MACHINE LEARNING

Nghiem Thi Ngoc Minh
MSV 11202545

HW week2

1 Problem 1

Now we will minimize:

$$P = \frac{\beta}{2}\sum_{n=1}^{N}\{y(x_n, w) - t_n\}^2$$

Suppose that: $y(x_n, n) = w_1 x_n + w_0$ for linear regression problem

$$x = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}; \qquad t = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}; \qquad w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

Then,

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} w_1 x_1 + w_0 \\ w_2 x_2 + w_0 \\ \vdots \\ w_n x_n + w_0 \end{bmatrix} = x.w$$

$$t - y = \begin{bmatrix} t_1 - y_1 \\ t_2 - y_2 \\ \vdots \\ t_n - y_n \end{bmatrix}$$

$$\begin{aligned} \longrightarrow \|t - y\|_2^2 &= (t_1 - y_1)^2 + \cdots + (t_n - y_n)^2 \\ &= \sum (t_i - y_i)^2 = P \end{aligned}$$

$$\longrightarrow P = \|t - y\| = \|t - xw\| = (xw - t)^T (xw - t)$$

Take the derivate of P:

$$\frac{\partial(P)}{\partial(w)} = 2x^T(t - xw) = 0$$

$$\begin{aligned} &\leftrightarrow x^t = x^T x w \\ &\leftrightarrow w = (x^T x)^{-1} x^T.t \end{aligned}$$

2 Problem 4

Proof $X^T X$ invertible khi X full rank. If X is full rank, X is linear independent.

$$\begin{aligned}
&\Rightarrow \vec{v}^T X^T X \vec{v} = \vec{v}^T \overrightarrow{0} \\
&\Rightarrow (X\vec{v})^T X \vec{v} = 0 \\
&\Rightarrow (X\vec{v}) \cdot (X\vec{v}) = 0 \\
&\Rightarrow X\vec{v} = \overrightarrow{0}
\end{aligned}$$

We have: if $\vec{v} \in N\left(X^T X\right) \Rightarrow \vec{v} \in N(X)$

$\Rightarrow \vec{v}$can only be $\overrightarrow{0} \Rightarrow N\left(X^T X\right) = N(X) = \{\overrightarrow{0}\}$

$\Rightarrow X^T X$ is linearly independent; and $X^T X$ is a square matrix $\Rightarrow X^T X$ is invertible