Course Name: Math for ML
Name: Nguyen Thu Thao

Assignment 3

September 29, 2021

Question 1. *Find the output of linear regression:* $w = \left(x^T x\right)^{-1} x^T t$

My Answer 1.

We have:
- Set a observation $x = (x_1, x_2, \dots, x_N)^T$
- Total observation N
- Target values $t = (t_!, t_2, \dots t_N)^T$

Suppose that the observations are drawn independently from a Gaussiandistribution.

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

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

$$p(t \mid x, w, \beta) = N\left(t \mid y(x,w), \beta^{-1}\right)$$

We now use the training data x, t to determine the values of the unknown parameters w and by maximum likelihood. If the data are assumed to be drawn independently from the distribution then the likelihood function:

$$p(t \mid x, w, \beta) = \prod_{n=1}^{N} N\left(t \mid y(x,w), \beta^{-1}\right)$$

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

$$\begin{aligned}\text{logp}(t \mid x, w, \beta) &= \textstyle\sum_{n=1}^{N} \log\left(N\left(t \mid y(x,w), \beta^{-1}\right)\right) \\ &= \textstyle\sum_{n=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\beta^{-1}}} e^{-\frac{(t_n - y(x_n,w))^2 B}{2}}\right)\end{aligned}$$

we have:

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

$$\leftrightarrow x^T t = x^T x w$$

$$\leftrightarrow w = \left(x^T x\right)^{-1} x^T t$$

Question 2. *Proof that $X^T X$ is invertible when X is full rank.*

My Answer 2.

We note that $N(X) \subseteq N\left(X^T X\right)$ (N denoting the null space).
Conversely, if $v \in N\left(X^T X\right)$, then $X^T X v = 0$, so also $v^T X^T X v = 0$, which is to say that $(Xv)^T(Xv) = 0$. This implies $Xv = 0$, so $v \in N(X)$.
But $X^\top v = 0$ and $v \neq 0$ if and only if X has linearly dependent rows. Thus, $XX^\top$ has nullspace $\{0\}$ (i.e. $XX^\top$ is invertible) if and only if X has linearly independent rows.