Exercise 1:

We have : $t = y(w,x) + \varepsilon$
while $\varepsilon \approx N(0, \sigma^2)$
If $\mu(\varepsilon) \neq 0$ we just adjust bias of y : $\mu(\varepsilon) = 0$
$\Rightarrow P(t) = N(t|y(w,x), \sigma^2)$
Suppose: $t_n = y(x_n, w) + \varepsilon$
$\Rightarrow P(t_n) = N(t_n|y(x_n, w), \varepsilon^2)$
Generality: Maximum for all point, we use maximum likelihood function:
$P(t|x, w, \beta) = \prod_{n=1}^{N} N(t_n|y(x,w), \beta^{-1})$
Simplize:
$\log(P(t|x,w,\beta)) = \sum_{n=1}^{N}(log(N(t_n|y(x,w), \beta^{-1})))$
$= \dfrac{-\beta}{2} \sum_{n=1}^{N}(y(x_n, w) - t_n)^2 + \dfrac{N}{2}\log(\beta) - \dfrac{N}{2}\log(2\pi)$
Maximum likelihood:
Max $\log(P(t|x,w,\beta))$ = -Max $\dfrac{\beta}{2}\sum_{n=1}^{N}(y(x_n,w) - t_n)^2$
= Min $\dfrac{1}{2}\sum_{n=1}^{N}(y(x_n,w) - t_n)^2$
We minimize P = $\dfrac{1}{2}\sum_{n=1}^{N}(y(x_n,w) - t_n)^2$ to find w
Suppose:

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ ... & ... \\ 1 & x_n \end{pmatrix}; t = \begin{pmatrix} t_1 \\ t_2 \\ ... \\ t_n \end{pmatrix}; w = \begin{pmatrix} w_0 \\ w_1 \end{pmatrix}$$

By minimizing P, we can find w. P is called Mean Squared Error Loss(MSE):
$L = \dfrac{1}{N}\sum_{n=1}^{N}(t_n - y(x_n, w))^2$

we have:
$y(x_n, w) = w_1 x_v + w_0$

$$y = \begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{pmatrix} = \begin{pmatrix} w_1 x_1 & w_0 \\ w_2 x_2 & w_0 \\ ... & ... \\ w_n x_n & w_0 \end{pmatrix} = XW$$

$$t - y = \begin{pmatrix} t_1 - y_1 \\ t_2 - y_2 \\ ... \\ t_n - y_n \end{pmatrix}$$

$$\Rightarrow L = \|t - y\|_i^2 = \|t - Xw\|_i^2 = (t - Xw)^T(t - Xw)$$

$$\frac{\partial(L)}{\partial(w)} = 2X^T(t - Xw) = 0$$

$$\Leftrightarrow X^T t = X^T X w$$

$$\Leftrightarrow w = (X^T X)^{-1} X^T t$$

Exercise 4:

We know that when the rows and columns are independent, then matrix A would be invertible and would have non-zero determinant.
We have an nxm matrix X
If n = m, It is enough for check $det(X) \neq 0$
if $n \neq m$ Precisely, When the rank of X is m (Which forces n >= m Observation: For $v \epsilon R^m, Xv = 0$ if and only if $X^T Xv = 0$. For the non-trivial implication. If $X^T Xv = 0$ then $v^T X^T Xv = 0$, that is $(Xv)^T Xv = 0$, Which implies that $Xv = 0$.

If $rank(X) = m$, this mean that X is one-to-one when acting on R^m. So by the observation, $X^T X$ is one-to-one, Which makes it Invertible (as it is square).

Conversely, If $rank(X) < m$ there exist $v \epsilon R^m$ with $Xv = 0$, and $X^T X$ can not Invertible.