① Linear Regression

Given the matrix $X_{m\times(d+1)}$ and the vector $\vec{y} \in \mathbb{R}^{d+1}$

To minimise the loss/objective function:

$$Loss(f, w) = \frac{1}{m}\sum_{i=1}^{m} (y_i - \bar{w}^T \bar{x}_i)^2$$

where $\bar{w} = \begin{bmatrix} b \\ \vec{w} \end{bmatrix} = \begin{bmatrix} b \\ w_1 \\ \cdots \\ w_d \end{bmatrix}$; $\bar{x}_i = \begin{bmatrix} 1 \\ \vec{x}_i \end{bmatrix} = \begin{bmatrix} 1 \\ x_{i,1} \\ \cdots \\ x_{i,d} \end{bmatrix}$

Take $J(\bar{w}) = \sum_{i=1}^{m} (y_i - \bar{w}^T \cdot \bar{x}_i)^2$

$$= (X\bar{w} - \vec{y})^T (X\bar{w} - \vec{y})$$

$$= \bar{w}^T X^T X \bar{w} - \boxed{\bar{w}^T X^T y} - \boxed{y^T X \bar{w}} + \vec{y}^T \vec{y} \quad \text{same!}$$

$$= \bar{w}^T X^T X \bar{w} - 2\bar{w}^T X^T y + \vec{y}^T \vec{y}$$

$$\nabla J_{\bar{w}}(\bar{w}) = 2X^T X \bar{w} - 2X^T \vec{y} = 0$$

$\Rightarrow \quad \bar{w}^* = (X^T X)^{-1} X^T \vec{y}$ if X has full rank.

② ~~Class~~ Linear classification

I. Binary classification: $y_i \in \{-1, 1\}$

For any $X\bar{w} > 0 \rightarrow y_i = 1$; otherwise, $y_i = -1$

$$\bar{w}^* = (X^T X)^{-1} X^T \vec{y}.$$

II. multi-class classification

let $\vec{y}_i$ be $1 \times (d+1)$ size.

$\bar{w}^* = (X^T X)^{-1} X^T Y$, $\quad \bar{w}^* \begin{bmatrix} 1 \\ X_{new} \end{bmatrix}^T$, which column is largest, it belongs to what the column symbolizes.