Name:______________________________________ Roll Number:_______________

Quiz-3

Max points: 20 **Max Time: 20 mins**

Q.1. [6 points]

Mark True (T) or False (F), fill in the blanks, or choose the correct choice for the statements below.

1. If conditional independence assumption does not hold, $\hat{c}_{MAP} = \hat{c}_{NB}$.

(A) True (B) False

2. Consider a dataset that has four attributes and a class label representing one of the three possible classes. If two of the attributes are Boolean and the other two are real-valued, the total number of probability estimates needed for classification of a test sample for a Naïve Bayes classifier would be:

(A) 8 (B) 14 (C) 10 (D) None of the given options

3. Naïve Bayes classifier can be described as a generative model.

(A) True (B) False

4. In decision trees, the more the entropy, the more the information gain with respect to classification.

(A) True (B) False

5. Split information for a binary attribute with equal samples in a dataset is __________.

(A) 2 (B) $\log n$ (C) 0 (D) None of the given options

6. ID-3 is a ____________algorithm.

(A) Iterative (B) Recursive (C) Greedy (D) Both 'B' and 'C'

Q.2. [4+5+5 points]

a) How do you estimate probabilities in the Naïve Bayes classifier using Bayesian approach.

$$P(A=v \mid Class=C_K) = \frac{n_c + mp}{n+m}$$

where; n_c = # of instances where $A=v$ & class $= c_k$

n = # of instances where class $= C_k$

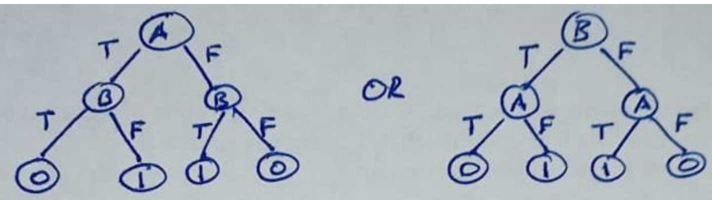
m = equivalent/virtual/hallucinated sample size

p = prior ~~distrib~~ probability distribution of attribute 'A's values

Name:____ Roll Number:__________________

b) Give a decision tree corresponding to the following Boolean function.

$$A \oplus B$$



Note: Both 'A' & 'B' have equal information gain of '0'.

c) Compute the information gain of the attribute a_2 in the following dataset.

Instance	Classification	A_1	A_2
1	+	T	T
2	+	T	T
3	−	T	F
4	+	F	F
5	−	F	T
6	−	F	T

$$Gain(A_2, S) = E(S) - \sum_{v=\{T,F\}} \frac{|S_v|}{|S|} \cdot E(S_v)$$

$$E(S) = -\frac{3}{6}\log_2\frac{3}{6} - \frac{3}{6}\log_2\frac{3}{6} = 1$$

$$E(S_{v=T}) = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = 1$$

$$E(S_{v=F}) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

$$|S_{v=T}| = 4 \quad ; \quad |S_{v=F}| = 2$$

$$\therefore \; Gain(A_2, S) = 1 - \left(\frac{4}{6}\cdot(1) + \frac{2}{6}(1)\right)$$

$$Gain(A_2, S) = 1 - \left(\frac{2}{3} + \frac{1}{3}\right) = 1 - 1 = 0$$