Exercise 1

$$t = y(x,w) + noise$$
$$noise \sim N(0,\ \sigma^2)$$
$$\rightarrow p(t) \sim N(t|y(x,w),\ \sigma^2)$$

For $t \approx y(x,w)$

$$p(t|x,w,\beta)\ max$$
$$\rightarrow \prod \left(t_i|y(x_i,\ w),\ \sigma^2\right)\ max$$
$$log\ p(t|x,w,\beta) = \sum log\left(N\left(t_i|y(x_i,w),\beta^{-1}\right)\right)$$
$$= -\frac{\beta}{2}\sum [y(x_i,w) - t_i]^2 + const$$
$$max\ log\ p(t|x,w,\beta) = max\ -\frac{\beta}{2}\sum [y(x_i,w) - t_i]^2$$
$$= min \sum [y(x_i,w) - t_i]^2 \qquad (1)$$

To minimize (1), suppose

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ ... & ... \\ 1 & x_n \end{bmatrix} \qquad w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$
$$\rightarrow (1) = min\ ||Xw - t||^2$$
$$S = ||Xw - t||^2 = (Xw - t)^T(Xw - t)$$
$$= t^Tt - (Xw)^Tt - t^T(Xw) + (Xw)^T(Xw)$$
$$= t^Tt - 2(Xw)^Tt + (Xw)^T(Xw)$$
$$= t^Tt - 2w^TX^Tt + w^TX^TXw$$

To minimize S, w must satisfy

$$\frac{\partial S}{\partial w} = -2X^Tt + 2X^TXw = 0$$
$$\leftrightarrow 2X^TXw = 2X^Tt$$
$$\leftrightarrow w = \left(X^TX\right)^{-1}X^Tt$$

Exercise 4

$$
\begin{gathered}
X^T X \textit{ invertible} \\
\leftrightarrow |X^T X| \neq 0 \\
\leftrightarrow |X^T||X| \neq 0 \\
|X^T| = |X| \leftrightarrow |X|^2 \neq 0 \\
\leftrightarrow |X| \neq 0 \\
\leftrightarrow X \textit{ full rank}
\end{gathered}
$$