HW3 - Linear Regression

Phương Quỳnh Chi

October 1, 2022

Exercise 1: $t = y(x,w) + noise \Rightarrow w = (X^TX)^{-1}X^Tt$
We have: $t = y(x,w) + noise = N(y(x,w), \sigma^2)$
$\Rightarrow \quad p(t) = N(t|y(x,w), \beta^{-1}) \qquad \text{with } \beta = \dfrac{1}{\sigma^2}$
Making assumption that these data points are drawn independently from the distribution then we obtain the likelihood function:

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

It is convenient to logarithm of the likelihood function

$$\begin{aligned} log\ p(\mathbf{t}|\mathbf{X}, w, \beta) &= \sum_{n=1}^{N} log(N|t_n|y(x_n,w), \beta^{-1}) \\ &= \frac{-\beta}{2}\sum_{n=1}^{N}\Big(y(x_n,w) - t_n\Big)^2 + \frac{N}{2}log\ \beta - \frac{N}{2}log(2\pi) \end{aligned}$$

$$\begin{aligned} max_w log\ p(\mathbf{t}|\mathbf{X}, w, \beta) &= -max_w \frac{\beta}{2}\sum_{n=1}^{N}\Big(y(x_n,w) - t_n\Big)^2 \\ &= min\frac{1}{2}\sum_{n=1}^{N}\Big(y(x_n,w) - t_n\Big)^2 = L \end{aligned}$$

We suppose:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ ... & ... \\ 1 & x_n \end{bmatrix}, \quad w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, \quad \mathbf{t} = \begin{bmatrix} t_1 \\ t_2 \\ ... \\ t_n \end{bmatrix} \rightarrow Y = \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix} = \begin{bmatrix} w_1x_1 & w_0 \\ w_2x_2 & w_0 \\ ... & ... \\ w_nx_n & w_0 \end{bmatrix} = \mathbf{X}w$$

$$\begin{aligned} \Rightarrow L &= \sum_{n=1}^{N}(t_n - y_n)^2 = ||t-y||_2^2 = ||t - Xw||_2^2 \\ &= (t - Xw)^T(t - Xw) \\ &= (t^T - Xw)^T(t - Xw) \\ &= t^Tt - t^TXw - w^TX^Tt + w^TX^TXw \\ \Rightarrow \frac{\partial L}{\partial w} &= 0 - X^Tt - X^Tt + X^TXw + (w^TX^TX)^T \\ &= -2X^Tt + 2X^TXw = 0 \end{aligned}$$

$\Leftrightarrow \quad \mathrm{X}^T t = X^T X w$
$\Leftrightarrow \quad \mathrm{w} = (\mathrm{X}^T X)^{-1} X^T t$

Exercise 4: Prove that $X^T X$ is invertible when X is full rank
For any nonezero matrix X, $X^T X$ has the same rank as X. If X is m x n matrix, a necessary and sufficient condition for $X^T X$ to be invertible is that rank(X) = n. Then, it need to have full column rank (column linear independent).
Suppose

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

$Xv = 0$ if and only if v is in null space of $X^T X$.
But $Xv = 0$ and $v \neq 0$ if and only if X has linearly independent columns. So $X^T X$ í invertible if and only if X has full column rank.