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

$$= \frac{P(D,\theta)}{\int P(D,\theta)\, d\theta}$$

COIN TOSS

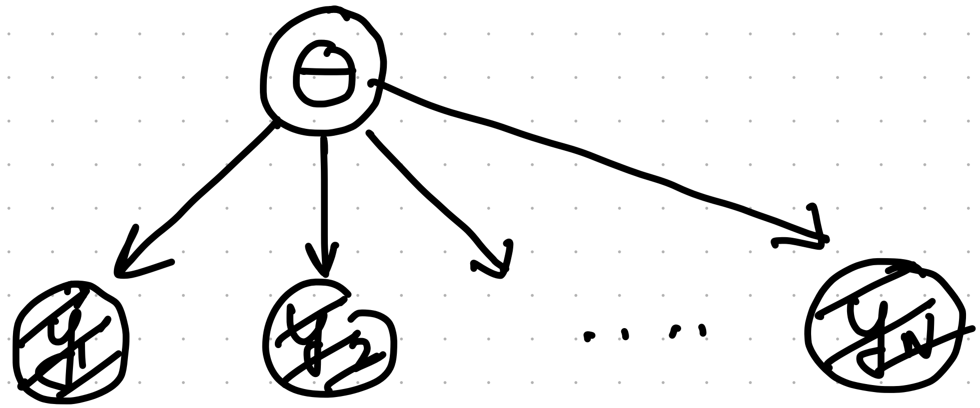
$$D = \{y_1, y_2, \ldots\ldots y_N\}$$

$$y_i = \begin{cases} 1 ; H \\ 0 ; T \end{cases}$$

$$P(y_i \mid \theta) = \theta^{y_i} (1-\theta)^{1-y_i}$$

$$y_i \sim \text{Bernoulli}(\theta)$$

# Graphical Model




SHADED : observed

: Unshaded circle : Random

a : Scalars

# Graphical Model

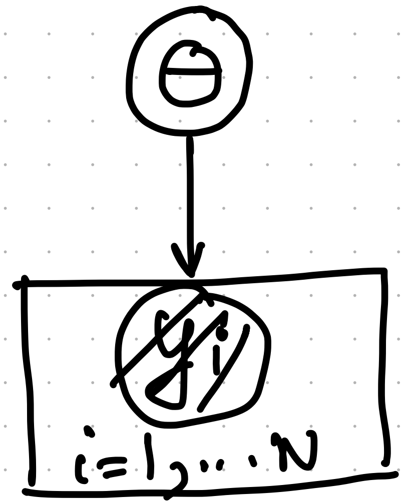



Plate Notation

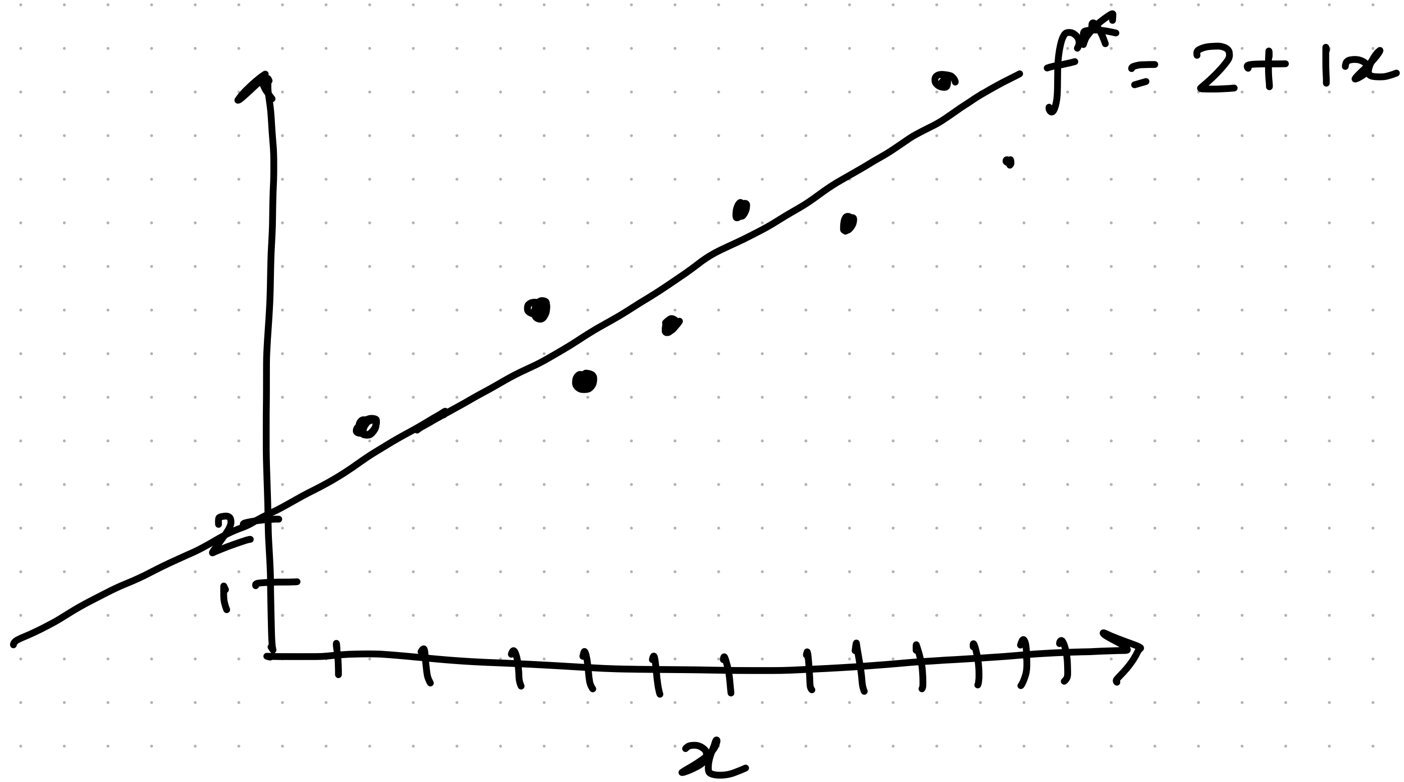
SHADED : observed

: Unshaded circle : Random

a: Scalers | Known Values

: Plate

# Linear Regression

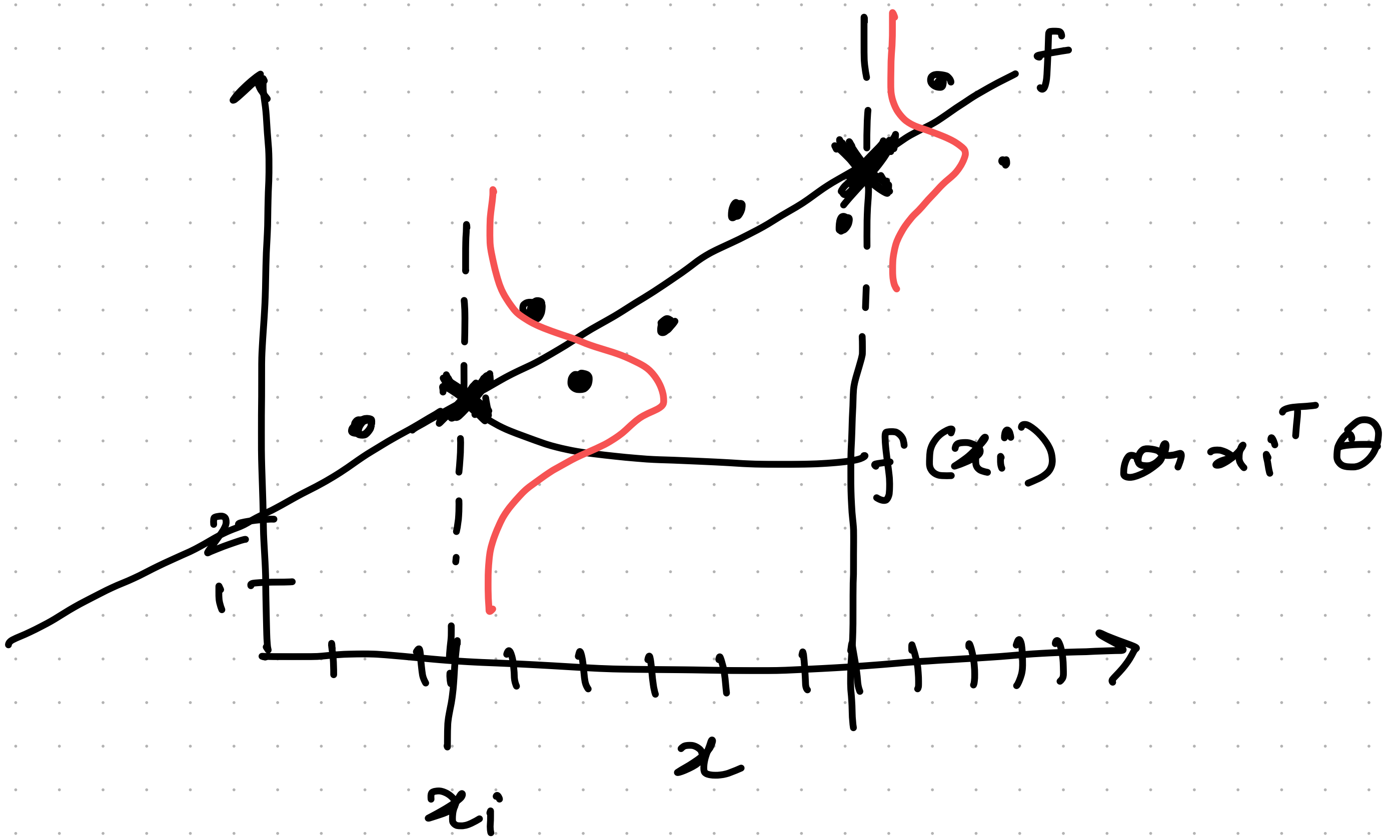



$f^*$ = TRUE FUNC$^n$ = $2 + 1x$

$y_{OBS} = f^* + e_y$

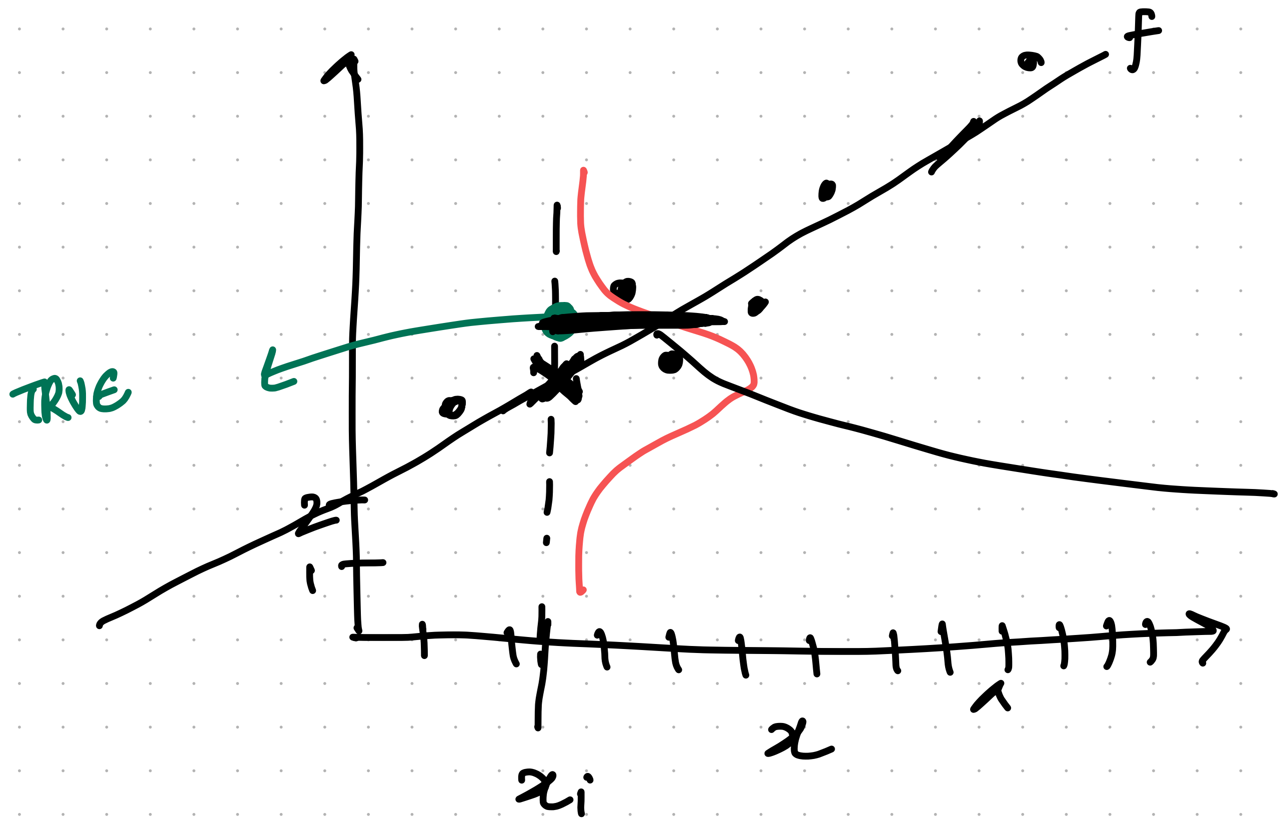
$e_y \sim N(0, \sigma^2)$

# Linear Regression



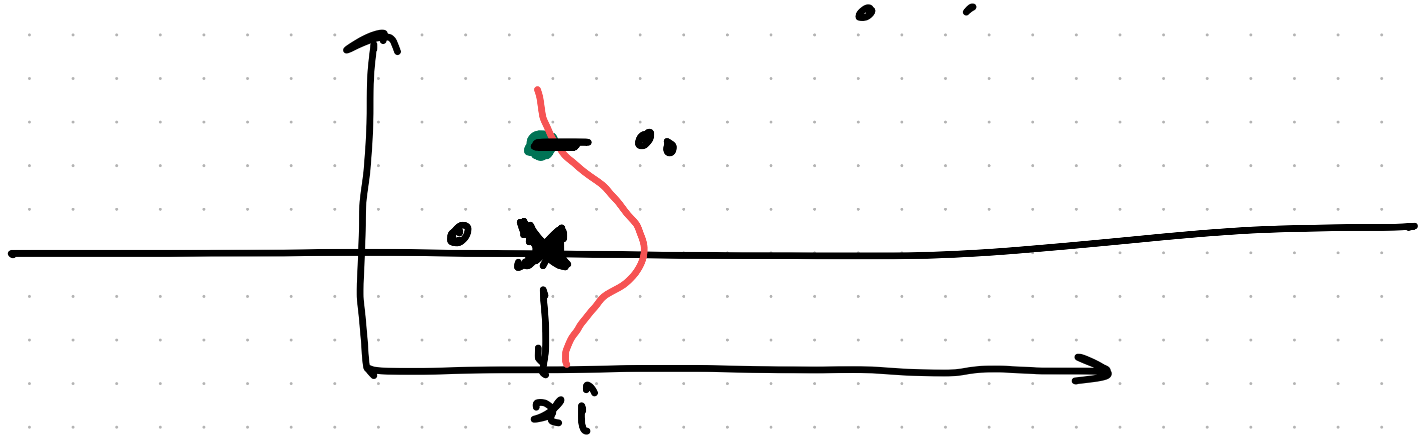

$$\boxed{\begin{gathered} y_i \sim N(f(x_i), \sigma^2) \\ f(x_i) = x_i^T\theta \end{gathered}}$$

# Linear Regression




$$\theta = \begin{bmatrix} \theta_A \\ \theta_B \end{bmatrix}$$

$$P(y_i \mid x_i^T\theta, \sigma^2)$$




$$\theta = \begin{bmatrix} \theta_c \\ \theta_d \end{bmatrix}$$

Problem:

Learn $\Theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$

$$f = \theta_0 + \theta_1 * x$$

$$y_{OBS} = f + e_y$$
$$= \theta_0 + \theta_1 x + N(0, \sigma^2)$$

$$y_{OBS} \sim N(\theta_0 + \theta_1 x, \sigma^2)$$

$$\boxed{y \sim N(x^T \Theta, \sigma^2)}$$

## General

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_d \end{bmatrix} \; ; \quad x = \begin{bmatrix} \cancel{x_0} \to 1 \\ \vdots \\ x_d \end{bmatrix}$$

$$\theta_0 x_0 + \theta_1 x_1 + \dots \quad \theta_d x_d$$

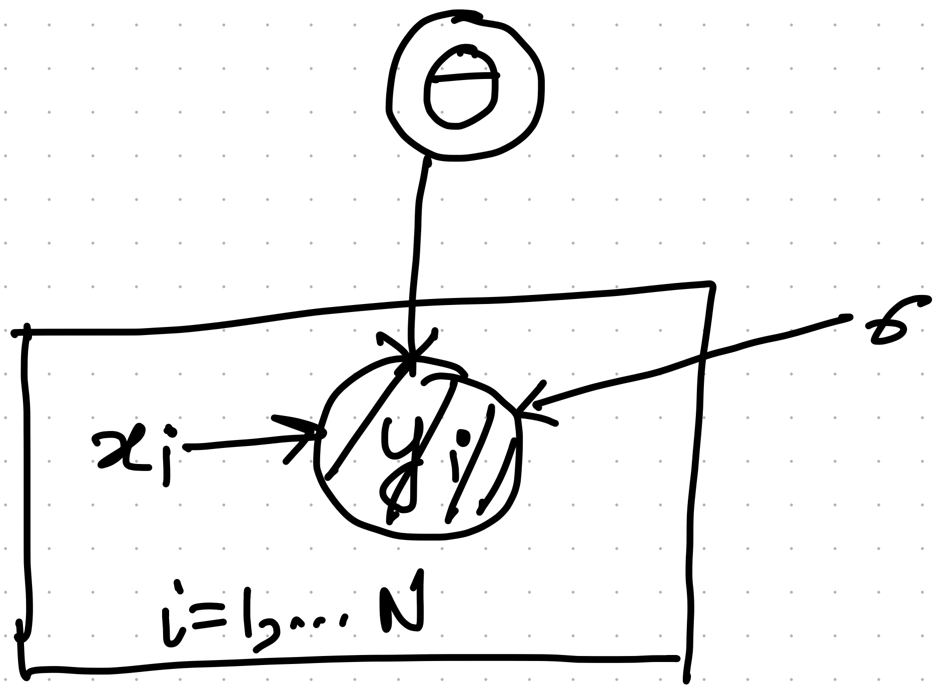
$$x^T \theta$$

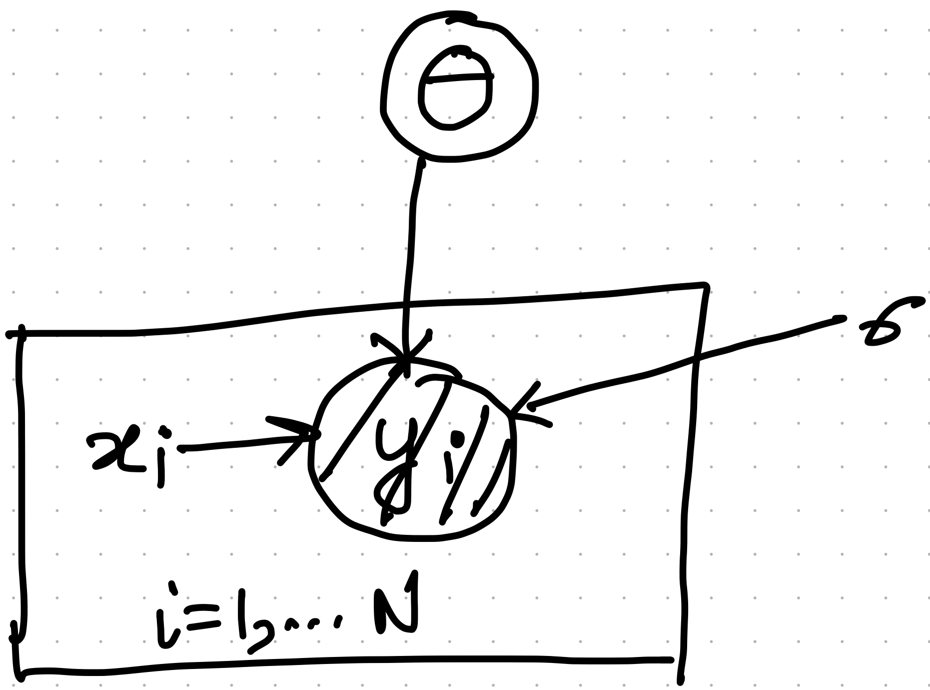
$$x^T \theta = \theta_0 * 1 + \theta_1 * x_1 + \dots \quad \theta_d * x_d$$

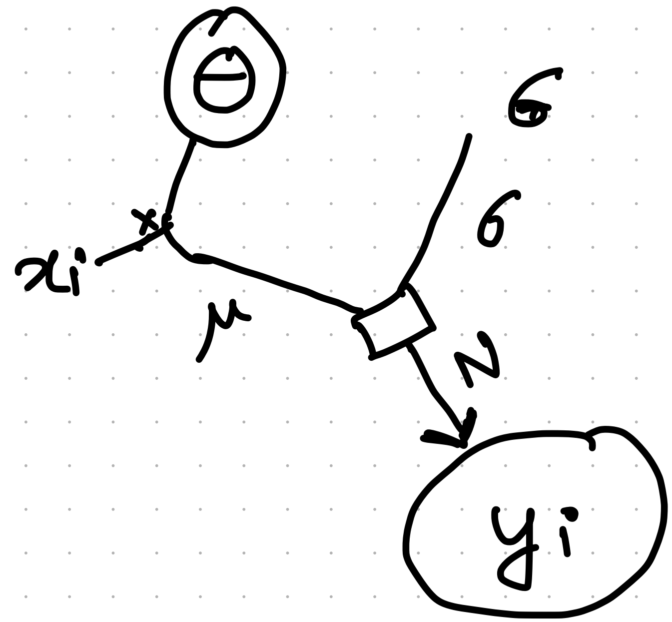
Likelihood: $P(D \mid \theta)$

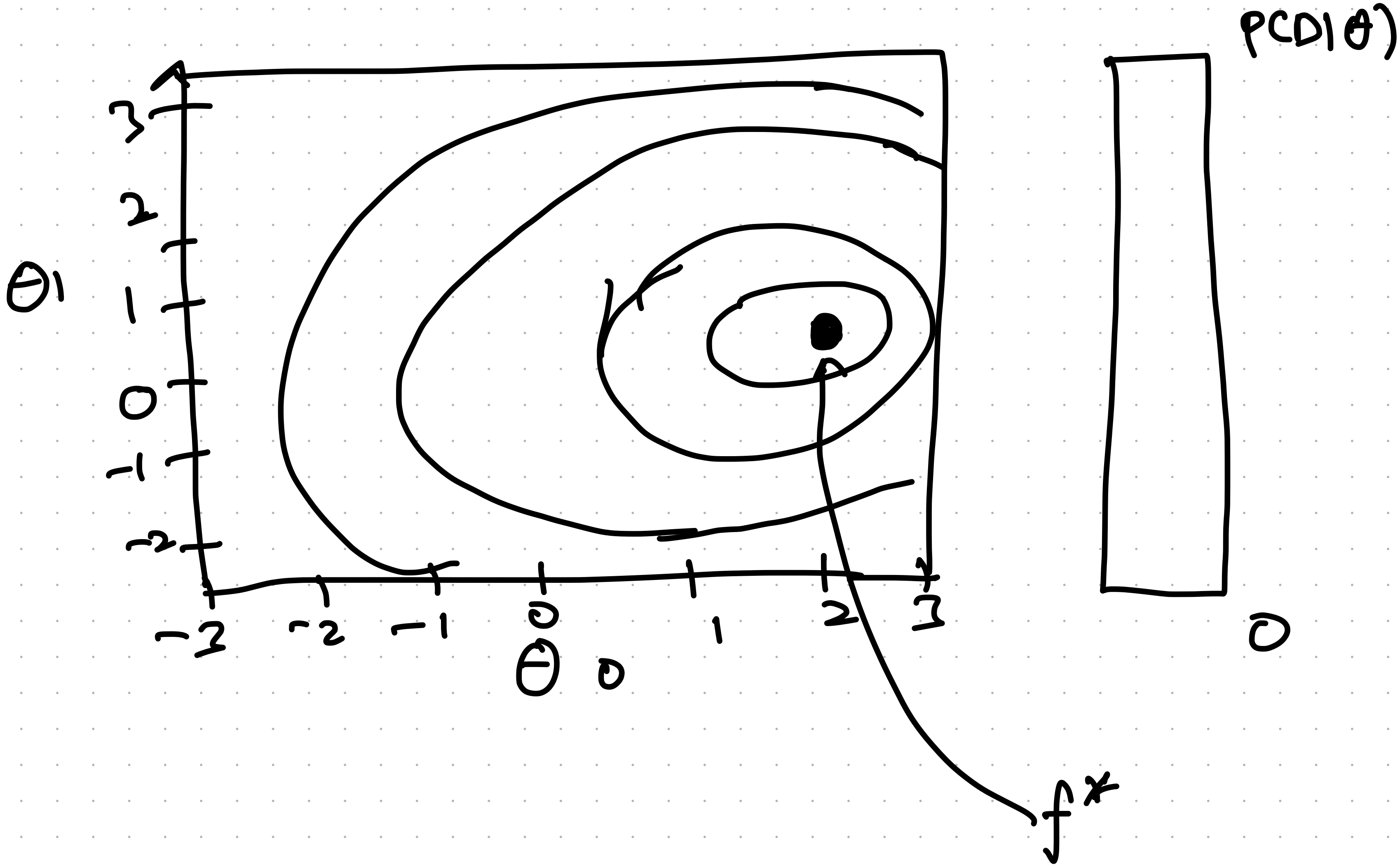
$$= P(D_1, \ldots D_N \mid \theta)$$

$$D = \{(x_1, y_1), (x_2, y_2) \ldots (x_N, y_N)\}$$




θ
$x_i$
$y_i$
σ
i=1,... N


θ
$x_i$
μ
σ
σ
N
$y_i$


P(D|θ)
θ1
3
2
1
0
-1
-2
-3
-2
-1
0
1
2
3
θ0
0
f*

## MLE for Lin. Reg

$$\hat{\theta}_{MLE} = \underset{\theta}{\arg\max}\; P(D \mid \theta)$$

$$= \underset{\theta}{\arg\max}\; \log P(D \mid \theta)$$

$$= \underset{\theta}{\arg\min}\; \underbrace{-}_{N.}\; \underbrace{\log}_{L.}\; \underbrace{P(D \mid \theta)}_{L.}$$

$$X \sim N(\mu, \sigma)$$

$$\mathrm{PDF}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

$$P(D_1|\theta) = ?$$

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

$$= P(y_1 | x_1; \theta) = N(y_1 | x_1^T\theta, \sigma^2)$$

$$P(D_1|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(y_1 - x_i^T\theta)^2}{\sigma^2}}$$

$$\log P(D_1|\theta) = \text{Constant} - \frac{1}{2}\frac{(y_1 - x_1^T\theta)^2}{\sigma^2}$$

$$P(D|\theta) = P(D_1, \ldots D_N|\theta)$$

$$= P(D_1|\theta) \ldots P(D_N|\theta)$$

$$\log P(D|\theta) = \log P(D_1|\theta) + \ldots \log P(D_N|\theta)$$

$$-\log P(D|\theta) = -\log P(D_1|\theta) \ldots -\log P(D_N|\theta)$$

$$NLL(\theta) = \sum_{i=1}^{N} \frac{(y_i - x_i^T\theta)^2}{2\sigma^2}$$

$$NLL(\theta) \propto \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \quad ; \hat{y}_i = x_i^T \theta$$

$$\underset{\theta}{\operatorname{argmax}} \; P(D|\theta)$$ = MAXIMIZING Likelihood

= MAX. LOG. LIKELIHOOD

= MIN. NLL.

= $J(\theta)$

= Squared Error LOSS