

**Problem 1**

Transform linear regression by Latex, from  $t = y(x, w) + noise \Rightarrow w = (X^T X)^{-1} X^T t$

**Solution**

We have:

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

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

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:

$$X = \begin{bmatrix} 1 & x_1 \\ 2 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix}, w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

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

$$\nabla P_w = X^T(Xw - t) = X^T Xw - X^T t$$

Setting this gradient to zero, we have:

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

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

**Problem 2**

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

**Solution**

We have : 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  $X^T X$ .

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.