1.1
The number of training data < dimensionality of ω

1.2
We know
$$RSS = \sum_n (y - (\omega^T x_n + b))^2$$
Then take the derivative
$$\frac{\partial RSS}{\partial b} = \sum_n 2(\omega^T x_n + b - y_n) = 0$$
$$= \sum_n 2(\omega^T x_n) + \sum_n 2(b - y_n) = 0$$
From the question we know:
$$\frac{1}{N} \sum_n x_{nd} = 0$$
$$1_N X = 0_N$$
$$1_N X \omega = \sum_n (\omega^T x_n) = 0$$
plug in:
$$\sum_n 2(b^* - y_n) = 0$$
$$b^* = \frac{1}{N} \sum_n y_n$$

2.1
The error function:
$$min\epsilon(b) = -\sum_n [y_n log\sigma(b) + (1 - y_n) log(1 - \sigma(b))]$$
Take derivatives:
$$\frac{\partial \epsilon}{\partial b} = \sum_n [\frac{y_n \sigma'(b)}{\sigma(b)} - (1 - y_n)\frac{\sigma'(b)}{1 - \sigma(b)}] = 0$$
$$\sum_n [\frac{y_n}{\sigma(b)} - \frac{1 - y_n}{1 - \sigma(b)}] = 0$$

$$\sum_n [y_n - \sigma(b)] = 0$$

$$b = \sigma^{-1}(\frac{\sum_n y_n}{N}) = log(\frac{\frac{\sum_n y_n}{N}}{1 - \frac{\sum_n y_n}{N}})$$

The optimal classifier will be assign the sample to y=1 in probability $\frac{\sum_n y_n}{N}$,y=0 in probability $1 - \frac{\sum_n y_n}{N}$ The probability of assigning a sample to y=1 is $\frac{\sum_n y_n}{N}$