Machine Learning Homework Week 4

Trinh Thi Huong Lien

October 2022

1 Question 1: Linear regression model

We have:
+ Set an observation

$$x = (x_1, x_2, .., x_N)^T \tag{1}$$

+ Total observations N
+ Target values

$$t = (t_1, t_2, .., t_N)^T \tag{2}$$

Suppose that the observations are drawn independently from a Gaussian distribution.

$$t = y(x, w) + N(0, \beta^{-1})t = N(y(x, w), \beta^{-1}) \tag{3}$$

with $\beta = \frac{1}{\sigma^2}$

p(t |x, w, β) = N(t |y(x, w), β^{-1})
We now use the training data x, t to determine the values of the unknown parameters w by maximum likelihood function. If the data are assumed to be drawn independently from the distribution, then the likelihood function:

$$p(t|x, w, \beta) = \prod_{n=1}^{N} N(t|y(x, w), \beta^{-1}) \tag{4}$$

We have to maximise the logarithm of the likelihood function:

$$logp(t|x, w, \beta) = \prod_{n=1}^{N} log(N(t|y(x, w), \beta^{-1})) \tag{5}$$

$$logp(t|x, w, \beta) = \sum_{n=1}^{N} log\left(\frac{1}{\sqrt{2\pi\beta^{-1}}} e^{-\frac{(t_n - y(x_n, w))^2}{2}}\right) \tag{6}$$

$$logp(t|x,w,\beta) = \sum_{n=1}^{N}\left[\frac{1}{2}log(2\pi\beta^{-1} - t_n - y(x_n,w)^2 - \frac{\beta}{2}\right] \tag{7}$$

$$logp(t|x,w,\beta) = -\sum_{n=1}^{N}(t_n - y(x_n,w)^2) \tag{8}$$

We eliminate $\frac{1}{2}log(2\pi\beta^{-1})$ and $\frac{\beta}{2}$ because they are not elements that we want to optimize. We have to minimize $(t_n - y(x_n,w)^2)$
Set

$$L = \frac{1}{2N}\sum_{n=1}^{N}(t_n - y(x_n,w)^2) \tag{9}$$

with

$$x = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ .. & .. \\ 1 & x_n \end{bmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ .. \\ y_n \end{bmatrix} = \begin{bmatrix} w_1x_1 + w_0 \\ w_1x_2 + w_0 \\ w_1x_3 + w_0 \\ .. \\ w_1x_n + w_0 \end{bmatrix} = xw \tag{10}$$

we have

$$\frac{\partial L}{\partial w} = \begin{bmatrix} \frac{\partial L}{\partial w_0} \\ \frac{\partial L}{\partial w_1} \end{bmatrix} = \begin{bmatrix} t - xw \\ x(t - xw) \end{bmatrix} = x^T(t - xw) = 0 \tag{11}$$

$\iff$ x$^Tt = x^Txw$
$\iff$ w = (x$^Tx)^{-1}x^Tt$

2 Question 2: Proof X^TX is invertible when X full rank

Supppose $X^Tv = 0$, then $XX^Tv = 0$ too.
On the other hand, suppose $XX^Tv = 0$, then $v^TXX^Tv = 0$ too. Therefore, $(X^Tv)^T(X^Tv) = 0$. It also means $X^Tv = 0$
Hence, we also know that $X^Tv = 0 \iff$ v is in the nullspace of X^TX
But $X^Tv = 0$ and v $\neq 0 \iff$ X has linearly dependent rows.
Therefore, X^TX is invertible when X has full row rank.