Problem 1
Re-transform linear regression on the latex output layer, from

$$t = y(x, w) + noise- > w = (X^T X) - 1X^T t$$

Solution

We have:
+, Set a observation $x = (x_1, x_2, ..., x_N)^T$
+, Total observation N
+, Target values $t = (t_!, t_2, ...t_N)^T$

Suppose that the observations are drawn independently from a Gaussiandistribution.

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

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

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

We now use the training data x, t to determine the values of theunknown parameters w and by maximum likelihood. If the data areassumed to be drawn independently from the distribution then thelikelihood 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:

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

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

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

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

$$\longrightarrow we \quad minimize \quad (t_n - y(x_n, w))^2$$

Set:

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

with:

$$x = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ ... & ... \\ 1 & x_n \end{bmatrix}, \quad w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ ... \\ y_n \end{bmatrix} = \begin{bmatrix} w_1 x_1 + w_0 \\ w_1 x_2 + w_0 \\ w_1 x_3 + w_0 \\ ... \\ w_1 x_n + w_0 \end{bmatrix} = xw$$

we have:

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

$$\leftrightarrow x^T t = x^T xw$$

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

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.