

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -2 EXAMINATION-2022

M.Tech.-I Semester (CS/IT/DS)

COURSE CODE (CREDITS): 22MIWCI133 (3)

MAX. MARKS: 25

COURSE NAME: Introduction to Statistical Learning

COURSE INSTRUCTORS: Dr. Nancy Singla

MAX. TIME: 1 Hour 30 Min

*Note: All questions are compulsory. Marks are indicated against each question in square brackets.*

Q1.	<p>Suppose you are using the Decision Tree Learning algorithm to learn a 2-class classification variable, <math>C</math>, and you must decide which attribute to assign to a node in the tree. At this node there are 100 examples; 30 are positive and 70 are negative. If attribute <math>A</math> is selected, its first child will get 18 positive and 22 negative examples, and its second child will get 12 positive and 48 negative examples.</p> <p>Use <math>\log 0.1 = -3.32</math>, <math>\log 0.2 = -2.32</math>, <math>\log 0.3 = -1.74</math>, <math>\log 0.33 = -1.59</math>, <math>\log 0.4 = -1.32</math>, <math>\log 0.45 = -1.15</math>, <math>\log 0.5 = -1.0</math>, <math>\log 0.55 = -0.86</math>, <math>\log 0.6 = -0.74</math>, <math>\log 0.67 = -0.58</math>, <math>\log 0.7 = -0.51</math>, and <math>\log 0.8 = -0.32</math>, <math>\log 0.9 = -0.15</math>, and <math>\log 1 = 0</math>, where all logs are to base 2.</p> <p>(a) What is the entropy of <math>C</math>, i.e., <math>H(C)</math>, at the node?            (b) What is the conditional entropy of choosing attribute <math>A</math> at the node? That is, compute <math>H(C   A)</math>.            (c) How overfitting in decision trees can be reduced?</p>	[2+2+2] CO3
Q2.	<p>(a) What is dimensionality reduction? How PCA is different from LDA?            (b) What are the steps involved in PCA algorithm? Using PCA, determine the Eigen values for the given data <math>\{2, 3, 4, 5, 6, 7 ; 1, 5, 3, 6, 7, 8\}</math>.</p>	[3+6] CO2, CO4
Q3.	Suppose you found that your model is suffering from low bias and high variance. Which algorithm you think could tackle this situation and Why?	[3] CO1, CO3
Q4.	<p>Best subset, forward stepwise and backward stepwise selection is performed on a single data set. For each approach, we obtain <math>p + 1</math> models, containing 0, 1, 2, .. <math>p</math> predictors. Explain your answers:</p> <p>(a) Which of the three models with <math>k</math> predictors has the smallest <i>training</i> RSS?            (b) Which of the three models with <math>k</math> predictors has the smallest <i>test</i> RSS?</p>	[2+2] CO2
Q5.	<p>(a) Explain how boosting is performed with the help of a suitable example.            (b) Why do we want to use "weak" learners when boosting?</p>	[2+1] CO3