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

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:

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

$$\Rightarrow P = \|Xw - t\|_2^2$$
$$\triangledown P_w = X^T(Xw - t) = X^TXw - X^Tt$$

Setting this gradient to zero, we have:

$$X^TXw - X^Tt = 0$$
$$\Leftrightarrow w = (X^TX)^{-1}X^Tt$$

Problem 2
Prove that X^TX is invertible when X is full rank

Solution
We have : Suppose $X^Tv = 0$.
Then, of course, $XX^Tv = 0$ too.
Conversely, suppose $XX^Tv = 0$.
Then $v^TXX^Tv = 0$, so that $(X^Tv)^T(X^Tv) = 0$.
This implies $X^Tv = 0$.
Hence, we have proved that $X^Tv = 0$ if and only if v is in the nullspace of X^TX.
But $X^Tv = 0$ and $v \neq 0$ if and only if X has linearly dependent rows.
Thus, X^TX is invertible if and only if X has full row rank.