CS 4824 / ECE 4424 Homework 1 (Written Portion) Due: Feb 20, 2020

Note: Practice questions will not be graded.

Question 1 [12 points]

Consider the dataset shown in Table 1 for a binary classification problem.

Customer ID	Housing Type	Gender	Marital Status	Class
1	Apartment	Male	Married	C0
2	House	Male	Single	C1
3	House	Female	Married	C1
4	Apartment	Female	Single	C0
5	Apartment	Male	Married	C0
6	Hostel	Male	Single	C1
7	House	Female	Married	C1
8	Apartment	Female	Single	C0
9	Apartment	Male	Married	C0
10	House	Male	Single	C1
11	Hostel	Female	Married	C1
12	Hostel	Female	Single	C0
13	House	Male	Married	C0
14	Hostel	Male	Single	C1
15	Hostel	Female	Married	C1
16	Apartment	Female	Single	C0

Table 1

a. **[1.5 points]** Compute the Gini index, entropy, and misclassification error for the overall data.

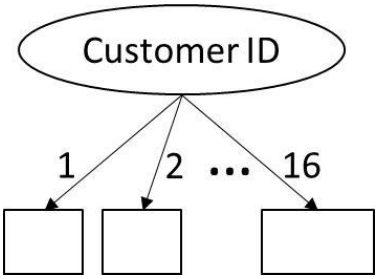
b. **[6 points]** Compute the Gini index, entropy, and misclassification error for each of the four attributes (consider a multi-way split using each unique value of an attribute).

c. **[3 points]** Compute the Information Gain (IG) obtained by splitting the overall data using each of the four attributes. Which attribute provides the highest IG, and which attribute provides the lowest IG.

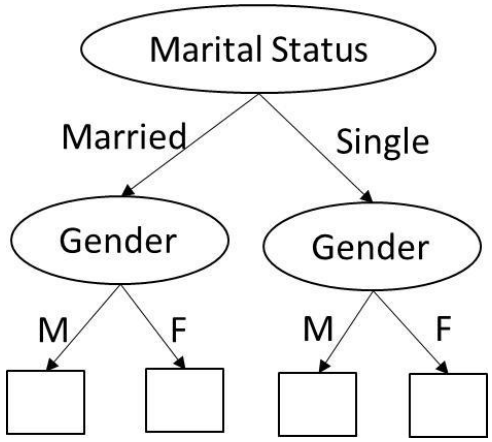
d. **[2.5 points]** Compute the Gain Ratio for splitting over each of the four attributes. Which attribute provides the highest Gain Ratio?

e. **[2 points]** For splitting at the root node, would you choose the attribute that provides the maximum IG, or the attribute that provides maximum Gain Ratio? Briefly explain your choice.

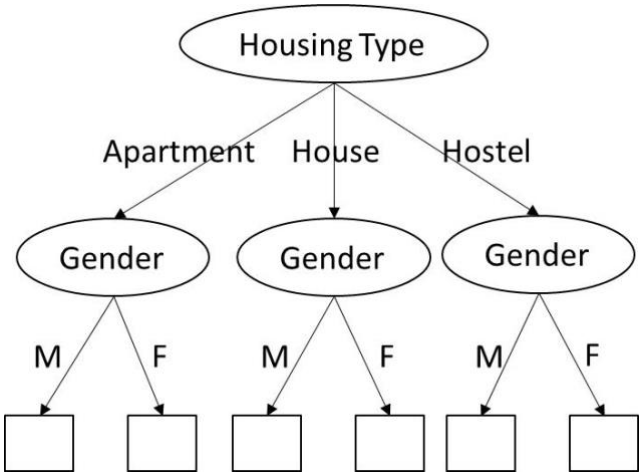
f. **[3 points]** Consider the following 3 decision trees:



Tree 1



Tree 2



Tree3

Compute the difference between the entropy of overall data with the weighted entropy of the leaves for each of the three trees. Based on these differences, which tree would you choose for performing classification? Is the attribute chosen at the root of this tree same as the attribute chosen for splitting in (e)? Briefly comment on the nature of your results, and the properties of the impurity measure used while constructing decision trees.

Question 2 [14 points]

Consider the training data given in Table 2 for classification, where the two classes of interest are '-' and '+'. We want to apply binary decision trees as our chosen algorithm for classifying this data.

	X=1	X=2	X=3	X=4	X=5
Y=5	-	-	-	-	-
Y=4	-	-	-	-	-
Y=3	-	-	-	+	+
Y=2	-	+	-	+	+
Y=1	-	-	-	+	+

Table 2

a. **[6 points]** Find a decision tree which uses minimum number of splits (decision boundaries at internal nodes) to perfectly classify each training data instance of Table 2. Hint: The minimum number of splits that you need to create a perfect classifier is 6. You are ***not*** required to compute the Information Gain at each split for constructing the decision tree, but to arrive at your solution by visually inspecting the data.

b. **[5 points]** Use a pessimistic estimate of the generalization error to prune this tree using sub-tree replacement post-pruning method. Use $\Omega = 2$ as the cost of adding a leaf node while calculating the pessimistic estimate. In case there is a tie in determining the majority class of a leaf node, use '-' as the default majority class.

c. **[3 points]** Compare the nature of the original tree (with perfect classification on training data) with the pruned true by visual inspection. Would you use the original tree or the pruned tree for classifying any future instance of the data? State the name of the phenomena being explored in this question.

Question 3 [16 points]

Consider the dataset shown in Table 3.

Instance	A	B	C	Class
1	0	0	1	-
2	1	0	1	+
3	0	1	0	-
4	1	0	0	-
5	1	0	1	+
6	0	0	1	+
7	1	1	0	-
8	0	0	0	-
9	0	1	0	+
10	1	1	1	+

Table 3

a. **[3 points]** Estimate the conditional probabilities for P(A = 1|+), P(B = 1|+), P(C = 1|+), P(A = 1|−), P(B = 1|−), and P(C = 1|−).

b. **[2 points]** Use the conditional probabilities in part (a) to predict the class label for a test sample (A = 1, B = 1, C = 1) using the naïve Bayes approach.

c. **[2 points]** Compare P(A = 1, B = 1|Class = +) against P(A = 1|Class = +) and P(B = 1|Class = +). Are the variables conditionally independent given the class?

d. **[3 points]** Let us consider the data instance (A=1, B=1, C=1). Compute the probability of this instance belonging to Class = + using

   i. no attributes (i.e. calculate prior probability)

   ii. attribute A [P(Class = +|A=1)]

   iii. attributes A and B [P(Class = +|A=1, B=1)]

   iv. attributes A, B and C [P(Class = +|A=1, B=1, C=1)]

   Comment on the change in probability values as we proceed from (i) to (iv).

Now, consider Table 4

Instance	A	B	C	Class
1	0	0	1	-
2	0	0	1	+
3	0	0	0	-
4	1	0	0	-
5	0	0	1	+
6	0	0	1	+
7	1	0	0	-
8	0	0	0	-
9	0	1	0	+
10	0	1	1	+

Table 4.

e. **[3 points]** Estimate the conditional probabilities for P(A = 1|+), P(B = 1|+), P(C = 1|+), P(A = 1|−), P(B = 1|−), and P(C = 1|−) using Table 4.

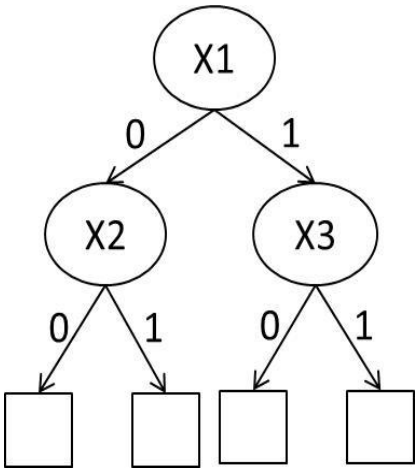
f. **[1 point]** For a new data instance, **x** = (A = 1, B = 1, C = 1), compute the posterior probabilities, P(+|x) and P(-|x) using the Naïve Bayes approach.

g. **[2 points]** What kind of problems will you encounter in predicting the class of **x** using the posterior probabilities computed in (e) and how can you resolve them?

Practice Questions

Question 4

Consider a classification scenario where the data consists of 15 binary attributes, and 10 observations. We are interested in employing complete decision trees of depth 2 (3 internal nodes, 4 leaf nodes) as our chosen algorithm for performing classification. An example of a depth-2 complete decision tree is shown below:



It is also known that each such decision tree has a 0.5 probability of accurately classifying a single data instance in the above scenario.

a. What is the total number of all such possible trees that can be constructed using the given data?

b. What is the probability that such a decision tree will accurately classify all 10 data instances?

c. What is the probability that at least one such decision tree will accurately classify all 10 data instances?

d. Comment on the quality of the classifiers by looking at the result obtained in (c)? State the name of the phenomena being experienced in (c).

Question 5

Consider two decision trees, T1 and T2, where T2 is obtained by pruning some of the subtrees in T1. T1 is the most exhaustive tree that can be created using classification error as the splitting criterion. List all of the following statements that are always true and give a brief explanation.

a. Training error of T1 $\geq$ Training error of T2.

b. Training error of T1 $\leq$ Training error of T2.

c. Test error of T1 ≤ Test error of T2.

d. Test error of T1 ≥ Test error of T2.

Question 6

Consider a data set with four binary attributes X1, X2, X3 and X4. The attribute X4 takes exactly the same value as X3 for each record, i.e., X4 is equal to X3.

i. We build two decision trees: one using all four attributes X1, X2, X3 and X4 and the other using only three attributes X1, X2 and X3.

   a. Do you expect the predictions for a new test instance using the two decision trees to be same or different? Briefly explain.

   b. Do you expect the decision tree built using all 4 attributes to be better or worse than the decision tree built using 3 attributes? Briefly explain.

ii. We build two Naive Bayes models: one using all four attributes X1, X2, X3 and X4 and the other using only three attributes X1, X2 and X3.

   a. Do you expect the predictions for a new test instance using the two Naive Bayes models to be same or different? Briefly explain.

   b. Do you expect the naive Bayes classifier built using all 4 attributes to be better or worse than the naive Bayes classifier built using 3 attributes? Briefly explain.