COM S 474 - Homework 3

Stamati Morellas

1. [1pt] Given the dataset (where each sample has 3 feature values: a, b, and c) below, compute the gini impurity for the condition $S1 : a > 10$. Please show the estimations of $Pr(class = +1|a > 10)$ and $Pr(class = -1|a > 10)$. If you do not show these two but a final result, you will get no point.

The samples that pass the test for condition $S1 : a > 10$ are samples S_1 and S_8

For $Pr(class = +1|a > 10) = Pr(class = +1|S_1 = +1, S_8 = -1) = 1/2$

For $Pr(class = -1|a > 10) = Pr(class = -1|S_1 = +1, S_8 = -1) = 1/2$

where I used the notation S_i for sample i

2. [1pt] Do the same for a condition $S2 : a \leq 5$. Again, intermediate steps need to be shown.

The samples that pass the test for condition $S2 : a \leq 5$ are samples S_2 and S_3

For $Pr(class = +1|a \leq 5) = Pr(class = +1|S_2 = +1, S_3 = +1) = 2/2 = 1$

For $Pr(class = -1|a \leq 5) = Pr(class = -1|S_2 = +1, S_3 = +1) = 0/2 = 0$

3. [1pt] Based on the results from the two problems above, compute the expectation for gini impurity for the feature a and the threshold 5. Please show the estimatons of the probabilities of both conditions, i.e., $P(a > 5)$ and $P(a \leq 5)$. If you just show a final result, no point.

For the *left* group, the samples that pass the condition $a \leq 5$ are the following:

$$S_2 = +1, S_3 = +1$$

For the *right* group, the samples that pass the condition $a > 5$ are the following:

$$S_1 = +1, S_4 = +1, S_5 = -1, S_6 = -1, S_7 = -1, S_8 = -1$$

The gini impurity for the left and right groups are as follows:

$$G_{left} = 1 * (1 - 1) + 0 * (1 - 0) = 0$$

$$G_{right} = 2/6 * (1 - 2/6) + 4/6 * (1 - 4/6) = 4/9$$

The expectation for the gini impurity is the following:

$$G_{total} = 2/8 * G_{left} + 6/8 * G_{right} = 2/8 * 0 + 6/8 * 4/9 = 2/12 = 1/6$$

4. [1pt] Using the decision tree below, decide the classification outcomes for all samples in Problem 1. Left branch is True and right branch is False. Present your result as a two-column table.

Sample	Prediction
1	+1
2	+1
3	1
4	1
5	1
6	1
7	1
8	1