Linear Regression

Pham Van Hong – 11201638

October 2022

1. Problem 1:

To transfer $t = y(x, w) + \varepsilon$ to $w = (X^T X)^{-1} X^T t$, we minimize the loss function:

$$L = \frac{1}{N} \times \sum_{i=1}^{N} \left(t_n - y(x_n, w)\right)^2$$

Note that: $y(x_n, w) = w_1 x_n + w_0$

We have some vectors:

The data set of observations: $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$

The weight: $w = \begin{bmatrix} w_1 \\ w_0 \end{bmatrix}$

The corresponding target value: $\vec{t} = \begin{bmatrix} t_1 \\ t_2 \\ \dots \\ t_n \end{bmatrix}$

Prediction vector: $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} w_1 x_1 + w_0 \\ w_2 x_2 + w_0 \\ \dots \\ w_n x_n + w_0 \end{bmatrix} = x \times w$

We calculated:

$$\vec{t} - \vec{y} = \begin{bmatrix} t_1 - y_1 \\ t_2 - y_2 \\ \dots \\ t_n - y_n \end{bmatrix} = (\|t - y\|_2)^2$$

$$= (t_1 - y_1)^2 + (t_2 - y_2)^2 + \dots + (t_n - y_n)^2$$

$$= \sum_{i=1}^{N} (t_i - y_i)^2$$

$$= L$$

$$= (\|t - y\|_2)^2$$

$$= (\|t - xw\|_2)^2$$

Then, we derivates:

$$\frac{\partial L}{\partial w} = 2.X^T.(t - X_w)$$

$$= 0$$

$$<=> X^T t = X^T X w$$

$$<=> w = (X^T X)^{-1}.X^T.t$$

2. Problem 4:

When X is full rank, X is linear independent, so:

$$\vec{v}^T.X^T.X.\vec{v} \quad = \quad \vec{v}^T.\vec{0} = 0$$

$$\gg (X.\vec{v})^T.X.\vec{v} = 0$$

$$\gg (X.\vec{v}).(X.\vec{v}) = 0$$

$$\gg X.\vec{v} = \vec{0}$$

From that, if:

$\vec{v} \in N(X^T X)$

$\gg \vec{v} \in N(X)$

$\gg N(X^T X) = N(X) = \vec{0}$

$\gg X^T X \ if \ linear \ independent \ and \ X^T X \ is \ a \ square \ matrix$

$\gg X^T X \ is \ invertible$