Homework 3

Tong Thi Van Anh - 11200365

September 2022

1 Exercise 1

We have: observation $\mathbf{x}$, the target values $\mathbf{t}$ and the model $y(\mathbf{x}, \mathbf{w})$, noise ϵ:

$$\mathbf{t} = y(\mathbf{x}, \mathbf{w}) + \epsilon$$

$$\Rightarrow \mathbf{t} \sim \mathcal{N}(y(\mathbf{x}, \mathbf{w}), \sigma^2)$$

$$\Rightarrow p(\mathbf{t}) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}), \sigma^2)$$

Then:

$$t_n = y(x_n, \mathbf{w}) + \epsilon$$

and

$$p(t_n) = \mathcal{N}(t_n|y(x_n, \mathbf{w}), \sigma^2)$$

The likilihood function:

$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma) = \prod_{n=1}^{N} \mathcal{N}(t_n|y(x_n, \mathbf{w}), \sigma^2)$$

It's convenient to maximize the log of the likelihood function:

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

Maximizing above $\Leftrightarrow$ minimizing $\mathcal{L} = \sum_{n=1}^{N}(y(x_n, \mathbf{w}) - t_n)^2$.

Recall: $\mathbf{x} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$, $\mathbf{t} = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$,

$$\text{and } y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} w_0 + w_1 x_1 \\ w_0 + w_1 x_2 \\ \vdots \\ w_0 + w_1 x_n \end{bmatrix} = \mathbf{xw}$$

$$\Rightarrow \mathbf{y} - \mathbf{t} = \begin{bmatrix} y_1 - t_1 \\ y_2 - t_2 \\ \vdots \\ y_n - t_n \end{bmatrix} = \mathbf{xw} - \mathbf{t}$$

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

$$\Rightarrow \mathcal{L} = ||\mathbf{y} - \mathbf{t}||_2^2 = (\mathbf{xw} - \mathbf{t})^T (\mathbf{xw} - \mathbf{t})$$

We have: $\frac{\partial x^T x}{\partial x} = 2x$.

$$\Rightarrow \frac{\partial u^T u}{\partial x} = \frac{\partial u^T u}{\partial u} \frac{\partial u}{\partial x} = (2u)^T u'$$

Therefore,

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = 2\mathbf{x}^T (\mathbf{xw} - \mathbf{t}) = 0$$

$$\Leftrightarrow \mathbf{x}^T \mathbf{xw} = \mathbf{x}^T \mathbf{t}$$

$\mathbf{x}^T \mathbf{x}$ is invertible.

$$\Rightarrow \mathbf{w} = (\mathbf{x}^T \mathbf{x})^{-1} \mathbf{x}^T \mathbf{t}$$

2 Exercise 4

Let X be a n x m matrix with columns are linearly independent (X is full rank) and $\mathbf{u}$ be a m x 1 vector.

$$X = \begin{bmatrix} | & | & \dots & | \\ \mathbf{x_1} & \mathbf{x_2} & \dots & \mathbf{x_m} \\ | & | & \dots & | \end{bmatrix}$$

$\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_m}$ are column vectors and linearly independent.

$\Rightarrow u_1 \mathbf{x_1} + u_2 \mathbf{x_2} + ... + u_m \mathbf{x_m} = \mathbf{0}$ has only solution: $\mathbf{u} = \mathbf{0}$.

$\Leftrightarrow X\mathbf{u} = \mathbf{0}$ has only solution $\mathbf{u} = \mathbf{0}$.

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

We have $X^T X$ is a m x m square matrix.

Let $\mathbf{v} \in N(X^T X)$. $\Rightarrow X^T X \mathbf{v} = \mathbf{0}$

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

$$\Leftrightarrow (X\mathbf{v})^T X \mathbf{v} = 0$$

$$\Leftrightarrow ||X\mathbf{v}||^2 = 0 \Rightarrow X\mathbf{v} = \mathbf{0}$$

$\Rightarrow \mathbf{v} \in N(A) \Rightarrow \mathbf{v} = \mathbf{0}$
$\Rightarrow N(X^TX) = N(X) = \{\mathbf{0}\}$
$\Rightarrow (X^TX)\mathbf{u} = \mathbf{0}$ has only solution $\mathbf{u} = 0$
$\Rightarrow$ columns of X^TX are linearly independent. In addition, X^TX is a square matrix.

Therefore, the reduced row echelon form of X^TX is equal to the m x m identity matrix. Equivalently, X^TX is invertible.