

Mid-Sem Test

Date: 08.03.2018

Max. Time: 90 minutes

Maximum Marks: 30

Note: Answers to all questions must be analytical, precise and complete.

1. Write short answers for the following :
 - a. What is boosting? State why it may improve the accuracy of decision tree induction.
 - b. Which one is a more expensive scheme to classify a test record and Why? Ordered or unordered rule based scheme.
 - c. What do you mean by M-estimate for Conditional Probability? What is M known as and what it determines?

[3*3=9]

2. Write an algorithm, in *pseudocode* or in your favorite programming language for k-nearest neighbor classification given k and n , the number of attributes describing each sample.

[6]

3. In this problem, use Bayes Rule, $p(y|x) = p(x|y)p(y)/p(x)$ to perform classification. Suppose we observe the following training data, with three binary features $x_1 \dots \dots \dots x_3$ and a binary class y :

x_1	x_2	x_3	y
1	0	0	0
0	1	1	0
0	0	1	0
1	0	0	1
0	0	1	1
0	1	0	1
1	1	0	1

Learn to predict y using a naive Bayes classifier; show your work.

- (a) After learning the model, what is the predicted probability $p(y = 0 | x_1 = 0, x_2 = 1, x_3 = 0)$?
- (b) Suppose that we *only* observe $x_1 = 0$. What is the predicted probability $p(y = 0 | x_1 = 0)$?

[2+2=4]

PTO.....

4. Consider the decision tree shown in Figure 1,

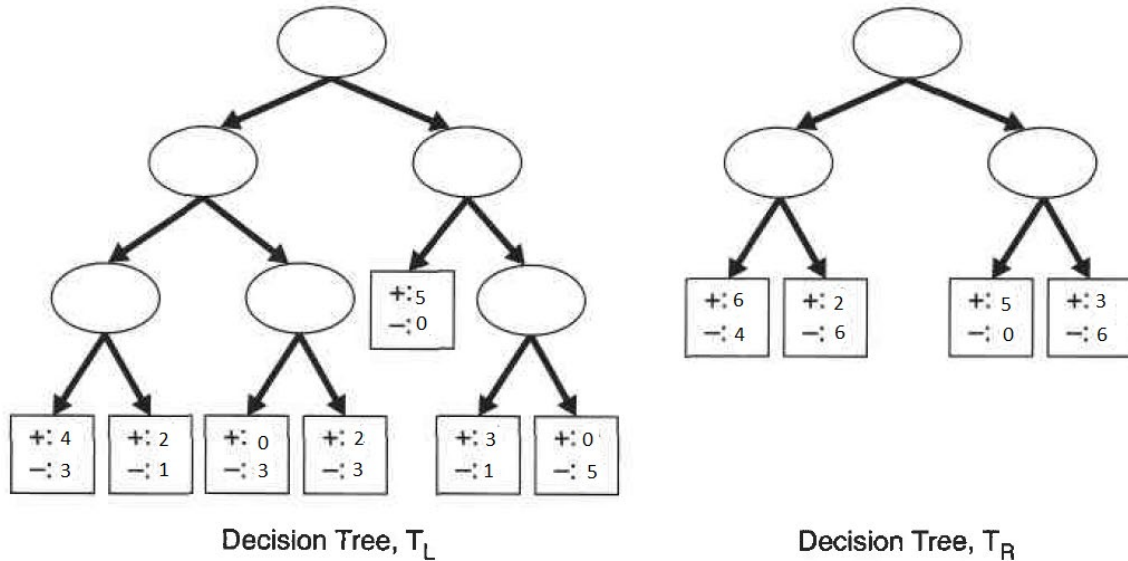


Figure 1

Estimate the generalization error rate of the tree using both the optimistic approach and the pessimistic approach. While computing the error with pessimistic approach, to account for model complexity, use a penalty value of 2 to each leaf node. Identify which tree is better in each case.

[5]

5. For the following given data set, identify the best split attribute based on Information gain. Show all Calculations. Draw the first level decision tree.

<u>Color</u>	<u>Size</u>	<u>Shape</u>	<u>Edible?</u>
Yellow	Small	Round	+
Yellow	Small	Round	-
Green	Small	Irregular	+
Green	Large	Irregular	-
Yellow	Large	Round	+
Yellow	Small	Round	+
Yellow	Small	Round	+
Yellow	Small	Round	+
Green	Small	Round	-
Yellow	Large	Round	-
Yellow	Large	Round	+
Yellow	Large	Round	-
Yellow	Large	Round	-
Yellow	Large	Round	-
Yellow	Small	Irregular	+
Yellow	Large	Irregular	+

[6]
