Homework 3: Linear Regression

Trần Minh Khoa

September 28, 2021

1.
t = y(x,w) + N(0, β^{-1})

$\Rightarrow$ t $\sim$ N(y(x,w), β^{-1})

$\Rightarrow$ p(t) = N(t|y(x,w),β^{-1})

If the data are assumed to be drawn independently from the distribution then the likelihood function is:

$$p(t|x, w, \beta) = \prod_{n=1}^{N} N(t_n|y(x_n, w), \beta^{-1})$$

It is convenient to maximize the logarithm of the likelihood function:

$$log\ p(t|x, w, \beta) = \sum_{n=1}^{N} log(N(t_n|y(x_n, w), \beta^{-1}))$$

$$= \sum_{n=1}^{N} log\Big[\frac{1}{\sqrt{2\pi\beta^{-1}}}\ exp\Big[-\Big(t_n - y(x_n, w)\Big)^2 \frac{\beta}{2}\Big]\Big]$$

$$= \sum_{n=1}^{N} \Big[log\ \frac{1}{\sqrt{2\pi\beta^{-1}}} + log\ exp\Big[-\Big(t_n - y(x_n, w)\Big)^2 \frac{\beta}{2}\Big]\Big]$$

$$= \sum_{n=1}^{N} \Big[-log\sqrt{2\pi\beta^{-1}}\ - \Big(t_n - y(x_n, w)\Big)^2 \frac{\beta}{2}\Big]$$

$$= \sum_{n=1}^{N} \left[-\frac{1}{2} \ log(2\pi\beta^{-1}) - \frac{\beta}{2}\Big(t_n - y(x_n, w)\Big)^2 \right]$$

$$= \sum_{n=1}^{N} \left[-\frac{1}{2}\Big(log \ 2\pi - log \ \beta\Big) - \frac{\beta}{2}\Big(t_n - y(x_n, w)\Big)^2 \right]$$

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

$$max \ log \ p(t|x, w, \beta) = max\left(-\sum_{n=1}^{N} \Big(t_n - y(x_n, w)\Big)^2 \right)$$

$$= min \ \sum_{n=1}^{N} \Big(t_n - y(x_n, w)\Big)^2 \quad (Loss function)$$

We suppose:

X = $\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ ... & ... \\ 1 & x_n \end{bmatrix}$, w = $\begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$, t = $\begin{bmatrix} t_1 \\ t_2 \\ ... \\ t_n \end{bmatrix}$

y = $\begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix}$ = $\begin{bmatrix} w_1x_1 + w_0 \\ w_1x_2 + w_0 \\ ... \\ w_1x_n + w_0 \end{bmatrix}$ = Xw

$\Rightarrow$ Loss function = $\sum_{n=1}^{N}(t_n - y_n)^2 = \big|\big|t - y\big|\big|_2^2 = \big|\big|t - Xw\big|\big|_2^2$

$$\Rightarrow L = (t - Xw)^T(t - Xw)$$

$$= (t^T - w^TX^T)(t - Xw)$$

$$= t^Tt - t^TXw - w^TX^Tt + w^TX^TXw$$

$$\Rightarrow \frac{\partial L}{\partial w} = 0 - X^T t - X^T t + 2X^T X w$$
$$= -2X^T t + 2X^T X w = 0$$
$$\Leftrightarrow X^T X w = X^T t$$
$$\Leftrightarrow w = (X^T X)^{-1} X^T t$$

2. Prove $X^T X$ invertible when X full rank (linearly independent)

X is linearly independent

$\Rightarrow$ The null space of X only has the trivial solution

$\Rightarrow$ N(X) = $\{\vec{0}\}$

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

$\Rightarrow X^T X \vec{v} = \vec{0}$

$\Rightarrow \vec{v}^T X^T X \vec{v} = \vec{v}^T\ \vec{0} = 0$

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

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

$\Rightarrow X\vec{v} = \vec{0}$

So we have: if $\vec{v} \in N(X^T X) \Rightarrow \vec{v} \in N(X)$

$\Rightarrow \vec{v}$ can only be $\vec{0}$

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

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

$\Rightarrow X^T X$ is invertible