1.a)

Open form of the perceptron equation is

$$w^T * x = w_0 + w_1 x_1 + w_2 x_2 \qquad (1)$$

Since the sign function gives output as +1 when it is greater than 0 and gives output as -1 when it is smaller than 0, the separation boundary will be according to the equation when its value is 0.

Hence, the mathematical expression of the decision boundary is

$$w_0 + w_1 x_1 + w_2 x_2 = 0 \qquad (2)$$

When we leave x_2 alone:

$$x_2 = -\frac{w_1}{w_2} x_1 - \frac{w_0}{w_2} \qquad (3)$$

Thus,

$$a = -\frac{w_1}{w_2} \text{ and } b = -\frac{w_0}{w_2}$$

2.b)

To reduce the dimension of the dataset, we decided to use Principal Component Analysis(PCA). PCA reduces the dimensions by determining principal components, which are the linear combinations of original features(gray-scaled image matrices in our case). Principal components will be chosen to be orthogonal to each other, which is a way to reinforce the independence of principal components.

After reducing the dimensions to a certain number, the RMSE value for the test dataset with new dimensions is relatively close to the original solution: 0.34946689514284945 for the original space and 0.356855932126467 for the new space. However, the time complexity is much smaller than the previous because we have fewer elements in matrix operations while minimizing the error function.

3.c)

The original expression:

$$E[w] = \frac{1}{N}\sum_{n=1}^{N} ln(1 + e^{(-y_n.w^T x_n)}) + \lambda||w||_2^2$$

Taking the gradient with respect to w:

$$\nabla(E[w]) = \nabla[\frac{1}{N}\sum_{n=1}^{N} ln(1 + e^{(-y_n.w^T x_n)})] + \nabla[\lambda||w||_2^2]$$

$$= \frac{1}{N}\sum_{n=1}^{N} \nabla[ln(1 + e^{(-y_n.w^T x_n)})] + \lambda\nabla[||w||_2^2]$$

$$= \frac{1}{N}\sum_{n=1}^{N} \frac{\nabla e^{(-y_n.w^T x_n)}}{1+e^{(-y_n.w^T x_n)}} + \lambda\sum_{i=1}^{m} \frac{\delta(w_i^2)}{\delta w_j}$$

$$= \frac{1}{N}\sum_{n=1}^{N} \frac{(\nabla(-y_n.w^T x_n))e^{(-y_n.w^T x_n)}}{1+e^{(-y_n.w^T x_n)}} + \lambda\sum_{i=1}^{m} 2[0\ 0\ 0 \ldots w_i \ldots\ 0\ 0]$$

$$= \frac{1}{N}\sum_{n=1}^{N} \frac{(-y_n x_n)e^{(-y_n.w^T x_n)}}{1+e^{(-y_n.w^T x_n)}} + 2\lambda w$$