

(2 ½ Hours)

[Total Marks: 60]

- N.B:**
- (1) All questions are compulsory.
  - (2) Figures to the **right** indicate full marks.
  - (3) **Assume additional data if necessary** but state the same clearly.
  - (4) Symbols have their usual meanings and tables have their usual standard design unless stated otherwise.

- Q.1** Attempt any two of the following (12)
- a) Explain the structure of learning. 6
  - b) What are the characteristics of machine learning tasks? Explain with a suitable example. 6
  - c) List and explain any 2 machine learning models. 6
  - d) Write a short note on feature selection. 6
- Q.2** Attempt any two of the following (12)
- a) Write a short note on PAC learnability. 6
  - b) Explain the term inductive bias. 6
  - c) Write a short note on generalized loss functions. 6
  - d) Define empirical risk minimization. 6
- Q.3** Attempt any two of the following (12)
- a) Compare linear regression and logistic regression. 6
  - b) Write a short note on boosting. 6
  - c) Describe the model selection and validation techniques. 6
  - d) Define with a suitable example “convex set”. 6
- Q.4** Attempt any two of the following (12)
- a) Describe the term VC dimension with a suitable example. 6
  - b) Explain the term generalization upper bound. 6
  - c) Given a sample space of  $N = 100$ , where the input space is contained inside the unit ball. Calculate the generalization error (GE) with 99% confidence level for a separating hyperplane with a margin of 0.2.  
(Note: Write the expression for GE by substituting the values of variables/parameters. Calculation of the accurate value is not needed.) 6
  - d) Explain Natarajan dimension. 6

- Q.5** Attempt any two of the following **(12)**
- a) Describe the geometric interpretation of normal distribution. **6**
- b) Define multinomial distribution. Calculate the multinomial distribution for a sequence of words: a b b b c a c a b. **6**
- c) Explain how multi-task learning improves the generalizability of a model. **6**
- d) Emails are represented by bit vectors in a multivariate Bernoulli model, as given in the following table. Calculate the estimated vectors for the classes + and - . **6**

Email	a?	b?	c?	Class
E1	0	1	0	+
E2	0	1	1	+
E3	1	0	0	+
E4	1	1	0	+
E5	1	1	0	-
E6	1	0	1	-
E7	1	0	0	-
E8	0	0	0	-

(2 ½ Hours)

[Total Marks: 60]

- N.B:**
- (1) **All questions are compulsory.**
  - (2) Figures to the **right** indicate full marks.
  - (3) **Assume additional data if necessary** but state the same clearly.
  - (4) Symbols have their usual meanings and tables have their usual standard design unless stated otherwise.
  - (5) Use of **calculators** and statistical tables are **allowed**.

- Q.1 Attempt **any two** of the following (12)
- a) Describe Confidence interval. 6
  - b) Explain the method of least squares and P-test. 6
  - c) Describe multiple regressions, and mention the assumptions of multiple regressions and explain them. 6
  - d) Define Bivariate, Trivariate, Outliner. 6
- Q.2 Attempt **any two** of the following (12)
- a) Explain fitting curvilinear relationship, when should a curvilinear relationship be used. 6
  - b) Mention why a linear transformation is written in matrix form. 6  
Translate the following to matrix  
 $T(x_1, x_2, x_3) = (2x_1 + x_2 - x_3, -x_1 + 3x_2 - 2x_3, 3x_2 + 4x_3)$
  - c) Explain Autocorrelation and the types, Describe any one type of autocorrelation 6
  - d) Justify transformation to linearize the model. 6
- Q.3 Attempt **any two** of the following (12)
- a) Define multicollinearity, with examples, equation, and explain why multicollinearity is a problem. 6
  - b) The eigen values of acetylene data are 6  
 $\lambda_1 = 4.2048, \lambda_2 = 2.1626, \lambda_3 = 1.1384, \lambda_4 = 1.0413, \lambda_5 = 0.3845,$   
 $\lambda_6 = 0.0495, \lambda_7 = 0.0136, \lambda_8 = 0.0051$  and  $\lambda_9 = 0.0001$   
Calculate the condition indices and mention which indicates severe multicollinearity.
  - c) Explain how to deal with multicollinearity. 6

- d) Define the terms 6
- 1) High multicollinearity.
  - 2) Perfect multicollinearity.
  - 3) Data based multicollinearity.

Q.4 Attempt **any two** of the following (12)

- a) Explain the Analysis of variance (ANOVA) test. 6
- b) Write a note on F-test and T-test 6
- c) Explain the three components of GLM 6
- d) Explain the modeling probability for binomial and continuous variable. 6

Q.5 Attempt **any two** of the following (12)

- a) Explain model misspecification. 6
- b) What is the need to collect fresh data for models. 6
- c) Write a note on Backward Elimination for variable selection. 6
- d) Explain Linearization. 6

\*\*\*\*\*