Machine Learning

Tran Kha Uyen - 11207413

October, 2022

1 Show that $t = y(w, x) + noise \to w = (X^T X)^{-1} X^T t$

We have

$$t = y(w, x) + noise \Rightarrow t = w_0 + w_1 \times x$$

And

$$L = \frac{1}{N} \sum_{i=1}^{N} (t - y)^2$$

In matrix form:

$$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & ... \\ 1 & x_n \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ ... \\ y_n \end{bmatrix}$$

$$\mathbf{t} = \begin{bmatrix} t_1 \\ ... \\ t_n \end{bmatrix}$$

$$\Rightarrow \mathbf{Xw} = \begin{bmatrix} w_0 + w_1 x_1 \\ ... \\ w_0 + w_1 x_n \end{bmatrix} = \begin{bmatrix} t_1 \\ ... \\ t_n \end{bmatrix} = \mathbf{t}$$

The loss function can be written as

$$L = ||t - y||_2^2 = ||\mathbf{Xw} - \mathbf{y}||_2^2 = (\mathbf{Xw} - \mathbf{y})^T (\mathbf{Xw} - \mathbf{y})$$

To minimize the lost function, take derivative and set to zero.

$$\frac{\partial L}{\partial \mathbf{w}} = \frac{\partial (\mathbf{Xw} - \mathbf{y})^T (\mathbf{Xw} - \mathbf{y})}{\partial \mathbf{w}}$$

$$= \frac{\partial (\mathbf{Xw} - \mathbf{y})^T (\mathbf{Xw} - \mathbf{y})}{\partial (\mathbf{Xw} - \mathbf{y})} \times \frac{\partial (\mathbf{Xw} - \mathbf{y})}{\partial \mathbf{w}}$$

$$= 2(\mathbf{X}\mathbf{w} - \mathbf{y})\mathbf{X}^T = 2\mathbf{X}^T(\mathbf{X}\mathbf{w} - \mathbf{y})$$

$$\begin{aligned}
&\rightarrow 2\mathbf{X}^T(\mathbf{X}\mathbf{w} - \mathbf{y}) = 0 \\
&\Leftrightarrow \mathbf{X}^T\mathbf{X}\mathbf{w} - \mathbf{X}^T\mathbf{y} = 0 \\
&\Leftrightarrow \mathbf{X}^T\mathbf{X}\mathbf{w} = \mathbf{X}^T\mathbf{y} \\
&\Leftrightarrow \mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}
\end{aligned}$$

2 Show that if $\mathbf{X}$ is full rank, $\mathbf{X}^T\mathbf{X}$ is invertible

$\mathbf{X}^T\mathbf{X}$ is positive semidefinite (PSD) because

$$\mathbf{X}^T\mathbf{X} = (\mathbf{X}^T\mathbf{X})^{T(1)}$$

and with $\forall a \in R^n$

$$a^T(\mathbf{X}^T\mathbf{X})a = (\mathbf{X}a)^T(\mathbf{X}a) = ||(\mathbf{X}a)||^2 \geqslant 0^{(2)}$$

$||(\mathbf{X}a)||^2 = 0 \Leftrightarrow (\mathbf{X}a) = 0 \Leftrightarrow \mathbf{X}$ are linearly dependent. Therefore, if $\mathbf{X}$ has full rank, then $\mathbf{X}^T\mathbf{X}$ is positive definite.

Every positive definite matrix is invertible. So if $\mathbf{X}$ has full column rank then $\mathbf{X}^T\mathbf{X}$ is invertible.