MASSACHVSETTS INSTITVTE OF TECHNOLOGY
Department of Electrical Engineering and Computer Science
6.034—Artificial Intelligence
Fall Semester, 1996

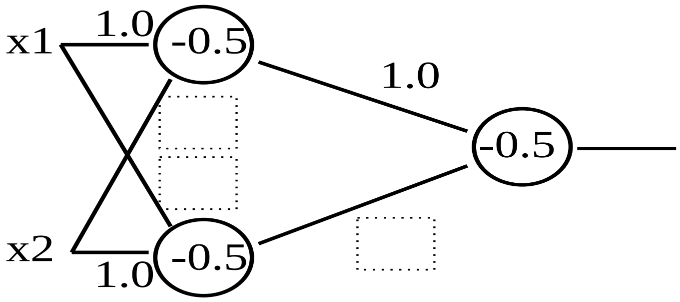
(Last !) Practice Problems – Dec 10, 1996

Problem 1: Artificial Neural Network

(a) We already know that a 1-layer network (i.e., a perceptron) cannot compute the logical function XOR. Are there any other logical functions of two boolean variables that a 1-layer network cannot compute? If yes, what are these functions? If no, explain why not.

Answer:

(b) Consider the 2-input 2-layer network:

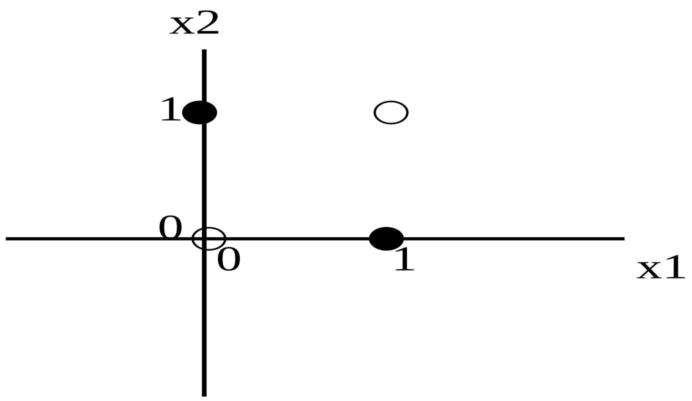


Assume all the weights are between -2 and 2. Fill in the three missing weights so that the network computes XOR.

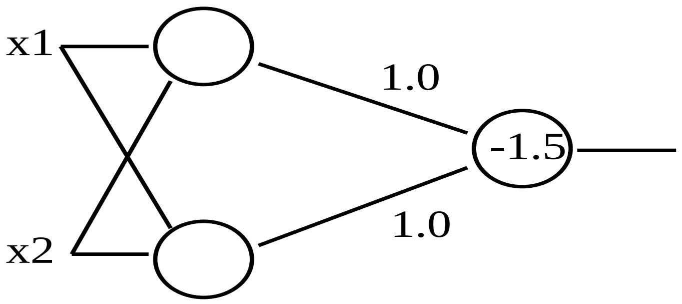
(c) Which logical functions of the inputs x_1 and x_2 do the hidden units compute?

Answer:

(d) Draw in the input space below the decision boundaries implied by the network in part (b). Shade the decision regions where the output of the network is 1.



(e) Consider the partial weight assignments for the following network. Is it possible to complete the weight assignments to give another implementation of XOR? If yes, fill in the missing weight assignments. If no, explain why not.



Problem 2: Learning

In this problem we related the performance of three learning procedures: procedure W, identification tree, and nearest neighbor. The training data set is attached on a separate sheet so that you can tear it off for handy reference.

To facilitate comparisons, we encode data into binary features. We will also assume the following AKO tree:

```
               Thing
               /   \
             Block  Curved
            /   \    /    \
       Brick  Wedge Cone  Cylinder
```

(i) Suppose the training sequence given to procedure W is:
Samples 1, 5, 6, 2, 3, 4

Draw the arch concept that procedure W learns after the training.

Answer:

(ii) Is it true that a different training sequence always lead procedure W to learn a different concept?

Answer:

(iii) Assume the disorder measure for an attribute to be $\frac{p}{q}$ where p = number of members of the less common class, and q = number of members of the most common class. Draw the identification tree for the training set.

Answer:

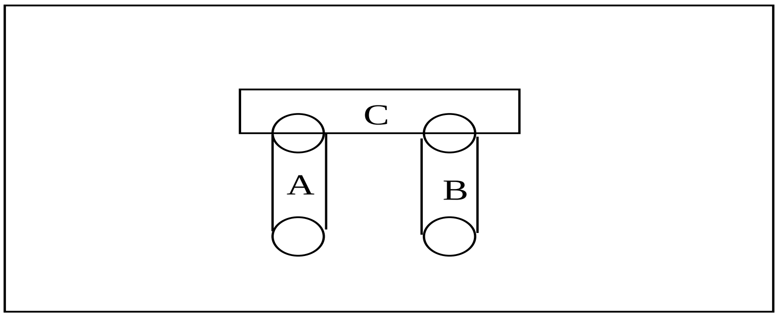
(iv) Is it true that one can always build an identification tree from a procedure W model? Why or why not? How are the required and forbidden links reflected in a identification tree?

Answer:

(v) Is it true that one can always build a procedure W model from an identification tree? Why or why not?

Answer:

(vi) Suppose we use the hamming distance of two binary feature vectors, i.e., number of bits that are different, as a metric. Consider the following sample:



What is the classification of this sample by the nearest neighbor algorithm?

Answer:

(vii) Compare the classification results obtained by procedure W, identification, and nearest neighbor on the sample in part (vi). Do they give the same result? If you think nearest neighbor gives a different answer, is there a way to adjust the distance metric to give an answer consistent with the other two learning methods?

Answer: