Problem 1. Re-transform linear regression to LaTex. Prove that $t = y(x, \omega) + noise \rightarrow \omega = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathrm{t}$

1 Solution

We have:
$y = w_0 + w_1 * x_1 + w_2 * w_2 + ... + w_i * x_i$
Suppose that the observations are drawn independently from a Gaussian distribution.

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

With $\beta = \frac{1}{\sigma^2}$

$$p(t|x, w, \beta) = N\left(y\left(x, w\right), \beta^{-1}\right) \tag{2}$$

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:
$p(t|x, w, \beta) = \prod_{n=1}^{N} N\left(y\left(x, w\right), \beta^{-1}\right)$
It is convenient to maximize the logarithm of the likelihood function:
$\log p(t|x, w, \beta) = \sum_{n=1}^{N} \log(N\left(y\left(x, w\right), \beta^{-1}\right))$
$= \sum_{n=1}^{N} \log(\frac{1}{\sqrt{2\pi\beta^{-1}}} e^{-\frac{(t_n - y(x_n, w))^2\beta}{2}})$

$$= \sum_{n=1}^{N} [\frac{1}{2} \log(2\pi\beta^{-1} - (t_n - y(x_n, w))^2 - \frac{\beta}{2}] \tag{3}$$

Since $\frac{1}{2}\log(2\pi\beta^{-1})$ & $\frac{\beta}{2}$ are just the numbers, so we ignore them and optimize the remain element $-\sum_{n=1}^{N}(t_n - y(x_n, w))^2)$. Concretely, we minimize $(t_n - y(x_n, w))^2$. So,

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

We also have:

$$x = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \ w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, \ y = \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix} = xw = \begin{bmatrix} w_0 + w_1 x_1 \\ \dots \\ w_0 + w_1 x_n \end{bmatrix} \quad (2)$$

Replace y in (1) by xw in (2). To minimize the loss function $\boldsymbol{L}$, its derivative equal to 0.

$$\frac{\delta L}{\delta w} = \begin{bmatrix} \frac{\delta L}{\delta w_0} \\ \frac{\delta L}{\delta w_1} \end{bmatrix} = \begin{bmatrix} t - xw \\ x(t - xw) \end{bmatrix} = 0 \tag{4}$$

Equivalently,

$$x^T(t - xw) = 0 \tag{5}$$

$$\leftrightarrow x^T t - x^T xw = 0 \tag{6}$$

$$\leftrightarrow x^T t = x^T xw \tag{7}$$

$$\leftrightarrow w = (x^T x)^{-1} x^T t \tag{8}$$

Problem 2. Prove that $X^T X$ invertible when X full rank.
Solution
If X is full rank, all the vectors that are in X will be linear independent.
Suppose $X^T v = 0 \rightarrow XX^T v = 0$
Conversely, suppose that XXT v = $0 \rightarrow v^T XX^T v = 0$, equivalent to $(X^T v)^T (X^T v) = 0$. This implies $X^T v = 0$
Hence, we have proved that $X^T v = 0 \Leftrightarrow v$ is in the nullspace of $XX^T v$.
But $X^T v = 0$ and $v \neq 0 \Leftrightarrow X$ has linearly dependent rows.
Thus XX^T has nullspace $0 \Leftrightarrow X$ has linearly independent rows.