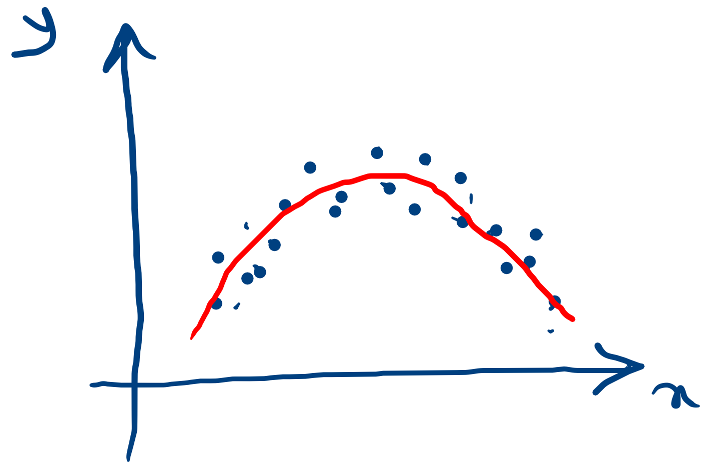
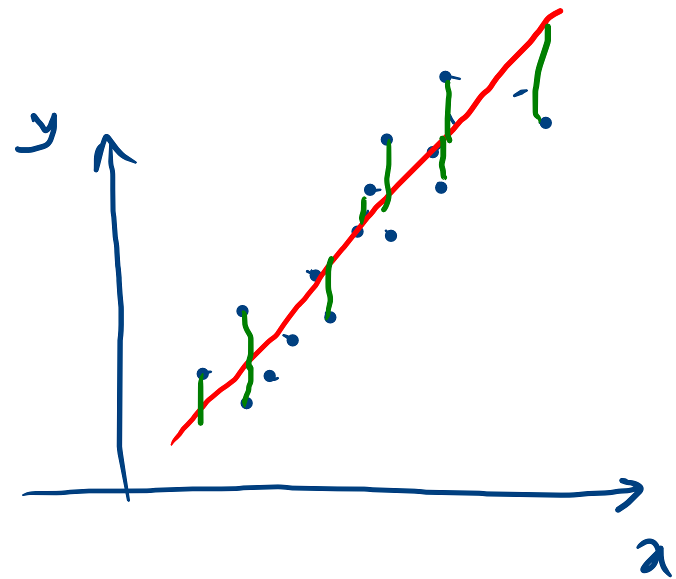
Regression

$$x \rightarrow \boxed{f} \rightarrow y$$



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

Linear Regression



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

Sum of Squared Error (SSE)

$$y = w_1 x + w_0$$

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

$$w = \begin{bmatrix} w_1 \\ \vdots \\ w_d \end{bmatrix} \qquad \underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$$

$$w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} \qquad x = \begin{bmatrix} 1 \\ x_1 \\ \vdots \\ x_d \end{bmatrix} \qquad w^T x$$

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

M

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

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

$$\nabla_w J = \sum_{i=1}^{n} x_i \left(w^T x_i - y_i \right) = 0$$

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

Design Matrix

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

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

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

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

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

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

$$\boxed{\frac{\partial\, w^T A w}{\partial w} = 2Aw \qquad \frac{\partial\, w^T x}{\partial w} = x}$$

$$\nabla_w J = \underbrace{X^T X}_{A} w - X^T y = 0 \implies \boxed{w = \underbrace{(X^T X)^{-1} X^T}_{\text{Pseudo inverse}} y}$$

$$\underset{d \times d}{X}\ \underset{d \times 1}{w} = \underset{d \times 1}{y} \longrightarrow \boxed{w = X^{-1} y}$$

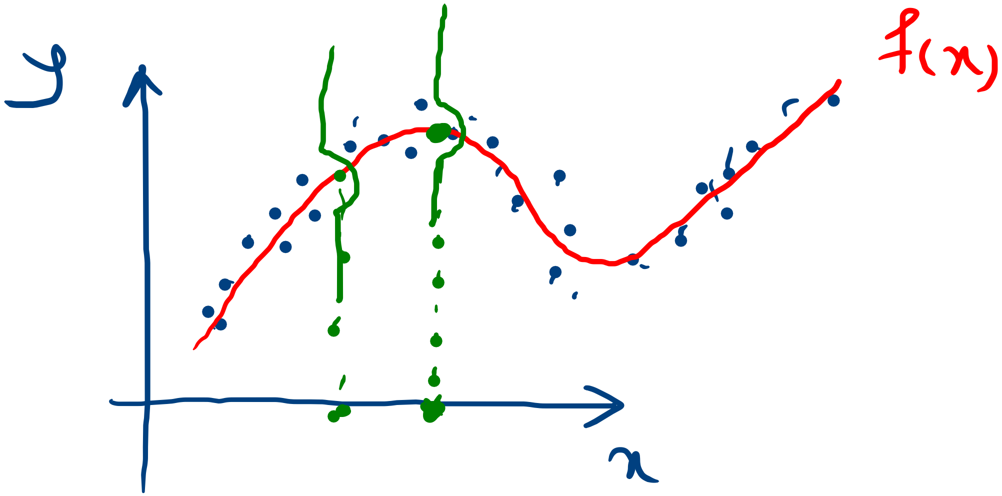
$$\boxed{\underset{n \times d}{X}\ \underset{d \times 1}{w} = \underset{n \times 1}{y} \qquad n > d} \longrightarrow \|Xw - y\|_2^2$$

$$Xw \doteq y$$

$$\underbrace{X^T X}\, w = X^T y$$

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

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



$$y = f(x) + \epsilon \qquad \epsilon \sim N(0, \sigma^2)$$

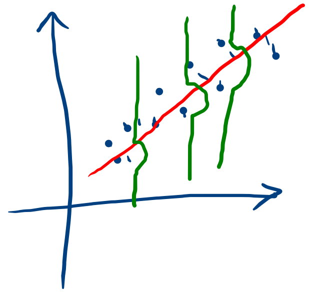
$$P(y|x) = N(y \mid f(x), \sigma^2)$$

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

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

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

$$= \boxed{\mathcal{N}(y \mid w^T x, \sigma^2)}$$



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

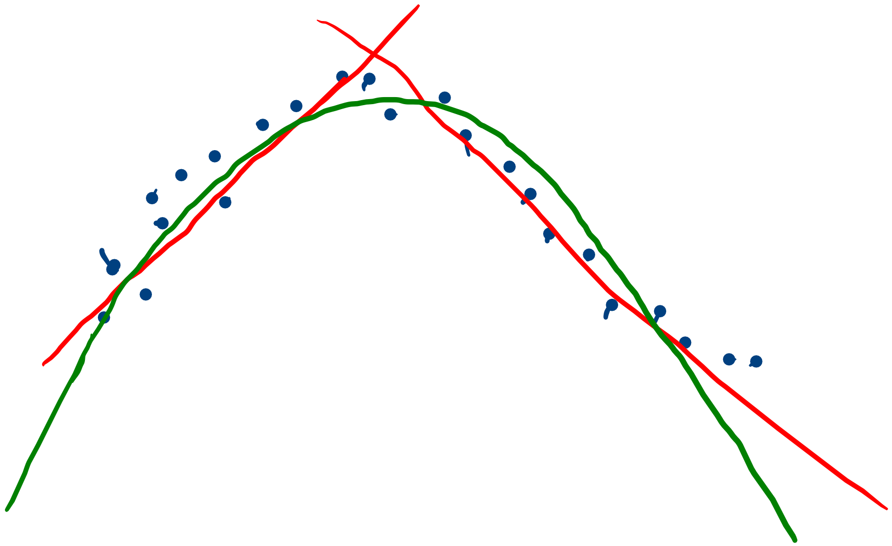
$$\hat{w}_{ML} = \arg\max_w \log P(D \mid w) = \arg\max_w \log P(y_1, \dots, y_n \mid x_1, \dots, x_n, w)$$

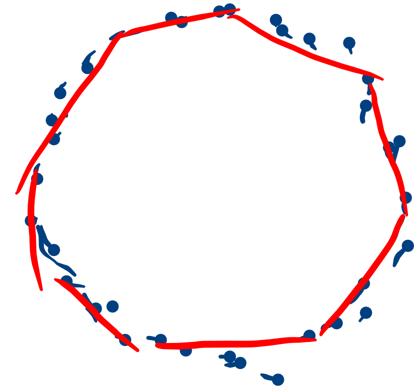
$$\overset{\text{i.i.d.}}{=} \arg\max_w \sum_{i=1}^{n} \log P(y_i \mid x_i, w)$$

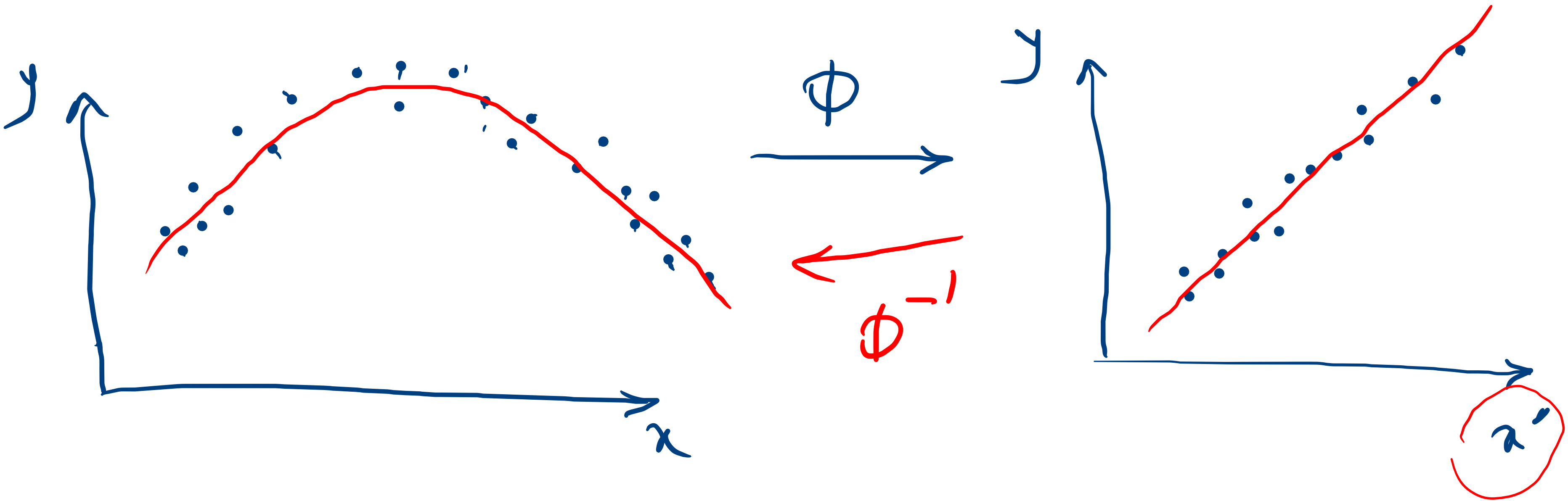
$$= \arg\max_w \sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2} \frac{(y_i - w^T x_i)^2}{\sigma^2}\right)$$

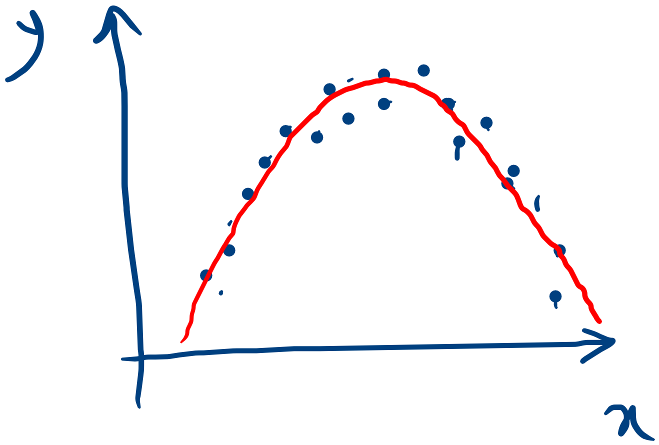
$$= \arg\max_w -\frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^{n} (y_i - w^T x_i)^2}_{SSE} + \text{const.}$$

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






y
x
Φ
Φ⁻¹
y
x'



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

$$x' = \begin{bmatrix}1\\x\\x^2\end{bmatrix} \qquad w = \begin{bmatrix}w_0\\w_1\\w_2\end{bmatrix}$$

$$\boxed{y = w^T x'}$$

$$\phi : \mathbb{R} \to \mathbb{R}^3$$

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

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