Problem 1
Biến đổi lại linear regression trên lớp ra latex, từ t = y(x,w) + noise -> w =$(X^TX)^{-1}X^Tt$

Proof: We have: $t = y(x, w) + noise = N(t|y(x, w), \beta^{-1})$ **in which** (noise = N(0, β^{-1})t)
We can write: $p(t|x, w, \beta) = N(t|y(x, w), \beta^{-1})$
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:

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

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

$$maxlogp(t|x, w, \beta) = -max\frac{\beta}{2}\sum_{n=1} ny(x_n, w) - t_n{}^2$$

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

We minimize $P = \frac{1}{2}\sum_{n=1} ny(x_n, w) - t_n{}^2$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:

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

$$\Rightarrow x^Tt = x^Txw$$

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

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

Problem 4
Prove $X^T X$ invertible when X full rank.

Proof: Suppose $X^T v = 0$

Then, $XX^T v = 0$ too.

Opposite, suppose $XX^T v = 0$. Then $v^T XX^T v = 0$

So that, $(X^T v)^T (X^T v) = 0$, that mean $X^T v = 0$

Hence, we have proved that $X^T v = 0$ if and only if v is in the nullspace of XX^T

But $X^T v = 0$ and $v \neq 0$ if and only if X has linearly dependent rows.

Thus, XX^T is invertible if and only if X has full rows rank.