Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY

(A constituent unit of MAHE, Manipal 576104)

I SEM M. Tech (MI) DEGREE END SEMESTER EXAMINATIONS, NOV/DEC-2023

SUBJECT: MACHINE LEARNING (BME 5116)

(REVISED CREDIT SYSTEM)

Thursday, 30th November, 2023; 9.30 AM- 12.30 PM

TIME: 3 HOURS **MAX. MARKS: 50**

Instructions to Candidates:

1. **Answer ALL questions**
2. **Draw diagrams wherever necessary**
3. **Missing data may be suitably assumed**

1A. How the linear regression method is different from logistic regression? Identify a suitable measure to estimate error during regression analysis and Justify. (3)

1B. How Naive Bayesian Theorem can be applied for classification problem? Determine the conditional probability **P(YES/Body_temperature)** from the given details of subjects in table 1(B) for various infection condition. (3)

Table 1 (B) Details of Patient symptoms

Body_temperature (Higher than 100.4 degree)	*Shortness of breath*	*Muscle pain*	*Loss of taste*	*Infection Present*
Yes	No	Yes	Yes	**YES**
Yes	No	Yes	No	**NO**
Yes	Yes	No	Yes	**YES**
No	No	Yes	Yes	**NO**
Yes	No	No	No	**NO**
Yes	No	Yes	Yes	**YES**
No	Yes	Yes	No	**NO**
No	No	Yes	No	**NO**
Yes	Yes	No	Yes	**YES**
Yes	Yes	Yes	Yes	**YES**

1C. Estimate the value of **p(X/A)** using matrix method; where X= [15 35]^T ; from the normally distributed classes A and B. The class features x and y are normally distributed and details are as follows: (4)

Class A: P(A)=0.6, μx=9, σx=2, μy=25, σy= 3, ρxy = 0.5

Class B: P(B)=0.4, μx = 18, σx =2, μy = 45, σy= 4, ρxy = 0.5

Visualize the **Class A** and interpret it.

2A. Develop an algorithm for solving classification problems with the concept of decision tree utilizing entropy/ information gain. (3)

2B. Determine the information *gain* utilizing the details from table 2B, consisting of symptoms of women suspected with ***Gestational Diabetes*** and identify the root node attribute for building the decision tree. (3)

Reg. No.										

Table 2B. Details of subjects

Frequent Urination	***Blurred Vision***	***Weakness***	***Gestational Diabetes***
Yes	Yes	Yes	**Yes**
Yes	No	Yes	**No**
Yes	Yes	No	**Yes**
No	No	Yes	**No**
Yes	No	No	**No**
Yes	Yes	No	**Yes**
No	Yes	Yes	**No**
Yes	Yes	Yes	**Yes**

2C. Develop an algorithm to describe workflow of a perception, and illustrate the usage of weight updating rules for training the network. (4)

3A Identify the challenges associated in the design of image-based classification system for pathology application. Illustrate this by considering a case study. (4)

3B. Explain identification of a support vectors in a given sample space. Illustrate the importance of them for building a machine learning model. (3)

3C. Discuss the principal components and its usage in the development of ML models. (3)

4A. Identify one of top-down clustering methods, and apply it on the following 5 subjects to generate 3 clusters: (10, 14), (11,19), (12,18), (27,29), (30,32). Finally find centroids of generated clusters. (4)

4B. Illustrate the stacking concept with an example, to improve the overall performance of a machine learning model. (3)

4C Design a workflow for analyzing the machine learning model for finding the abnormality with cardiac condition utilizing the ECG and heart rate variation. (3)

5A. Compare different machine learning methods. Justify that the explainability of a model are important today. (3)

5B Design a perceptron network to classify the two-dimensional input patterns "**C"** and "T" . The symbol "*" indicates the data representation to be " +1 " and " • " indicate data to be " -1 ". Consider a target of "+1 " for pattern **C** and "-1" for pattern **T**. (3)

```
      * * *                 * * *
C =   *   -   -       T =   - * -
      * * *                 - * -
```

Reg. No.										

5C. Analyze significance of accuracy, sensitivity, and specificity as performance parameters. (4)
Calculate the accuracy, sensitivity and specificity from the confusion matrix given:

Table 5 (B) Confusion matrix

		Predicted classes by ML model	
		Positive	**Negative**
Actual Class	**Positive**	1250	40
	Negative	60	1150