homework 3

Ngoc Anh Pham

September 2021

Ex1 : $t = y(x, w) + noise$

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

We have:

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

Precision parameter $\beta = \frac{1}{\sigma^2}$

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

The likelihood function:

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

It is convenient to maximize the logarithm of the likelihood function

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

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

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

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

We minimize $P = \frac{1}{2} \sum_{n=1}^{N} (y(x_n, w) - (t_n)^2)$ to find w. Suppose:

$$X = 1x_1 2x_2 ......1x_n, w = w_0 w_1$$

$$\Rightarrow P = \|Xw - t\|_2^2$$

$$\bigtriangledown P = 2X^T(Xw - t) = 2X^T Xw - 2X^T t$$

Setting this gradient to zero, we have:

$$X^T Xw - X^T t = 0$$

$$so : w = (X^T X)^{-1} X^T t$$

Ex4 : Prove that $X^T X$ is invertible when X is full rank

Suppose $X^T v = 0$. $\Rightarrow$XX$^T v = 0$
Conversely, suppose $XX^T v = 0$.
$\Rightarrow$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 $X^T X$.
But $X^T v = 0$ and $v \neq 0$ if and only if X has linearly dependent rows.
Thus, $X^T X$ is invertible if and only if X has full row rank.