JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION - February 2022

B.Tech-IV Semester (CSE&IT)

COURSE CODE: 119B1WCI740   MAX. MARKS: 15

COURSE NAME: Introduction to Statistical Learning

COURSE INSTRUCTORS: Dr. Hari Singh   MAX. TIME: 1 Hour

Note: All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means. Marks are indicated against each question in square brackets.

Q1. Provide a sketch of typical (squared) bias, variance, training error, test error, and Bayes (or irreducible) error curves, in a single plot, as we go from less flexible statistical learning methods towards more flexible approaches. The x-axis should represent the amount of flexibility in the method, and the y-axis should represent the values for each curves. There should be five curves. Make sure to label each one.   CO1 [3]

Q2. For a certain plat come from three different cities A, B and C where 30% are from A, 25% are from B and 45% are from C. Out of those 2% from A, 1% from B and 3% from C are found COVID positive. Find the probability that a worker chosen at random would be tested COVID positive and also find the probability that if a worker is tested COVID positive, what is the probability that he is coming from A or B or C?   CO1 [3]

Q3. Describe the following terms in detail along with appropriate figures: CO2 [3x2=6]
   (a) Bias-Variance Trade-Off
   (b) Heteroscedasticity

Q4. For a set of data of 100 observations containing a single predictor and a quantitative response, a linear regression and a cubic regression model is fitted. Suppose that the true relationship between X and Y is not linear, but we don't know how far it is from linear. Consider the training RSS for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.   CO2 [3]