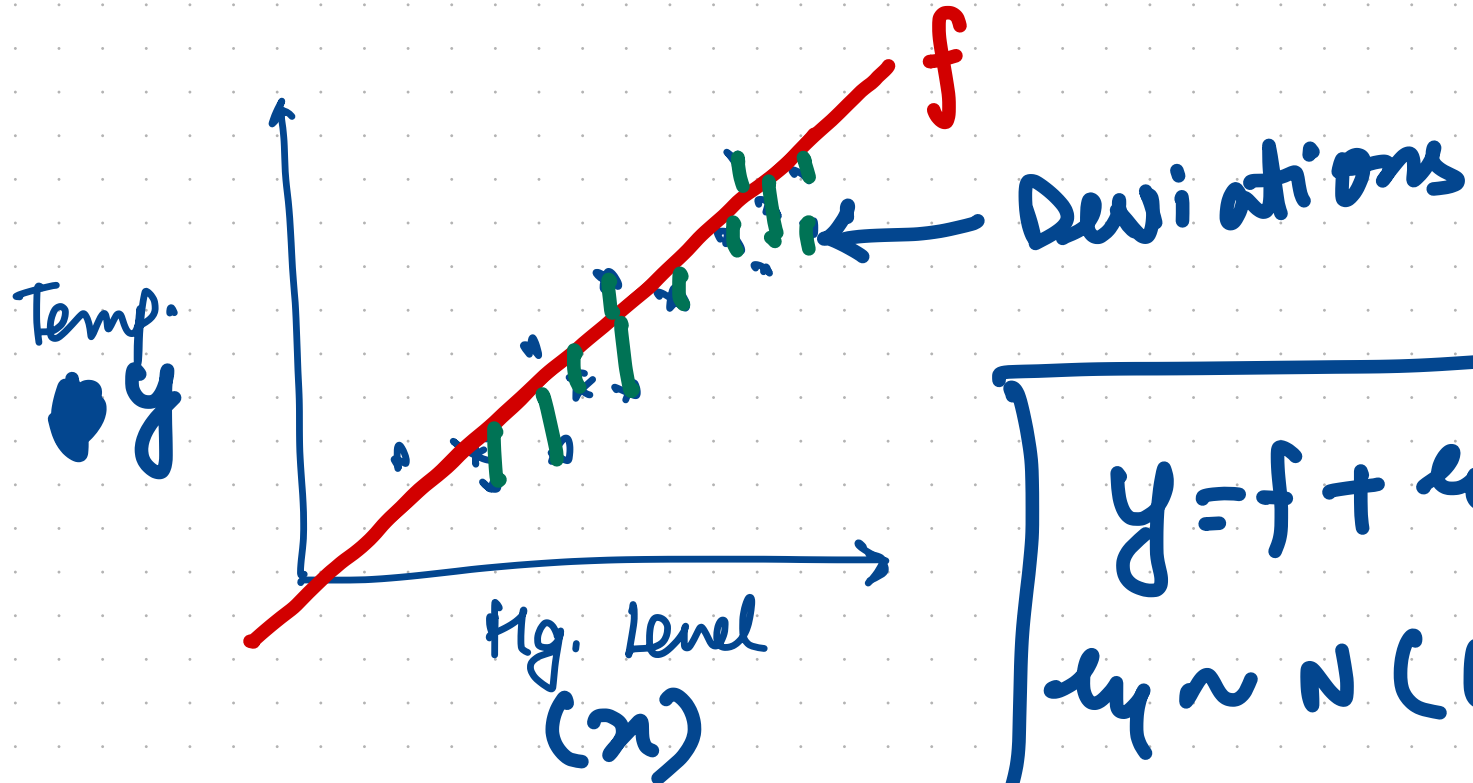
$$\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmin}} \left(\operatorname{NLL}(\theta) + N \cdot \log \operatorname{Prior}(\theta) \right)$$

$$= \underset{\theta}{\operatorname{argmin}} J(\theta) + \lambda \theta^T \theta (L_2)$$

$$\quad \quad \quad \propto \times \|\theta\|_1 (L_1)$$

$$p(\theta) \sim \operatorname{Lap}(\dots)$$

Linear Regression



$$\begin{aligned}\text{Temp.} &= f(\text{Hg}) \\ &= 4 + 3 * \text{Hg}\end{aligned}$$

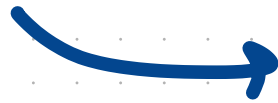
$$f = 4 + 3x$$

Given x

Calculate $f(x) = 4 + 3x$

$\therefore y(x) = f(x) + 4$

$x \sim N(0, 6)$



Sample

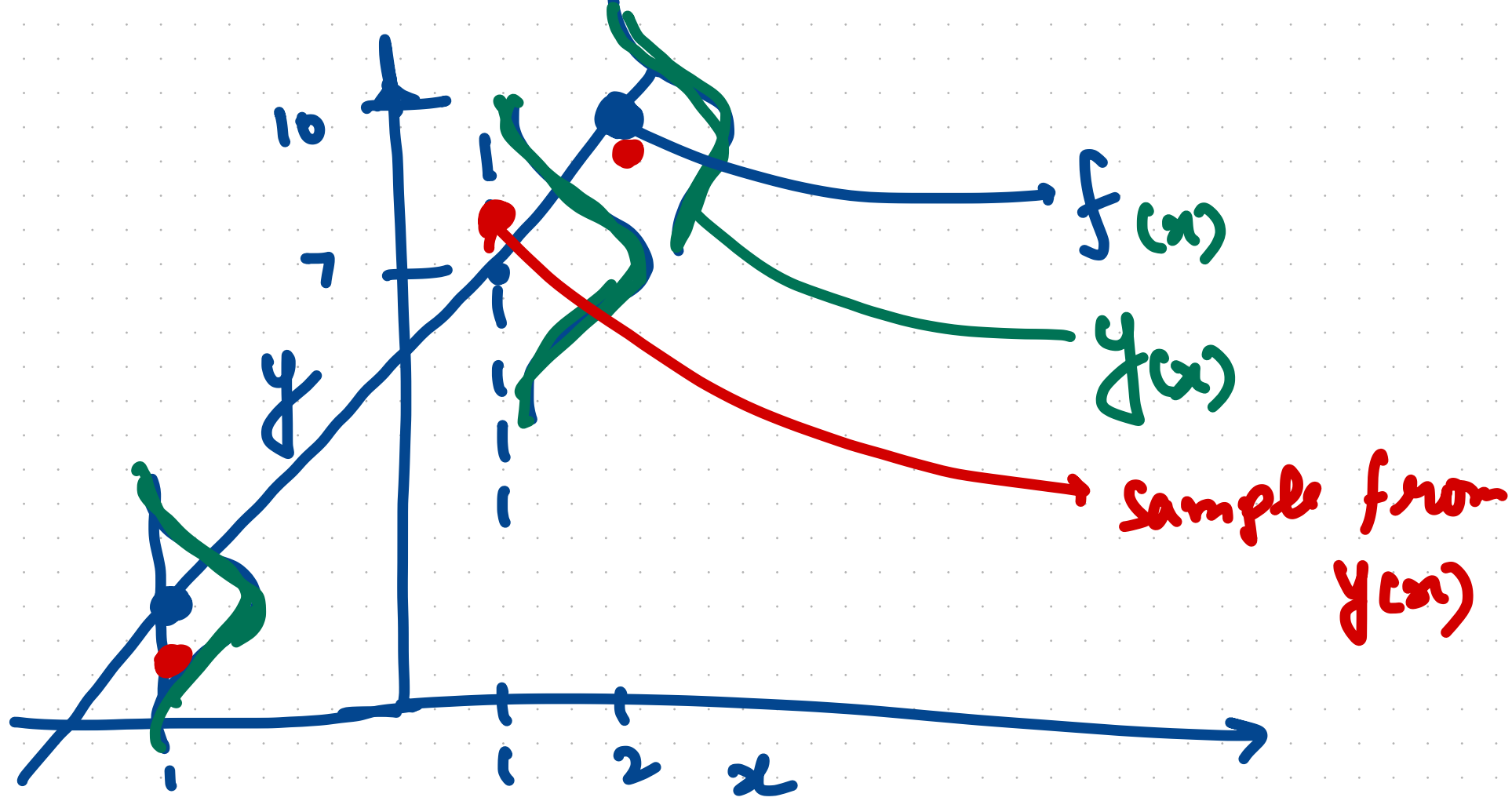
$\hookrightarrow 0.3$

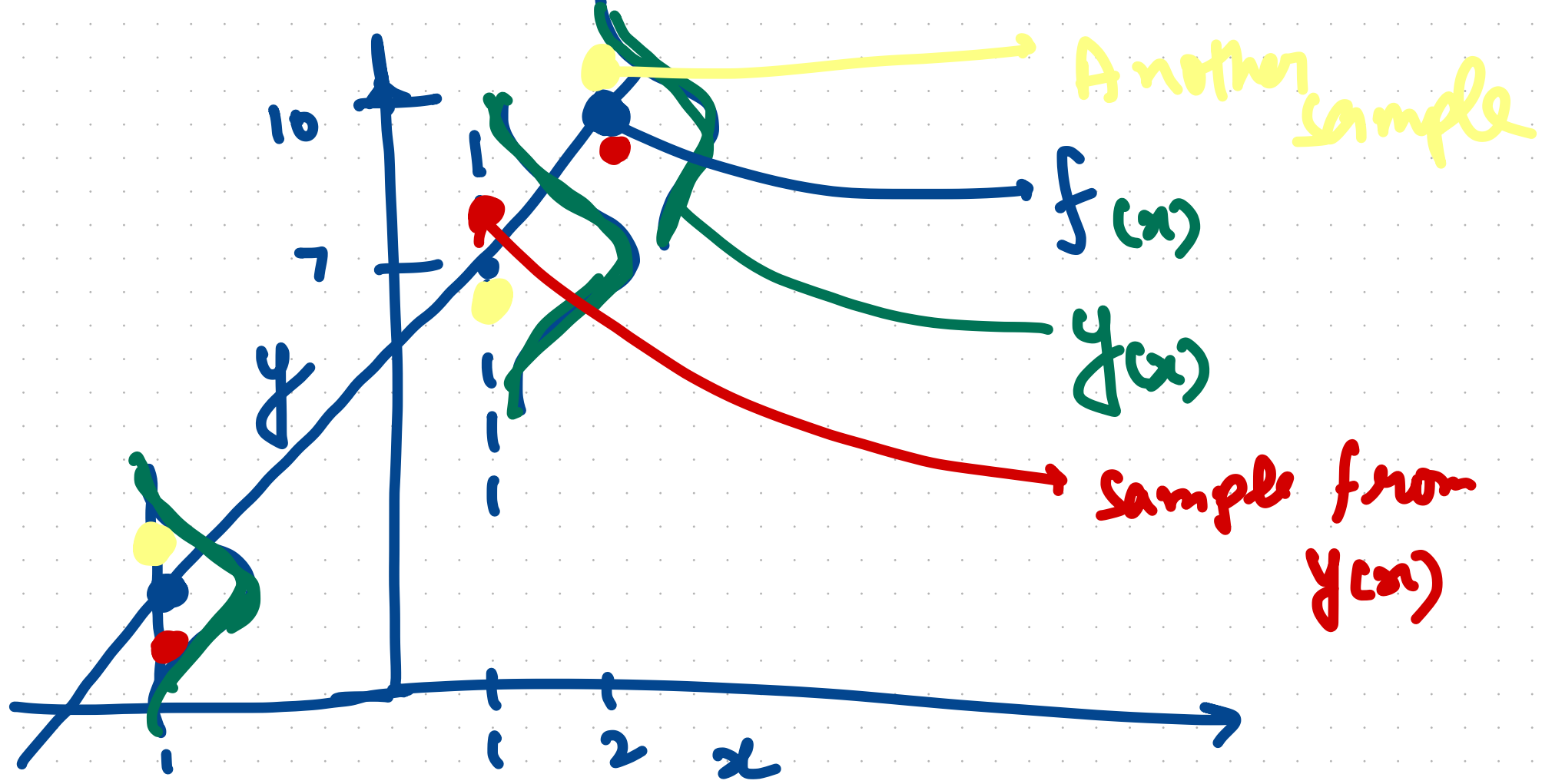
$\hookrightarrow -0.6$

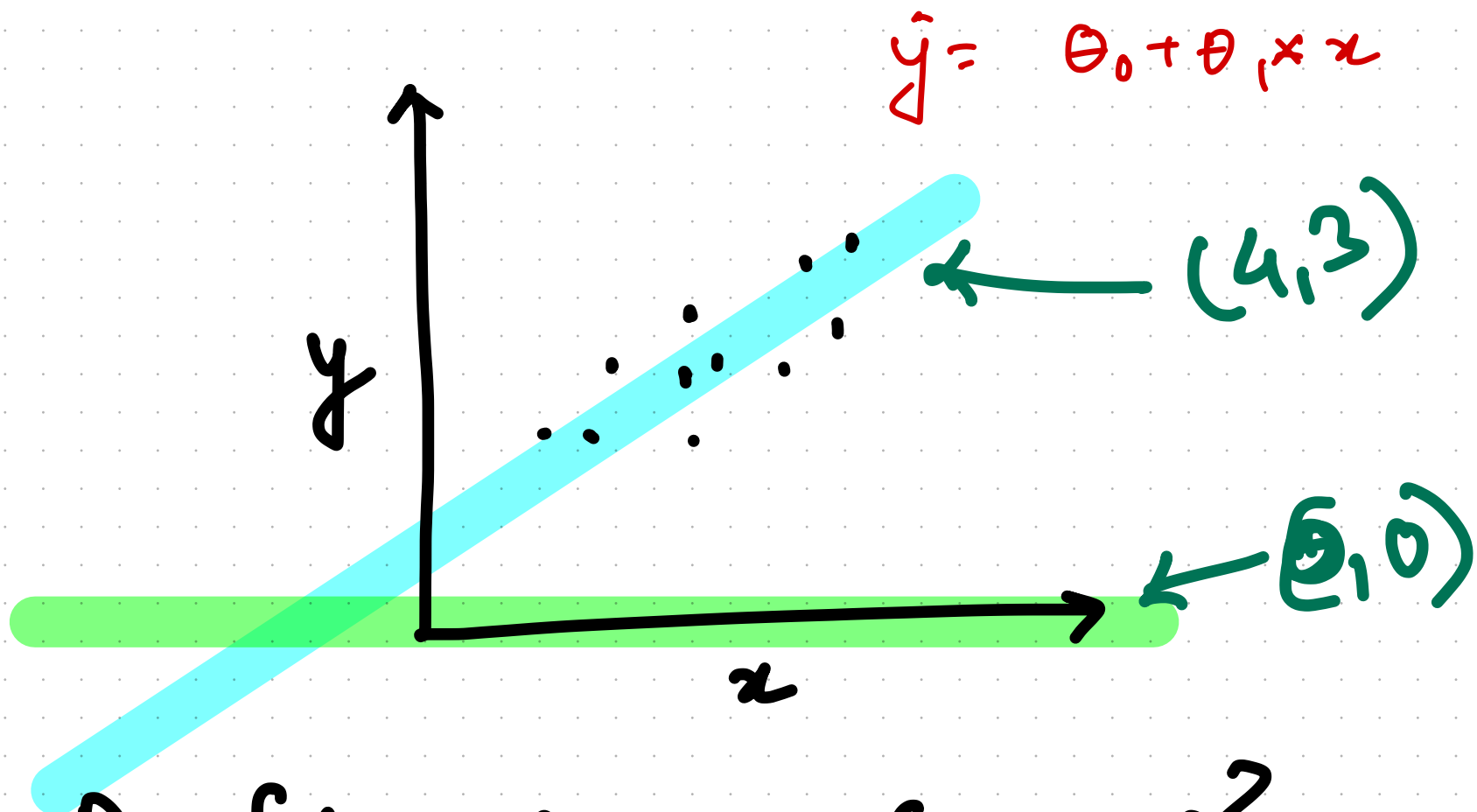
$$f = 4 + 3x$$

$$u \sim N(0, \sigma)$$

$$y_{f(x)} \sim N(f(x), \sigma)$$







$$D = \{(x_1, y_1), \dots, (x_N, y_N)\}$$

$$\text{Parameters} = \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

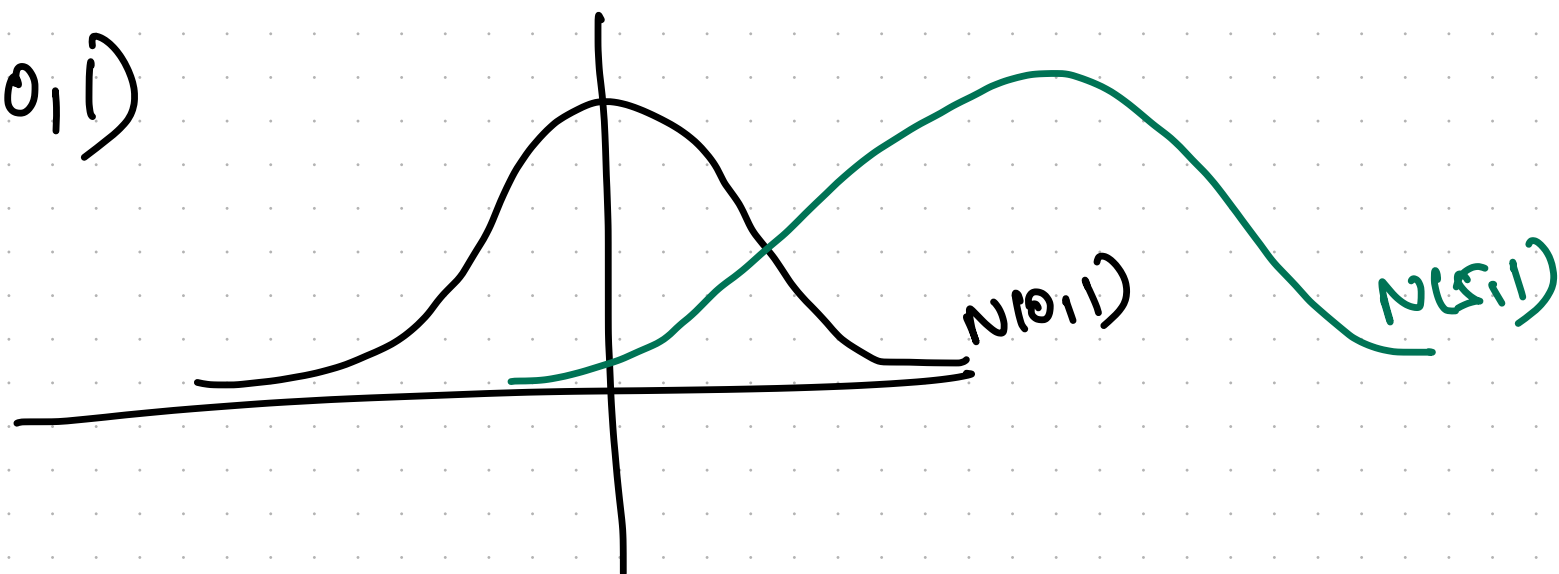
Goal

Find $\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$ s.t. $P(D|\theta)$ is maximized.

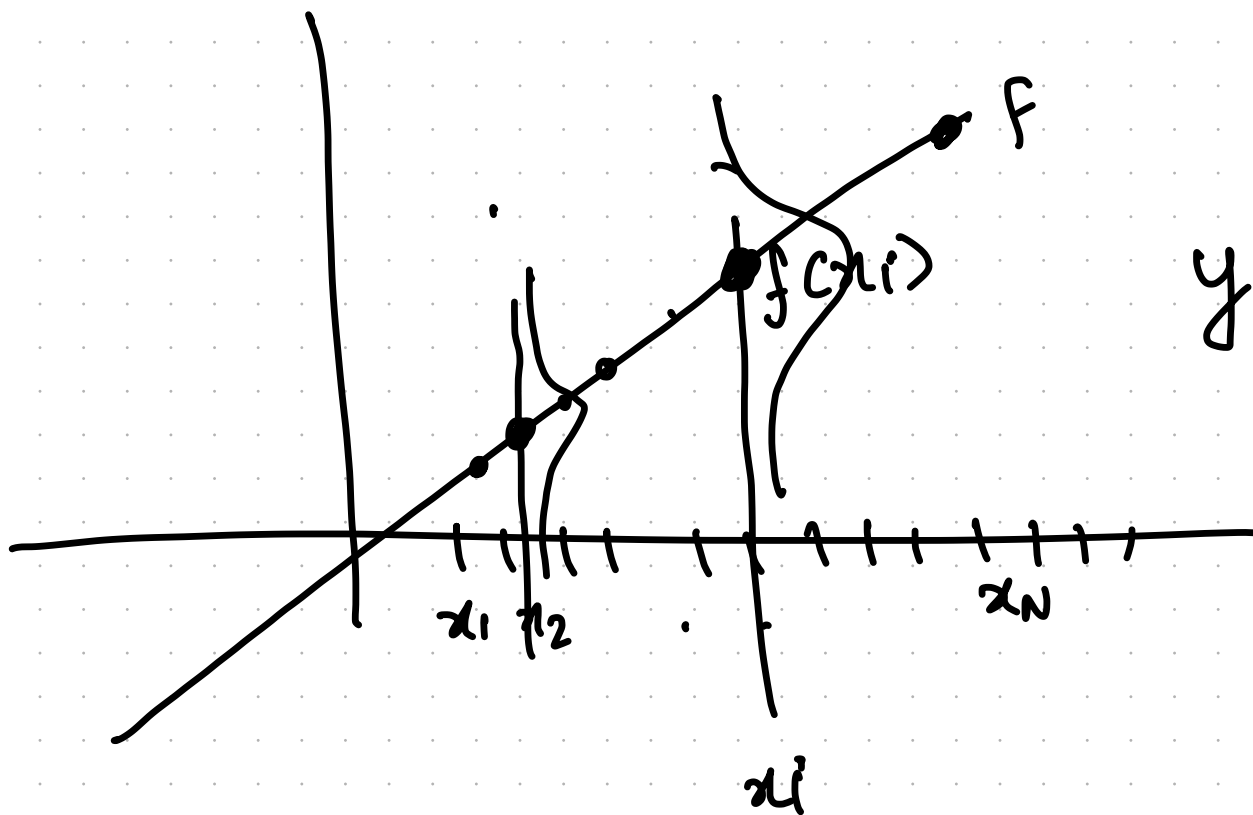
$$P(D|\theta) = \underbrace{P(D_1|\theta)}_{\text{I.I.D.}} \cdot P(D_2|\theta) \cdots P(D_n|\theta)$$

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} P(D|\theta)$$

$$X \sim N(0, 1)$$



$$X \sim N(5, 1)$$



$$y_i \sim N(f(x_i), \sigma^2)$$

$$P(D|\theta) = ?$$

$$D = \{(x_1, y_1), \dots, (x_n, y_n)\}$$

$$P(D|\theta) = P(y|x, \theta)$$

$$= P(y_1, y_2, \dots, y_n | x_1, x_2, \dots, x_n, \theta)$$

$$= P(y_1 | x_1, \theta) \cdot P(y_2 | x_2, \theta) \cdots P(y_n | x_n, \theta)$$

$$P(\theta) = P(y_1 | x_1, \theta) \cdot P(y_2 | x_2, \theta) \dots P(y_N | x_N, \theta)$$

$$P(y_1 | x_1, \theta)$$

$$y_1 \sim N(f(x_1), \sigma^2)$$

$$f(x) = ?$$

$$= \theta_0 + \theta_1 x$$

$$= \begin{pmatrix} x' & 1 \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}$$

Equal

$$\begin{bmatrix} 1 & x \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}_{2 \times 1}$$

$$x' = \begin{bmatrix} 1 \\ x \end{bmatrix}_{2 \times 1}$$

