JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -2 EXAMINATION - October 2022

B.Tech-IV Semester (CSE&IT)

COURSE CODE: 119B1WCI740 MAX. MARKS: 25

COURSE NAME: Introduction to Statistical Learning

COURSE INSTRUCTORS: Dr. Hari Singh MAX. TIME: 1.5 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. (a) Say buyers pay more for action movies than for other movies genre. A movie selling enterprise would be very much interested in identifying action movies correctly to promote to its buyers and make more revenues. Out of the various confusion metrics, accuracy, precision, and recall; which is more important here?

(b) Which one is more important among Accuracy, Precision, Recall, and F1-Score in music or video recommendation system and in medical cases such as identifying infected persons? CO2 [2x2=4]

Q2 (a) Suppose that an individual has a 16 % chance of defaulting on her credit card payment. What are the odds that she will default?

(b) If the Bayes decision boundary is linear, do we expect LDA or QDA to perform better on the training set? On the test set? CO2 [2x3=6]

Q3. Provide a sketch of typical (squared) bias, variance, training error, test error curves, as we go from validation approach to LOOCV approach to Cross-Validation approach. Make sure to label each one. Also explain each curve in brief. CO2 [6]

Q4. We perform best subset, forward stepwise, and backward stepwise selection on a single data set. For each approach, we obtain p + 1 models, containing 0, 1, 2,...,p predictors. CO2 [2x2=4]
Explain your answers:
(a) Which of the three models with k predictors has the smallest training RSS?
(b) Which of the three models with k predictors has the smallest test RSS?

Q5. Suppose we estimate the regression coefficients in a linear regression model by minimizing

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2 \quad \text{subject to} \quad \sum_{j=1}^{p}|\beta_j|\le s$$

for a particular value of s. For parts (a) through (e), indicate which of i. through v. is correct. Justify your answer. CO2 [5]

(a) As we increase s from 0, the training RSS will:
i. Increase initially, and then eventually start decreasing in an inverted U shape.
ii. Decrease initially, and then eventually start increasing in a U shape.
iii. Steadily increase.
iv. Steadily decrease.
v. Remain constant.

(b) Repeat (a) for test RSS.
(c) Repeat (a) for variance.
(d) Repeat (a) for (squared) bias.
(e) Repeat (a) for the irreducible error.