1 Prove that

$$t_n = y(x,w) + \epsilon \Longleftrightarrow W = \left(X^t X\right)^{-1} X^T t$$

Suppose that the observations are drawn independently from a Gaussian distribution we have:

$$p\left(t_n\right) = \mathcal{N}\left(t_n \mid y(x,w), \epsilon^2\right)$$
$$p(t \mid x, w, \beta) = \prod_{n=1}^{N} \mathcal{N}\left(t_n \mid y\left(x_n, w\right); \sigma^2\right)$$
$$\beta^- 1 = \frac{1}{\sigma^2}$$

Maximize the logarithm of the likelihood function:

$$\begin{aligned}
\log \prod_{i=1}^{N} \mathcal{N}\left(t_n \mid y\left(x_n, w\right), \beta^{-1}\right) &= \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\beta^{-1}}} e^{\left(t_n - y\left(x_n, w\right)\right)^2 \frac{\beta}{2}}\right. \\
&= \sum_{i=1}^{N}\left(\frac{-1}{2}\log\left(2\pi\beta^{-1}\right) - \left(t_n - y\left(x_n, w\right)\right)^2\right. \\
&= -\sum_{i=1}^{N}\left(t_n - y\left(x_n - w\right)\right)^2
\end{aligned}$$

Minimize: $\sum_{i=1}^{N}\left(t_n - y\left(x_n - w\right)\right)^2$

$$\mathrm{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{bmatrix}, \mathrm{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}, \mathrm{t} = \begin{bmatrix} t_1 \\ t_2 \\ \dots \\ t_n \end{bmatrix}$$
$$\mathrm{y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} w_1 x_1 + w_0 \\ w_1 x_2 + w_0 \\ \dots \\ w_1 x_n + w_0 \end{bmatrix} = \mathrm{Xw}, \mathrm{t} - \mathrm{y} = \begin{bmatrix} t_1 - y_1 \\ t_2 - y_2 \\ \dots \\ t_n - y_n \end{bmatrix}$$

$$\begin{aligned}
\Longrightarrow \|t - y\|_2^2 = \sum_{i=1}^{n}\left(t_i - y_i\right)^2 = \textit{Loss Function} \Longrightarrow L &= \|t - y\|_2^2 = \|t - XW\|_2^2 \\
&\Longrightarrow \frac{\partial L}{\partial W} = 2X^T(t - XW) = 0 \\
&\Longrightarrow X^T t = X^T X W \\
&\Longrightarrow W = \left(X^t X\right)^{-1} X^T t
\end{aligned}$$

2 Prove that $X^t X$ is invertable when X full rank

Solution
If X is full rank, X is linear independent.

$$\Rightarrow \mathrm{N(X)} = \{\overrightarrow{0}\}$$

Suppose $\vec{v} \in N\left(X^T X\right)$

$$\Rightarrow X^T X \vec{v} = \overrightarrow{0}$$

$$
\begin{aligned}
&\Rightarrow \vec{v}^T X^T X \vec{v} = \vec{v}^T \overrightarrow{0} = 0 \\
&\Rightarrow (X\vec{v})^T X\vec{v} = 0 \\
&\Rightarrow (X\vec{v}) \cdot (X\vec{v}) = 0 \\
&\Rightarrow X\vec{v} = \overrightarrow{0}
\end{aligned}
$$

So we have: if $\vec{v} \in N\left(X^T X\right) \Rightarrow \vec{v} \in N(X)$

$$
\begin{aligned}
&\Rightarrow \vec{v} \text{ can only be } \overrightarrow{0} \\
&\Rightarrow N\left(X^T X\right) = N(X) = \{\overrightarrow{0}\}
\end{aligned}
$$

$\Rightarrow X^T X$ is linearly independent; and $X^T X$ is a square matrix $\Rightarrow X^T X$ is invertible