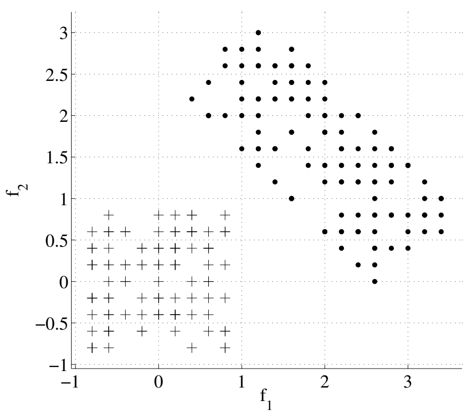
ARTIFICIAL INTELLIGENCE (E016350)
Prof. Aleksandra Pizurica
AY 2024/2025

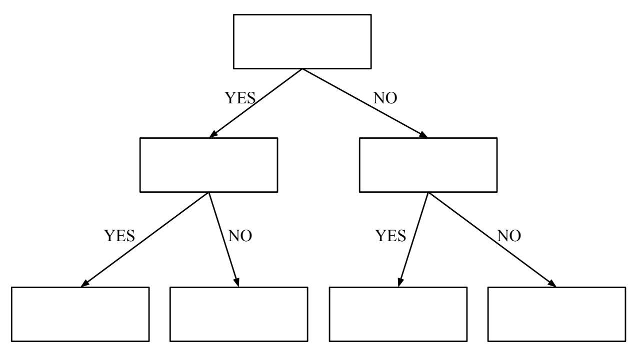
Solutions: Decision trees

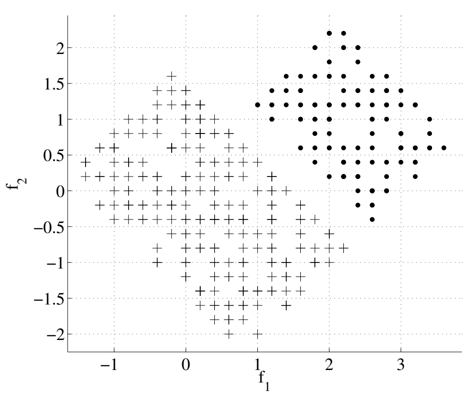
1. You are given points from 2 classes, shown as +'s and ·'s. For each of the following sets:

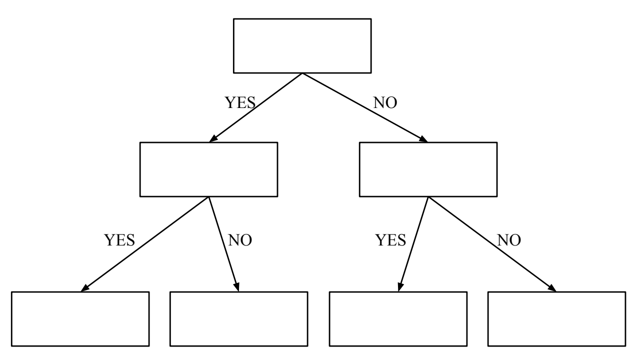
    1. Draw the decision tree of depth at most 2 that can separate the given data completely, by filling in binary predicates (which only involve thresholding of a single variable) in the boxes for the decision trees below. If the data is already separated when you hit a box, simply write the class, and 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.

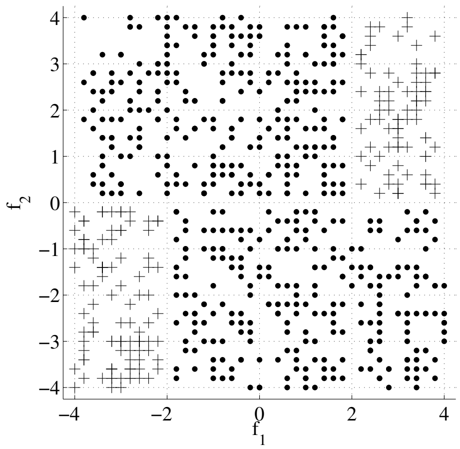
    If the data can not be separated completely by a depth 2 decision tree, simply cross out the tree template.



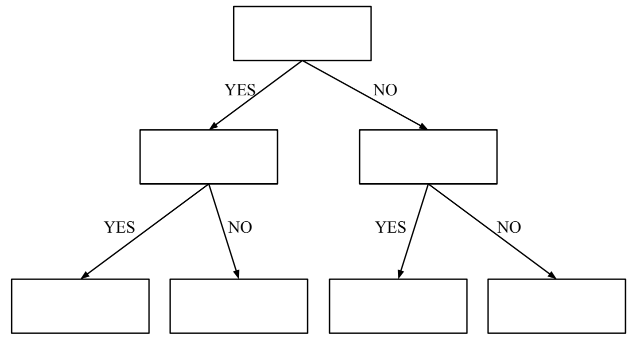






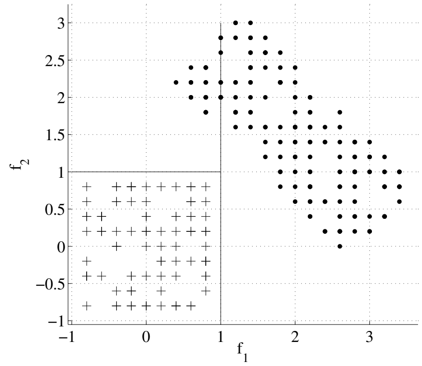


f_2
f_1

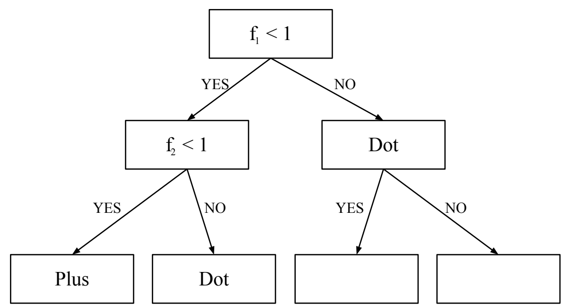


YES
NO
YES
NO
YES
NO

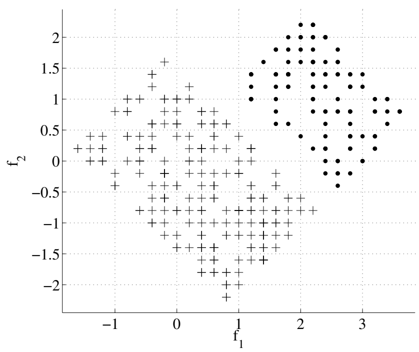
Solution:



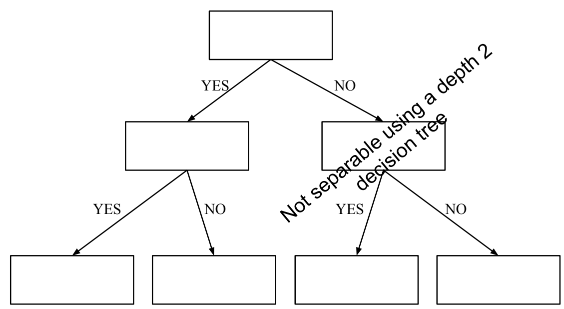
f_2
f_1



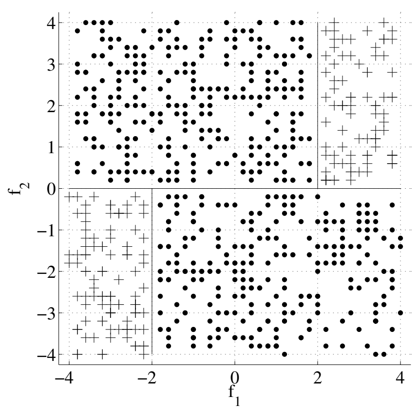
$f_1 < 1$
YES
NO
$f_2 < 1$
Dot
YES
NO
YES
NO
Plus
Dot



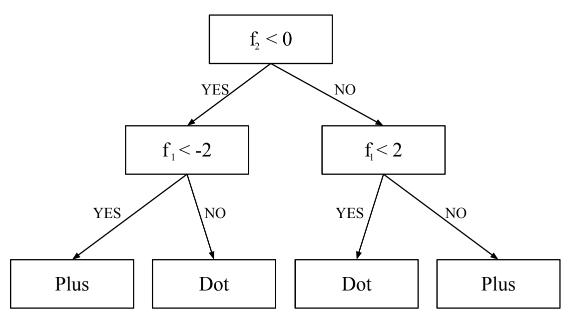
f_2
f_1



YES
NO
Not separable using a depth 2 decision tree
YES
NO
YES
NO



f_2
f_1



$f_2 < 0$
YES
NO
$f_1 < -2$
$f_1 < 2$
YES
NO
YES
NO
Plus
Dot
Dot
Plus

2. A friend asks you to babysit her cat for a few days. As you are inexperienced with animals but very good in machine learning, you build a decision tree to decide automatically whether you should feed the cat or not. Before she leaves you the cat, you spend a day with your friend and observe her to build the following data set. Ones and zeros stand respectively for *yes* and *no*.

Sample	Cat meowing	Cat rubbing against you	Min 2h since last meal	Food bowl empty	Feed the cat
1	0	0	1	0	0
2	0	1	1	0	0
3	1	1	0	1	0
4	1	0	1	0	1
5	1	1	1	0	1
6	0	0	0	1	0
7	1	0	1	1	1

(a) Use the entropy as a classification criterion to construct a minimal decision tree that predicts whether or not you should feed the cat. Show each step of the computation. Draw the final decision tree.

(b) Translate your decision tree into a collection of decision rules in plain language.

Solution:

(a) First we need to choose the attribute for the root node. We select the attribute with the minimum entropy, that is, minimum expected number of bits (EBS) $\Rightarrow$ maximum information gain.

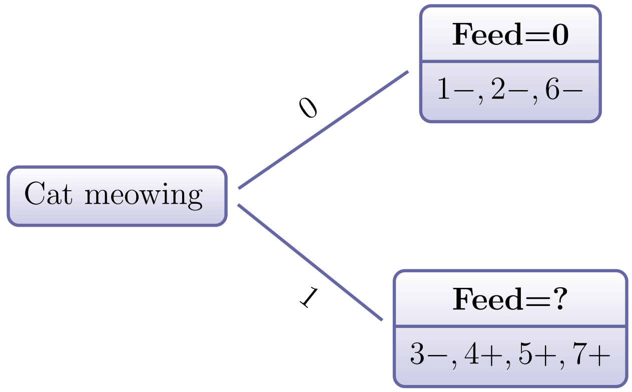
First, we compute the EBS for each attribute, see theory slides (Lec 4 p28). Say an attribute A splits the examples E into subsets E_i, and assume E_i has p_i positive and n_i negative samples. Then,

$$EBS(A) = \sum_i \frac{p_i + n_i}{p + n} H\left(\left\langle \frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i} \right\rangle\right).$$

Remember that $H(\langle 0, 1\rangle) = 0$ and $H(\langle 0.5, 0.5\rangle) = 1$.

- EBS(*Cat meowing*) $= \frac{3}{7}H(\langle \frac{3}{3}, \frac{0}{3}\rangle) + \frac{4}{7}H(\langle \frac{1}{4}, \frac{3}{4}\rangle)$
  $= 0 + \frac{4}{7}(-\frac{1}{4}\log_2(\frac{1}{4}) - \frac{3}{4}\log_2(\frac{3}{4})) \approx 0 + 0.46 = 0.46$
- EBS(*Cat rubbing*) $= \frac{4}{7}H(\langle \frac{2}{4}, \frac{2}{4}\rangle) + \frac{3}{7}H(\langle \frac{1}{3}, \frac{2}{3}\rangle) \approx 0.57 + 0.39 = 0.96$
- EBS(*Min 2h*) $= \frac{2}{7}H(\langle \frac{2}{2}, \frac{0}{2}\rangle) + \frac{5}{7}H(\langle \frac{2}{5}, \frac{3}{5}\rangle) \approx 0 + 0.69 = 0.69$
- For *Food bowl empty*, we have the same numbers as *Cat rubbing*
  EBS(*Food bowl empty*) $= \frac{4}{7}H(\langle \frac{2}{4}, \frac{2}{4}\rangle) + \frac{3}{7}H(\langle \frac{1}{3}, \frac{2}{3}\rangle) \approx 0.57 + 0.39 = 0.96$

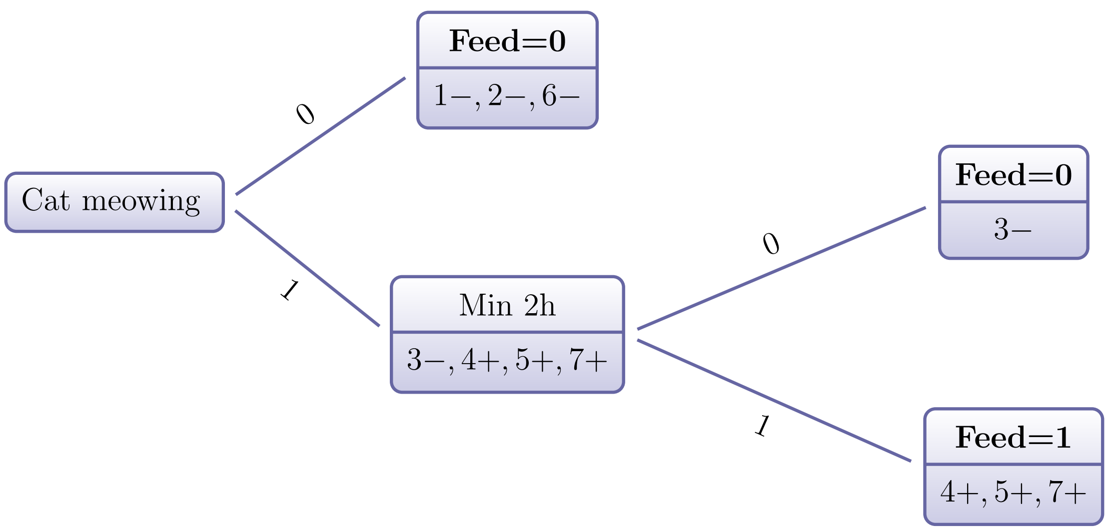
Since *Cat meowing* is the attribute with the lowest entropy, it is selected as the root node. Only the branch corresponding to *Cat meowing = 1* needs further processing, because for *Cat meowing = 0* we have *Feed the cat* always equal to 0.



No we are selecting the attribute to split the branch *Cat meowing = 1*. Only samples 3,4,5,7 need to be taken into consideration. Now we obtain that

- EBS(*Cat rubbing*) $= \frac{2}{4}H(\langle\frac{0}{2},\frac{2}{2}\rangle) + \frac{2}{4}H(\langle\frac{1}{2},\frac{1}{2}\rangle) = 0 + 0.5 = 0.5$
- EBS(*Min 2h*) $= 0 + 0 = 0$

Since the minimum possible entropy is 0 and *Min 2h* has that minimum possible entropy, it is selected as the best attribute to split the branch *Cat meowing = 1*. Note that we don't even have to calculate the entropy of the last attribute since it cannot possibly be better than 0.



(b) Decision rules in plain language:

- If the cat is not meowing, we do not feed him.
- If the cat is meowing and less than 2h have passed since last meal, we do not feed him.
- If the cat is meowing and minimum 2h have passed since last meal, we feed him.