CS 188 Spring 2021 | Introduction to Artificial Intelligence | Exam Prep 10

Q1. Decision Trees

(a) You are given a dataset for training a decision tree. The goal is to predict the label (+ or -) given the binary features A, B, and C.

A	*B*	*C*	label
0	0	0	+
0	0	1	+
0	1	0	+
0	1	1	-
1	0	0	-
1	0	1	-
1	1	0	+
1	1	1	-

First, consider building a decision tree by greedily splitting according to information gain.

(i) Which features could be at the root of the resulting tree? Select all possible answers.

A ☐
B ☐
C ☐

(ii) How many edges are there in the longest path of the resulting tree? Select all possible answers.

1 ☐
2 ☐
3 ☐
4 ☐
5 ☐
None of the above

Now consider building a decision tree with the smallest possible height.

(iii) Which features could be at the root of the resulting tree? Select all possible answers.

A ☐
B ☐
C ☐

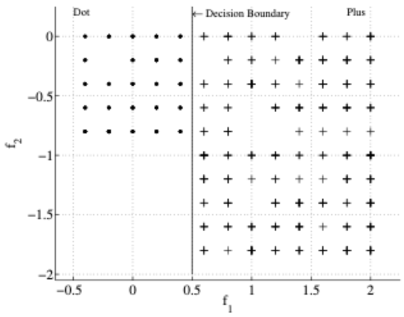
(iv) How many edges are there in the longest path of the resulting tree? Select all possible answers.

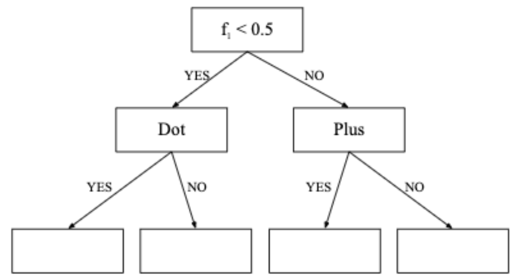
1 ☐
2 ☐
3 ☐
4 ☐
5 ☐
None of the above

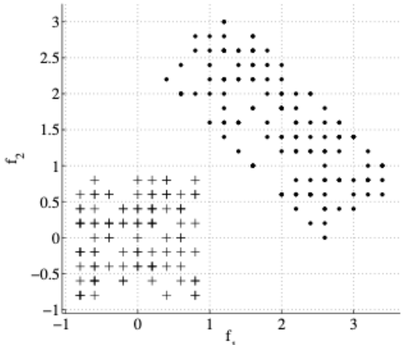
(b) You are given points from 2 classes, shown as +'s and ·'s. For each of the following sets of points,

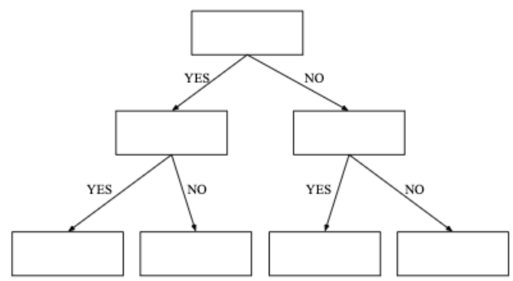
1. Draw the decision tree of depth at most 2 that can separate the given data completely, by filling in binary predicates (which only involve thresholding of a *single* variable) in the boxes for the decision trees below. If the data is already separated when you hit a box, simply write the class, and leave the sub-tree hanging from that box empty.
2. Draw the corresponding decision boundaries on the scatter plot, and write the class labels for each of the resulting bins somewhere inside the resulting bins.

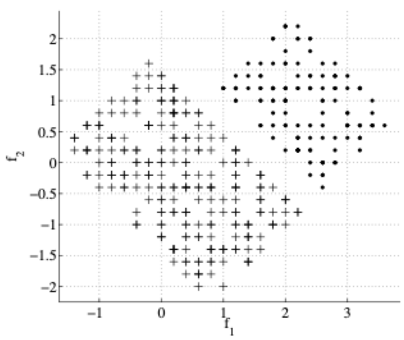
If the data can not be separated completely by a depth 2 decision tree, simply cross out the tree template. We solve the first part as an example.

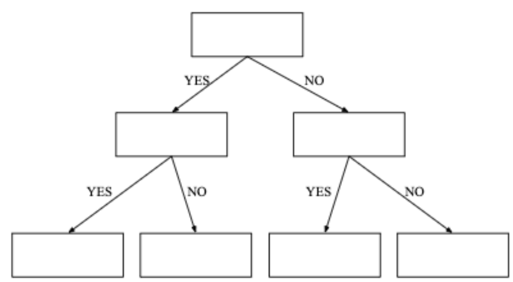


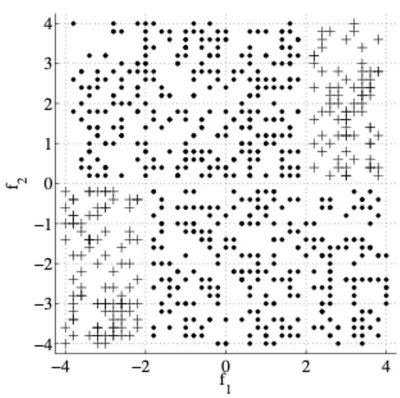


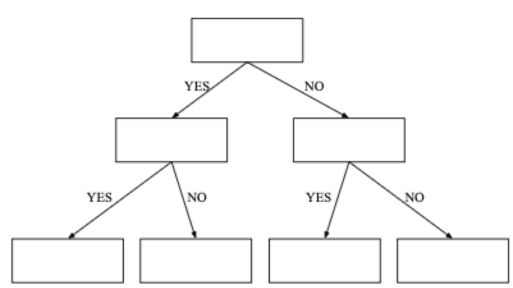








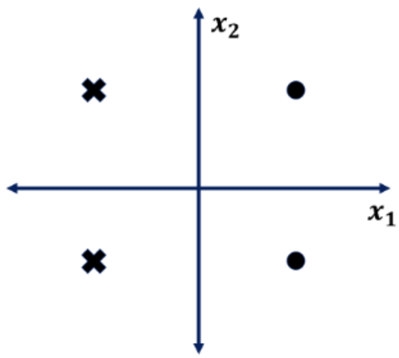




Q2. Perceptron Feature Maps

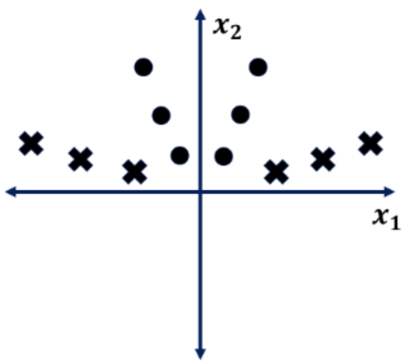
(a) For each of the datasets represented by the graphs below, please select the feature maps for which the perceptron algorithm can perfectly classify the data.

Each data point is in the form (x_1, x_2), and has some label Y, which is either a 1 (dot) or -1 (cross).



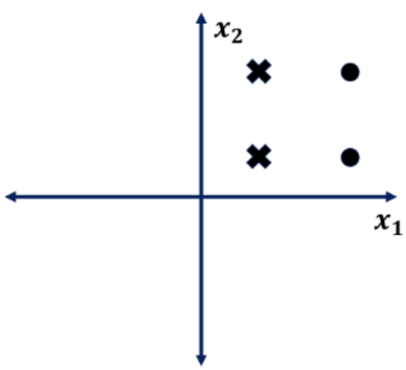
(i)

- ☐ $[x_1 \quad x_2 \quad 1]$
- ☐ $[x_1 \quad x_2 \quad x_1^2]$
- ☐ $[x_1 \quad x_2 \quad |x_1|]$
- ☐ $[x_1 \quad x_2 \quad Y]$
- ☐ $[x_1 \quad x_2]$



(ii)

- ☐ $[x_1 \quad x_2 \quad 1]$
- ☐ $[x_1 \quad x_2 \quad x_1^2]$
- ☐ $[x_1 \quad x_2 \quad |x_1|]$
- ☐ $[x_1 \quad x_2 \quad Y]$
- ☐ $[x_1 \quad x_2]$



(iii)

- ☐ $[x_1 \quad x_2 \quad 1]$
- ☐ $[x_1 \quad x_2 \quad x_1^2]$
- ☐ $[x_1 \quad x_2 \quad |x_1|]$
- ☐ $[x_1 \quad x_2 \quad Y]$
- ☐ $[x_1 \quad x_2]$