CSC242: Homework 4.3
AIMA 18.6.3–18.7

1. Briefly define the following terms:
   (a) Decision boundary
   (b) Linear separator
   (c) Linearly separable
2. Explain briefly the relationship between linear regression (learning a linear model that fits the data) and classification. You might want to use some of the terms you defined above.
3. Explain briefly the attraction(s) of logistic regression for classification. Is there a downside to using it?
4. The best way to understand neural networks is to actually implement them, which you can do as part of your Project 4. It is easy to implement a "unit" with inputs and weights that computes the thresholded, weighted sum of its inputs. A neural network is simply a graph of such units. You should understand that it's only possible to learn the weights for certain topologies of network.

   Here are some additional problems about neural networks for you.
5. Draw a diagram of a single-layer, fully-connected, feed-forward neural network with n input units and m output units. You only need to draw the connections between the input layer and the first unit of the output layer (the other output units are similar). Label these connections with their weights using a reasonable notation.
   (a) Give an expression for activation (output) of the j-th output unit if the activation function is g.
   (b) Suppose there are two input and two output units, and the units use a linear activation function. That is, for each unit the output is some constant c times the weighted sum of the inputs. (Note: normally you would have a nonlinear activation function). Give an explicit expression (no big sum or product symbols) for the activation of the first output unit.
   (c) How would you train such a network?
6. Draw a diagram of a fully-connected, feed-forward neural network with n input units, m output units, and one hidden layer of k units. You only need to draw the connections for the first units of the of hidden and output layers (the rest are similar). Label the connections with their weights using a reasonable notation.

7. Give an expression for activation (output) of the j-th output unit if the activation function is g.

   (a) Suppose there are two input, two hidden, and two output units, and the units use a linear activation function (as above). Give an explicit expression for the activation of the first output unit.

   (b) How would you train such a network?

8. Minksy and Papert (1969) showed that a single-layer perceptron network could not learn the XOR (exclusive-OR) function. Show how a feed-forward network with two inputs units, two hidden units, and one output unit (all with sigmoid thresholds) can compute the XOR function. Hint: Connection weights can be negative. These are called “inhibitory” connections.