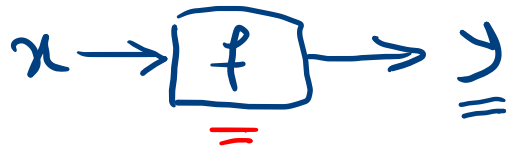


Regression

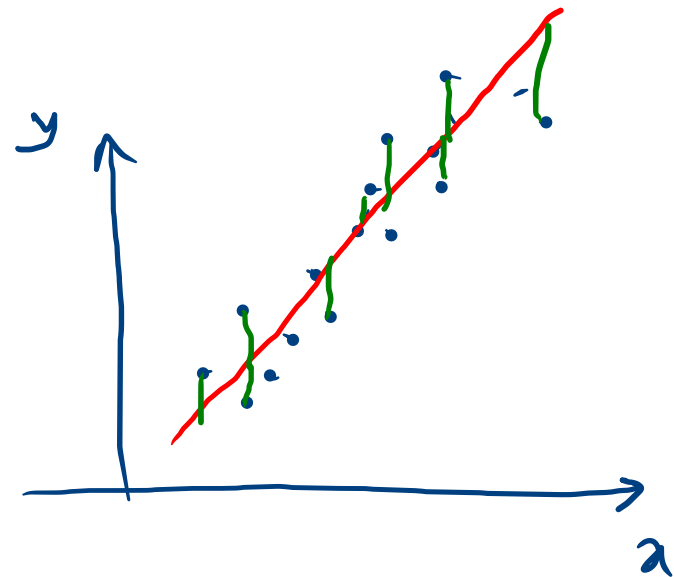
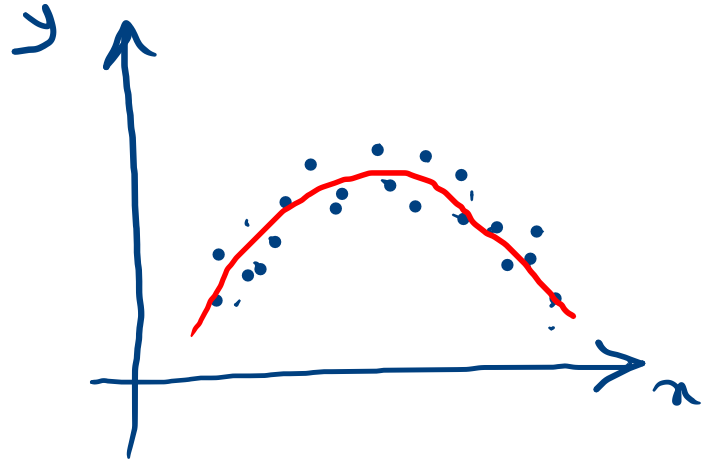


$$D = \{(x_1, y_1), \dots, (x_n, y_n)\}$$

Linear Regression

$$\min \sum_{i=1}^n (g(x_i) - y_i)^2$$

Sum of Squared Error (SSE)



$$y = w_1 x + w_0$$

$$y = w_1 x_1 + w_2 x_2 + \dots + w_d x_d + w_0 = \underline{\underline{w^T x}} + w_0$$

$$w = \begin{bmatrix} w_1 \\ \vdots \\ w_d \end{bmatrix}$$

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$$

$$w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}$$

$$x = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{bmatrix}$$

$$w^T x$$

$$g(x) = w^T x$$

w

$$\sum_{i=1}^n (g(x_i) - y_i)^2 = \sum_{i=1}^n (\omega^T x_i - y_i)^2$$

$$J(\omega) = \frac{1}{2} \sum_{i=1}^n (\omega^T x_i - y_i)^2$$

$$\nabla_{\omega} J = \sum_{i=1}^n x_i (\omega^T x_i - y_i) = 0$$

$$J(w) = \frac{1}{2} \sum_{i=1}^n (x_i^T w - y_i)^2 = \frac{1}{2} \left\| \begin{bmatrix} x_1^T w - y_1 \\ x_2^T w - y_2 \\ \vdots \\ x_n^T w - y_n \end{bmatrix} \right\|_2^2$$

Design Matrix

$$X = \begin{bmatrix} \text{---} x_1^T \text{---} \\ \text{---} x_2^T \text{---} \\ \vdots \\ \text{---} x_n^T \text{---} \end{bmatrix}_{n \times d}$$

$$\begin{matrix} X & w \\ \downarrow & \downarrow \\ n \times d & d \times 1 \end{matrix} = \begin{bmatrix} x_1^T w \\ x_2^T w \\ \vdots \\ x_n^T w \end{bmatrix}_{n \times 1}$$

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

$$J(w) = \frac{1}{2} \|Xw - y\|_2^2$$

$$J(w) = \min_w \frac{1}{2} \|Xw - y\|_2^2 = \min_w \frac{1}{2} (Xw - y)^T (Xw - y)$$

$$= \min_w \frac{1}{2} \left(\underbrace{w^T X^T X w}_A - 2 \underbrace{w^T X^T y} + y^T y \right)$$

$$\frac{\partial w^T A w}{\partial w} = 2Aw$$

$$\frac{\partial w^T x}{\partial w} = x$$

$$\nabla_w J = \underbrace{X^T X w}_A - X^T y = 0$$

$$\Rightarrow \boxed{w = \underbrace{(X^T X)^{-1}}_{\text{pseudo inverse}} X^T y}$$

pseudo inverse

$$\begin{matrix} \downarrow \\ X \omega = y \\ \begin{matrix} d \times d & & d \times 1 \\ & \searrow & \\ & d \times 1 & \end{matrix} \end{matrix}$$

$$\rightarrow \boxed{\omega = X^{-1} y}$$

$$\boxed{\begin{matrix} X \omega = y \\ \begin{matrix} n \times d & & n \times 1 \\ & \searrow & \\ & d \times 1 & \end{matrix} \end{matrix}}$$

$n > d$

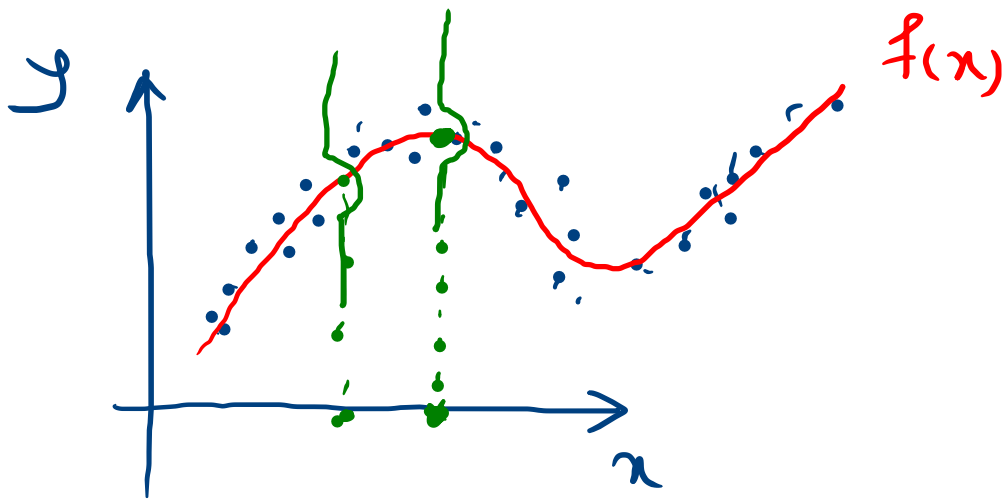
$$\|X\omega - y\|_2^2$$

$$X\omega = y$$

$$X^+ = \underbrace{(X^T X)^{-1}}_{d \times d} X^T_{d \times n}$$

$$\underbrace{X^T X}_{d \times d} \omega = X^T y_{d \times n}$$

$$\boxed{\omega = (X^T X)^{-1} X^T y}$$



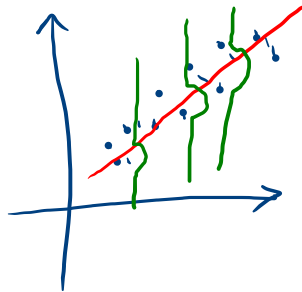
$$y = f(x) + \epsilon \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

$$P(y|x) = \mathcal{N}(y | \underbrace{f(x)}_y, \sigma^2)$$

$$E[y|x] = E[\underbrace{f(x) + \epsilon}_y | x] = f(x)$$

$$E[y^2|x] = \sigma^2$$

$$\begin{aligned} P(y|x) &= \mathcal{N}(y | f(x), \sigma^2) \\ &= \mathcal{N}(y | \underline{\underline{w^T x}}, \sigma^2) \end{aligned}$$



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

$$\hat{w}_{ML} = \arg \max_w \log P(D|w) = \arg \max_w \log P(y_1, \dots, y_n | x_1, \dots, x_n, w)$$

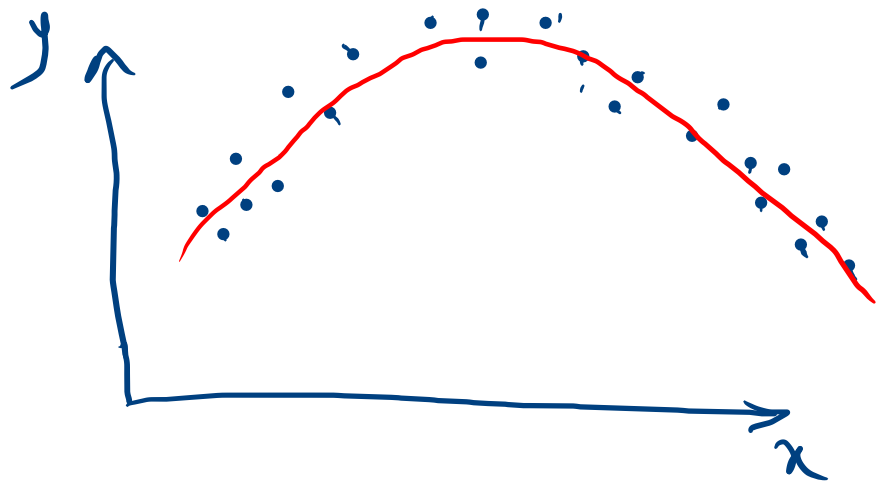
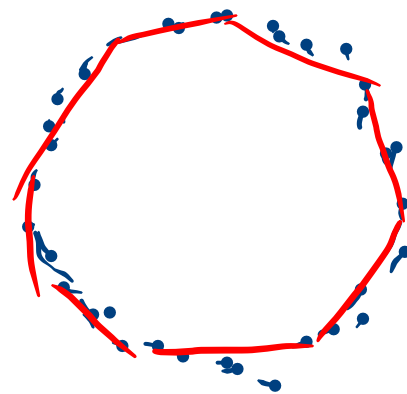
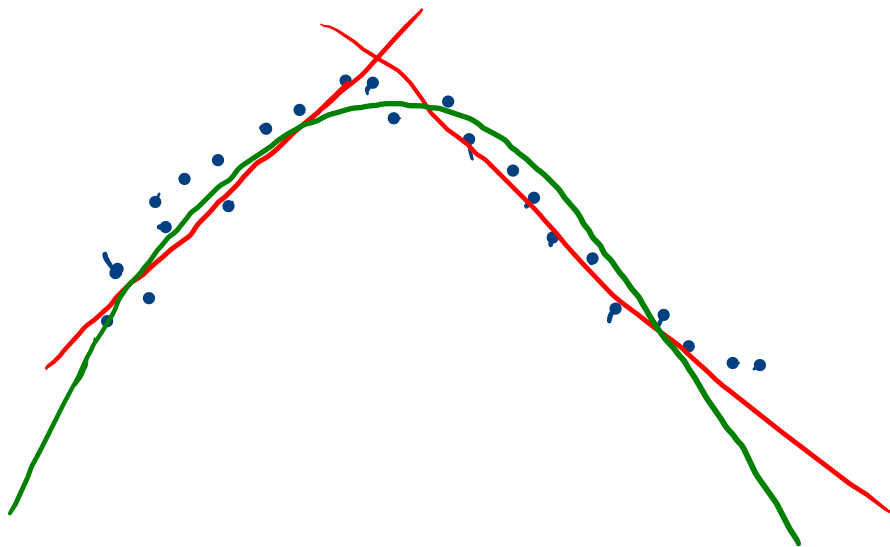
$$\underline{\underline{i.i.d.}} \arg \max_w \sum_{i=1}^n \log P(y_i | x_i, w)$$

$$= \arg \max_w \sum_{i=1}^n \log \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(y_i - w^T x_i)^2}{\sigma^2}\right)$$

$$= \arg \max_w - \frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^n (y_i - w^T x_i)^2}_{\text{SSE}} + \text{const.}$$

SSE

$$= \arg \min_w \frac{1}{2} \sum_{i=1}^n (y_i - w^T x_i)^2$$

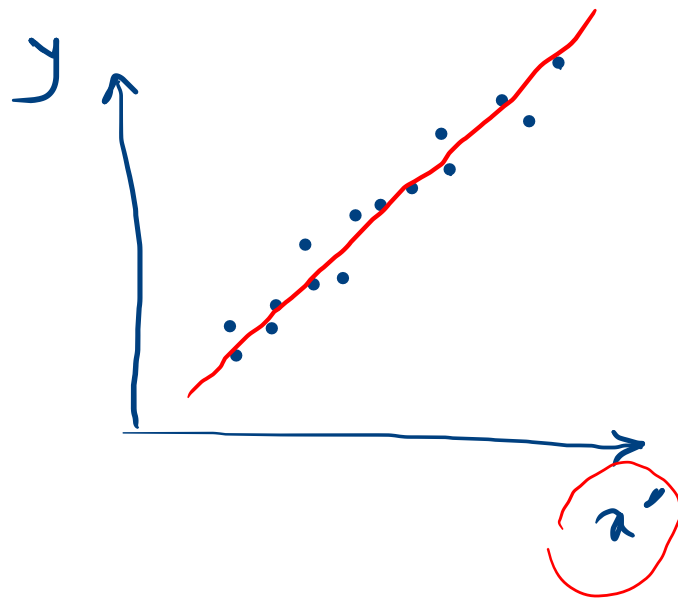


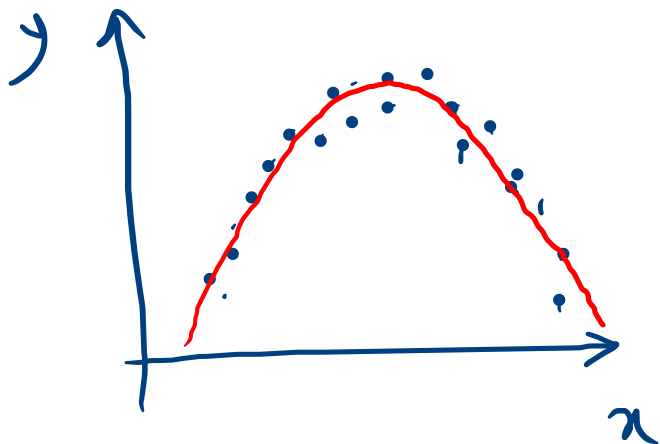
Φ

A blue arrow pointing from the left plot to the right plot, labeled with the symbol Φ .

Φ^{-1}

A red arrow pointing from the right plot back to the left plot, labeled with the symbol Φ^{-1} .





$$y = w_0 + w_1 x + w_2 x^2$$

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

$$w = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

$$y = w^T x'$$

$$\Phi: \mathbb{R} \rightarrow \mathbb{R}^3$$

$$\phi(x) = \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}$$

$$y = w_0 + w_1 x + \dots + w_{10} x^{10}$$