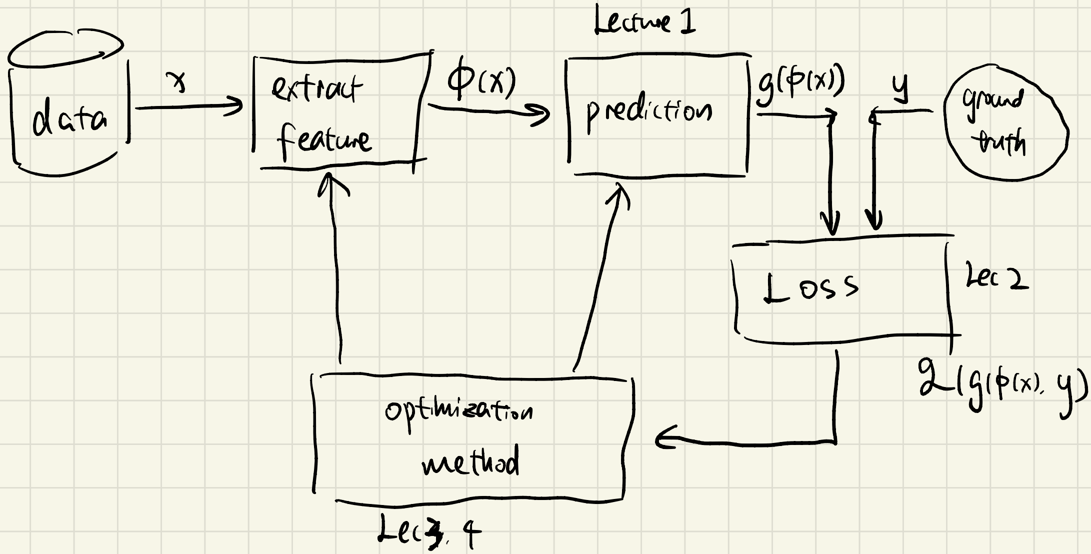
Linear Regression



Notation

$a, b, c \in \mathbb{R}$

$\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^d$

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_d \end{bmatrix}$$

$A, B, C \in \mathbb{R}^{N \times d}$. a_{ij}, $[A]_{ij}$

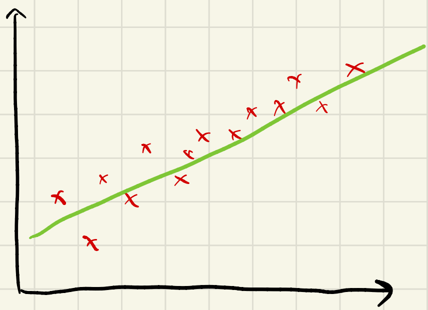
$$A = \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_N \\ | & | & & | \end{bmatrix}, \quad A = \begin{bmatrix} — & (\vec{x}^1)^T & — \\ — & (\vec{x}^2)^T & — \\ & \vdots & \\ — & (\vec{x}^d)^T & — \end{bmatrix}$$

I.

$$\vec{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}, \quad \vec{0}$$

$$\vec{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \; i\text{th}$$

Linear Regression Formulation



Regression:

- Given measurements: y^n, $n = 1, \ldots, N$
- Given inputs: $\vec{x}^n$
- Given model: $g_\theta(\vec{x}^n)$

$$y^n \approx g_\theta(\vec{x}^n)$$

Linear Regression:

$\vec{\theta}, \vec{x} \in \mathbb{R}^d$. $\quad \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}, \quad \vec{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_d \end{bmatrix}$

- $$g_\theta(\vec{x}) = \vec{x}^T \vec{\theta}, \quad = \sum_{j=1}^{d} x_j \theta_j$$

Solution

$y^n \approx g_\theta(\vec{x}^n)$

Loss function: $J(\vec{\theta}) = \sum_{n=1}^{N} (g_\theta(x^n) - y^n)^2$

$\sum_{n=1}^{N} |g_\theta(x^n) - y^n|$ 1-norm losses

Goal: $\hat{\vec{\theta}} = \arg\min_{\vec{\theta}} J(\vec{\theta})$

For $\vec{x}^{new}$. $y^{new} = g_{\hat{\theta}}(\vec{x}^{new}) = \vec{x}^{new\,T}\hat{\vec{\theta}}$

$$J(\vec{\theta}) = \sum_{n=1}^{N} (g_\theta(\vec{x}^n) - y^n)^2$$

$$\boxed{= \sum_{n=1}^{N} (\hat{\vec{\theta}}^T \vec{x}^n - y^n)^2} = \|A\vec{\theta} - \vec{y}\|^2$$

$\Downarrow$

$$A = \begin{bmatrix} \text{---} (\vec{x}^1)^T \text{---} \\ \text{---} (\vec{x}^2)^T \text{---} \\ \vdots \\ \text{---} (\vec{x}^N)^T \text{---} \end{bmatrix} \qquad \vec{y} = \begin{bmatrix} y^1 \\ \vdots \\ y^N \end{bmatrix}$$

$$\boxed{\begin{gathered} \nabla_{\vec{\theta}} J(\vec{\theta}) = 0 = 2(A)^T(A\vec{\theta} - y) \\ \vec{\theta} = (A^TA)^{-1}A^Ty \end{gathered}}$$

Theorem:

For a linear regression problem:

$$\hat{\vec{\theta}} = \underset{\vec{\theta}}{\operatorname{argmin}}\, J(\vec{\theta}) = \|A\vec{\theta} - \vec{y}\|^2$$

the minimizer $\hat{\theta} = (A^T A)^{-1} A^T y$.