# CSCI567 Homework 1

## Question 1.1

In the case where the matrix is not invertable, one can say that the w matrix, has rows that are dependent on one another,meaning its true dimensionality is less than that of each X.

## Question 1.2

Starting with the residual sum of squares objective function we get:

$$b^* = argmin_b \left\| \boldsymbol{y} - (\boldsymbol{wx} + b\boldsymbol{1_N}) \right\|^2$$

Then, taking a derivative with respect to b we get:

$$\boldsymbol{1_N^T} \ (\boldsymbol{y} - \boldsymbol{wx} - b^*\boldsymbol{1_N}) = 0$$

Then using:

$$\frac{1}{N} \sum_n x_n = 0$$

the equation is reduced to:

$$\boldsymbol{1_N^T} \ (\boldsymbol{y} - b^*\boldsymbol{1_N}) = 0$$

which is exactly the same equation as which was given in the 0 dimension case. With that, we find that:

$$b^* = \frac{1}{N} \boldsymbol{1_N^T y} = \frac{1}{N} \sum_n y_n$$

## Question 2.1

We start with the objective function:

$$\min_{w,b} \varepsilon(w,b) = \min_{w,b} - \sum_n \{y_n \log\left(\sigma\left(w^T x_n + b\right)\right) + (1 - y_n) \log\left(1 - \sigma\left(w^T x_n + b\right)\right)\}$$

Taking a derivative and simplifying we get:

$$0 = -\sum_n \{y_n - \sigma\left(w^T x_n + b^*\right) y_n + \left((-\sigma\left(w^T x_n + b^*\right) + y_n(\sigma\left(w^T x_n + b^*\right)\right)\}$$

$$0 = -\sum_n \{y_n - \sigma\left(w^T x_n + b^*\right)\}$$

The simplest way for the above statement to be true is for each of the summed elements to equal zero:

$$y_n = \sigma\left(w^T x_n + b^*\right)$$

with that we get the optimal b:

$$b = \log\left(\frac{y_n}{1 - y_n}\right) - w^T x_n$$