Modern Machine Learning
Homework #4

1. The book and slides use the sum-of-squared as the objective function (L_2 error function), equation (11.9). They can be used for regression or classification problems. In classification applications, cross-entropy is generally preferred versus some-of-squared. See equation (11.10). Also, we generally prefer the softmax activation function (which is a normalized version of the sigmoid). To become familiar with the widely used cross-entropy and softmax optimization, complete book problem 11.3.

2. Consider a neuron with 2 inputs, 1 output and a threshold activation function $g(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$. If the 2 input weights are $\theta_1 = \theta_2 = 1.5$, and the bias $\theta_0 = -2.0$, calculate the outputs to the following inputs

   (a) $\mathbf{x}^{(1)} = (0, 0)$

   (b) $\mathbf{x}^{(2)} = (1, 0)$

   (c) $\mathbf{x}^{(3)} = (0, 1)$

   (d) $\mathbf{x}^{(4)} = (1, 1)$

   Draw the discriminant function and write down its equation. Does it correspond to any particular logic gate?

   Hint: $y = g(\theta_0 + X_1\theta_1 + X_2\theta_2)$

   Given $\mathbf{x}^{(1)} = (0, 0)$, we have $y = g(-2.0+0\times1.5+0\times1.5) = g(-1.5) = 0$

3. Design a perceptron that emulates the behavior of:

   (a) NOT logic gate

(b) AND logic gate

Assume that the input vector $\mathbf{x} \in \mathbb{R}^2$ is binary.

Hint: (a) The perceptron has two inputs, two outputs.

$$\begin{aligned} y_1 &= g(\theta_{10} + X_1\theta_{11} + X_2\theta_{12}) \\ y_2 &= g(\theta_{20} + X_1\theta_{21} + X_2\theta_{22}) \end{aligned} \tag{1}$$

Since we are required to design a NOT logic gate, y_1 is only related to X_1, thus we have $\theta_{12} = \theta_{21} = 0$

$$\begin{aligned} y_1 &= g(\theta_{10} + X_1\theta_{11}) \\ y_2 &= g(\theta_{20} + X_2\theta_{22}) \end{aligned} \tag{2}$$

We have four cases for the input and output, thus we have four equations to solve the four parameters.

(b) The perceptron has two inputs, one output.

$$y = g(\theta_0 + X_1^1\theta_1 + X_2^1\theta_2) \tag{3}$$

4. Look up and list five commonly used activation functions — there are many resources online listing activation functions. Write 1-2 sentence arguments about each activation function, e.g., advantages and disadvantages.