CS360 Homework 5

Artificial Neural Networks

1) Can a decision tree represent the Boolean function $f(P, Q) \equiv P \Rightarrow Q$? What about a single perceptron with a step function (whose step is at zero) as threshold function? What about a network of perceptrons with step functions (whose steps are at zero) as threshold functions? Now answer the same three questions for the Boolean function $f(P, Q) \equiv P \Leftrightarrow Q$.

2) Develop the training rule for a perceptron with a sigmoid function as threshold function and the sum of errors to the power of four (instead of the sum of squared errors) as error function.

3) (Courtesy of Russell and Norvig) Suppose that a training set contains only a single example, repeated 100 times. In 80 of the 100 cases, the single output value is 1; in the other 20, it is 0. What will a neural network predict for this example, assuming that it has been trained on all training examples and reaches a global optimum? (Hint: To find the global optimum, differentiate the error function and set the resulting expression to zero.)

4) If we train a neural network for 1,000 epochs (one training example at a time), does it make a difference whether we present all training examples in turn for 1000 times or whether we first present the first training example 1000 times, then the second training example for 1000 times, and so on? Why?

5) You are given n numbers and have to determine their mean. Develop a gradient descent rule for this purpose.

6) Explain exactly why networks of perceptrons with linear activation functions are uninteresting (that is, networks of perceptrons where, for each perceptron, the output is some constant times the weighted sum of the inputs). Use equations if necessary.

7) Is overfitting more or less likely when the training set is small or large? Is overfitting more or less likely when the number of parameters to learn (such as the number of weights in a neural network) is small or large?