Homework 3

Thanh Hue

September 2021

1 Proof: $t = y(x, w) + noise \rightarrow \mathbf{w} = (\mathbf{x}^T x)^{-1} x^T t$

We have a data set of observations $x = (x_1, x_2, ..., x_N)^T$, representing N observations of the scalar variable x and their corresponding target values $t = (t_1, t_2, ..., t_N)^T$
Suppose that the observations are drawn independently from a Gaussian distribution.

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

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

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

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:

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

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

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

=

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

=

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

=

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

$=$

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

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

To $max\ logp(t|x, w, \beta)$ we minimize $(t_n - y(x_n, w))^2$

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

$$x = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}, w = \begin{bmatrix} w_1 \\ w_0 \end{bmatrix}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} w_1x_1 + w_0 \\ w_2x_2 + w_0 \\ . \\ . \\ . \\ w_nx_n + w_0 \end{bmatrix} = xw$$

$\Rightarrow$

$$L = \|t - xw\|_2^2$$

We have:

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

$\Rightarrow$

$$x^Tt = x^Txw$$

$\Rightarrow$

$$w = \frac{x^Tt}{x^Tx}$$

$\Rightarrow$

$$w = (x^Tx)^{-1}x^Tt$$

2 Proof: X^TX invertible when X full rank

Suppose:

$$X^Tv = 0$$

$\Rightarrow$

$$XX^Tv = 0$$

$\Rightarrow$

$$v^TXX^Tv = 0$$

$\Rightarrow$

$$(X^T v)^T (X^T v) = 0$$

That mean $X^T v = 0$ We have proved that $X^T v = 0$ if and only if v is in the null-space 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.