No. of Printed Pages : 4

MST–026

M. SC. (APPLIED STATISTICS)

(MSCAST)

Term-End Examination

June, 2026

MST-026 : INTRODUCTION TO MACHINE LEARNING

Time : 3 Hours *Maximum Marks : 50*

Note : *(i) Question No. **1** is compulsory.*

*(ii) Attempt any **four** questions from question numbers 2 to 6.*

(iii) Use of scientific (non-programmable) calculator in allowed.

(iv) Symbols have their usual meanings.

1. (a) In an unsupervised learning task, what is the term used for grouping similar data points together ?

   (b) Bias-variance tradeoff is a concept related to …… .

(c) A small value of k in k-NN regression can lead to …….. .

(d) Write any *four* benefits of neural networks. 2+2+2+4

2. Explain in detail single layer perception algorithm and its convergence. 10

3. For the dataset defined as follows :

```
dat < – function (n = 1000, p = 0.5, mu_1 = 0,
      mu_2 = 2, sigma_0 = 1, sigma_1 = 1)

      y < – rbinom (n, 1, p)

      f_0 < – rnorm (n, mu_0, sigma_0)

      f_1 < – rnorm (n, mu_1, sigma_1)

      x < – if else (y = = 1, f_1, f_0)
```

(a) Write the code to create an even split of the data into train and test partitions. 2

(b) For this dataset write R code to obtain accuracy of linear and logistic regression. 8

4. Consider the built in dataset tissue_gene_expression.

This dataset includes a matrix x with the gene expression measured on 500 genes for 189 biological samples representing

seven different tissues. The tissue type is stored in y.

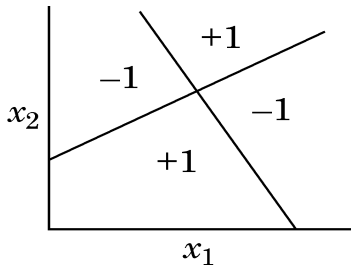
(a) Write R code to split this dataset in training and test sets in the ratio 80 : 20. 2

(b) Write R code to predict tissue type using kNN algorithm. Also, write R code to obtain accuracy. 8

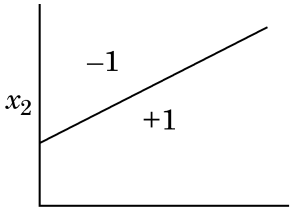
5. Explain the procedure of Naïve Bayes algorithm of machine learning. 10

6. (a) In a binary classification problem, out of 64 data points 29 belongs to Class I and 35 belongs to class II. What is the entropy of the dataset ? 5

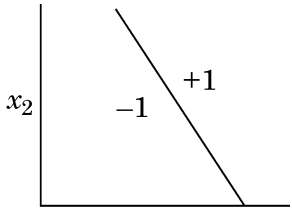
(b) A neural network can be considered as multiple simple equations stacked together. Suppose we want to replicate the function for the below mentioned decision boundary.



Using two simple inputs h_1 *and* h_2.



h_1



h_2

What will be the final equation ? 5

× × × × ×