Homework 3 Machine Learning

1. **Biến đổi lại linear regression trên lớp ra latex, từ t = y(x,w) + noise -> w = (X^TX)-1X^Tt**

$$t = y(x,w) + \varepsilon \quad \Rightarrow w = (X^T X) - 1X^T t$$

$N(\mu, \sigma^2) \quad => \quad \varepsilon \sim N(0, \sigma^2)$

$\Rightarrow t = y(x,w) + \varepsilon \sim N(y(x,w), \sigma^2)$

$\Rightarrow P(t) = N(t|y(x,w), \sigma^2)$

$Supppose: \ t_n = y(x_n, w) + \varepsilon$

$=> P(t_n) = N(t_n|y(x_n,w), \varepsilon^2$

$We\ use\ Maximum\ likelihood\ funcion:$

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

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

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

$Maximum\ likelihood:$

$$Max \log\big(P(t|x,w,\beta)\big) = -Max\frac{\beta}{2}\sum_{n=1}^{N}(y(x_n,w) - t_n)^2$$

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

$$we\ minimine\ \ P = \frac{1}{2}\sum_{n=1}^{N}(y(x_n,w) - t_n)^2$$

$$suppose: X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots & \cdots \\ 1 & x_n \end{pmatrix} \quad ; \qquad t = \begin{pmatrix} t_1 \\ t_2 \\ \cdots \\ t_n \end{pmatrix} \quad ; \qquad w = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}$$

$P\ is\ called\ Mean\ Squared\ Error\ Loss\ (MSE):$

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

We have:

$y(x_n, w) = w_1 x_v + w_0$

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} w_1 x_1 & w_0 \\ w_2 x_2 & w_0 \\ \dots & \dots \\ w_n x_n & w_0 \end{pmatrix} = XW$$

$$t - y = \begin{pmatrix} t_1 - y_1 \\ t_2 - y_2 \\ \dots \\ t_n - y_n \end{pmatrix} \quad => L = \|t - y\|_i^2 = \|t - Xw\|_i^2 = (t - Xw)^T (t - Xw)$$

$$\frac{\partial L}{\partial w} = 2X^T(t - Xw) = 0$$

$\Leftrightarrow$ $X^T t = X^T x w$

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

2. Chứng minh $\mathbf{XX^T}$ invertible khi X full rank.

The condition that X is a full rank matrix is not enough. It needs to have full row rank, i.e. it needs to have linearly independent rows.

For example, the matrix M = $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ has full rank, but MM^T is not invertible. The reason is that M does not have full row rank, but full column rank.

Assuming X has full row rank, then yes, XX^T will be invertible. The proof is the following.

Suppose $X^T v$ =0 . Then, of course, $XX^T v$ =0 too.

Conversely, suppose $XX^T v$ =0 . Then $v^T XX^T v$ =0 , so that $(X^T v)^T$ ($X^T v$)=0 . This implies $X^T v$ =0 .

Hence, we have proved that $X^T v$ =0 if and only if v is in the nullspace of XX^T. But $X^T v$ =0 and v≠0 if and only if X has linearly dependent rows. Thus, XX^T has nullspace {0} (i.e. XX^T is invertible) if and only if X has linearly independent rows.