PRACTICE QUESTIONS

1. Suppose you are given 6 data points in 1-dimension as follows.

x	y
-2	-1
-1	+1
0	+1
1	+1
2	-1
3	-1

In the above table, x represents the attribute and y represents the class label. Answer the following questions:

a. Can a perceptron achieve 0 training error on this dataset? If so, give the parameters.

b. Can a decision tree achieve 0 training error on this dataset? If so, draw the most compact tree.

c. Can a SVM with linear kernel achieve 0 training error on this dataset? If so, draw the maximum margin classifier. If not, suggest a kernel function that will achieve 0 training error

2. Suppose there is a classification problem f: X -> Y, where each instance X has 4 Boolean attributes i.e. X = (x1, x2, x3, x4).

a. What is the number of possible instances?

b. You decide to propose hypotheses of the form:

IF (x1, x2, x3, x4) = (b1, b2, b3, b4) THEN class = 1 ELSE class = 0
where each bi can be 0, 1, or ? (Don't care).

How many possible hypotheses can you generate?

3. Consider classification problem f : X -> Y , where each instance X has 4 attributes i.e. $X = (x_1, x_2, x_3, x_4)$ where each x_i is a boolean.

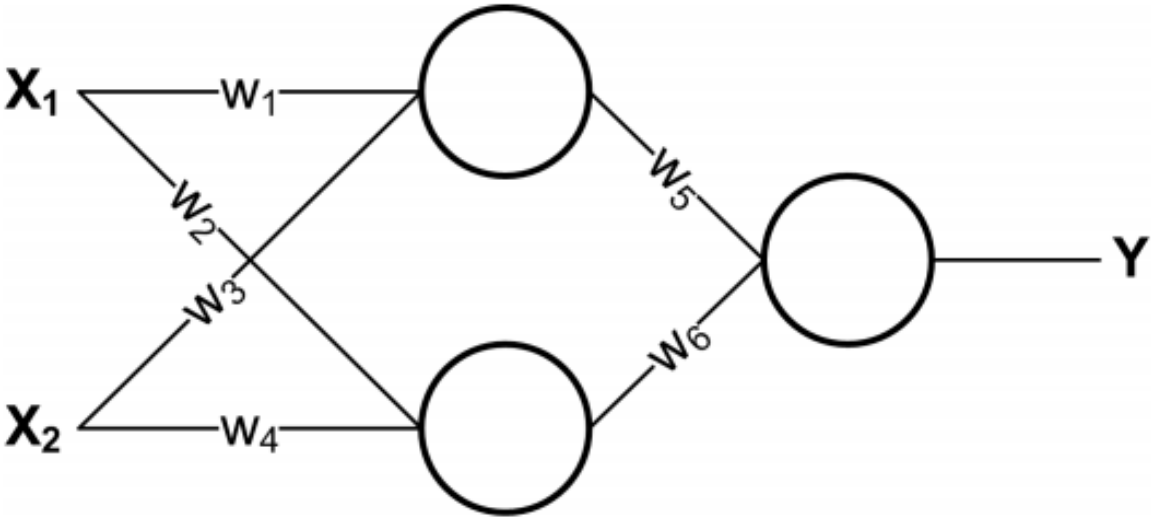
Also assume that the classifier consists of only AND-POSITIVE literals of variable length. Some examples of this type are:

h1(x) = (x1 ^ x2)
h2(x) = (x2 ^ x3 ^ x4)

Remember in AND-POSITIVE conjunctions, an attribute cannot be used with a negation.

What is the size of the hypothesis state i.e. how many hypotheses can you generate?

4. You would like to design a neural network to represent a function f: X->Y where X consists of two attributes i.e. X = (x1, x2). The neural network has one hidden layer with two nodes. The neural net can be represented as below:



The input to each node is $\sum_i w_i X_i$ where i is the set of inputs. For example, the top hidden node gets the input as $w_1 X_1 + w_3 X_2$. For the activation function of each node, you have two choices:

- S: Signed sigmoid function, which is defined as:

$$S(a) = sign(\sigma(a) - 0.5) = sign(\frac{1}{1 + \exp(-a)} - 0.5)$$

- L: Linear function, which is defined as:

$$L(a) = ca$$

In both cases above, a is the input.

a. You would like to represent the following function: $Y = \beta_1 X_1 + \beta_2 X_2$. Indicate for each of the three nodes whether you would use the S or L activation function. Also compute values of β_1 and β_2 in terms of weights on the network.

** Advanced question **

b. Next, you would like to represent the binary logistic function using this neural network. The function is defined as:

$$Y = \arg\max_{y} P(Y = y|X), \text{ where}$$

$$P(Y = 1|X) = \frac{1}{1+\exp(-(\beta_1 X_1+\beta_2 X_2))} \text{ and } \mathrm{P(Y=0|X)} = \frac{1}{1+\exp((\beta_1 X_1+\beta_2 X_2))}$$

Indicate whether each of the nodes would use S or L function. Also compute values of β_1 and β_2 in terms of weights on the network.