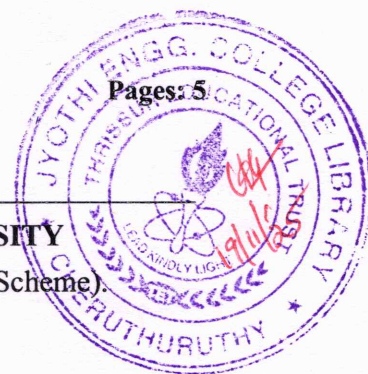


Reg No.: _____

Name: _____

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY
B.Tech Degree S7 (R,S) Examination November 2025 (2019 Scheme)

**Course Code: CST413****Course Name: MACHINE LEARNING****Max. Marks: 100****Duration: 3 Hours****PART A***Answer all questions, each carries 3 marks.***Marks**

- 1 A coin is tossed 100 times and lands heads 62 times. What is the maximum likelihood estimate for θ , the probability of heads. (3)
 - 2 Compute the ML estimate for the parameter λ in the Poisson distribution whose probability function is: $f(x) = e^{-\lambda} \lambda^x / x!$, $x=0,1,2,\dots$ (3)
 - 3 Describe the over fitting problem in machine learning and Identify some reason for this problem? (3)
 - 4 Obtain a linear regression for the data in table , assuming that y is the dependent variable (3)
- | | | | | | |
|---|------|------|------|------|------|
| x | 1.0 | 2.0 | 3.0 | 4.0 | 5.0 |
| y | 1.00 | 2.00 | 1.30 | 3.75 | 2.25 |
- 5 Define support vectors and list any two properties of the support vector (3)
 - 6 Suppose that you have a linear support vector machine (SVM) binary classifier. Consider a point that is currently classified correctly, and is far away from the decision boundary. If you remove the point from the training set, and re-train the classifier, will the decision boundary change or stay the same? Justify your answer. (3)
 - 7 Illustrate the strength and weakness of k-means algorithm. (3)
 - 8 Explain the different types of linkages in clustering. (3)
 - 9 Justify how bias and variance trade-off affect the machine learning algorithms? (3)
 - 10 Suppose there are three classifiers A, B and C. The (FPR, TPR) measures of the three classifiers are as follows – A (0, 1), B (1, 1), C (1, 0.5). Identify a perfect classifier for this problem and Justify your answer. (3)

PART B

Answer any one full question from each module, each carries 14 marks.

Module I

- 11 a) Suppose that X is a discrete random variable with the following probability mass function: where $0 \leq \theta \leq 1$ is a parameter. The following 10 independent observations were taken from such a distribution: (3, 0, 2, 1, 3, 2, 1, 0, 2, 1). What is the MLE of θ . (10)

X	0	1	2	3
$P(x)$	$2\theta/3$	$\theta/3$	$2(1-\theta)/3$	$(1-\theta)/3$

- b) Distinguish between classification and regression with an example (4)

OR

- 12 a) Suppose you have a three class problem where class label $y \in \{0, 1, 2\}$ and each training example X has 3 binary attributes $X_1, X_2, X_3 \in \{0, 1\}$. How many parameters (probability distribution) do you need to know to classify an example using the Naïve Bayes classifier? (7)

- b) Consider the geometric distribution, which has p.m.f $P(X=k) = (1-\theta)^{k-1}\theta$ (7)

Assume that an i.i.d data are drawn from that distribution.

- Write an expression for the log-likelihood of the data as a function of the parameter θ .
- Find the MLE of θ .
- Let θ has a beta prior distribution. What is posterior distribution of θ .

Module II

- 13 a) For a sunburn dataset given below, Find first splitting attribute for decision tree by using ID3 algorithm (8)

Name	Hair	Height	Weight	Lotion	Class
Sarah	Blonde	Average	Light	No	Sunburn
Dana	Blonde	Tall	Average	Yes	none
Alex	Brown	Tall	Average	Yes	None
Annie	Blonde	Short	Average	No	Sunburn
Emily	Red	Average	Heavy	No	Sunburn
Pete	Brown	Tall	Heavy	No	None
John	Brown	Average	Heavy	No	None
Katie	Blonde	Short	Light	Yes	None

- b) Assume we have a classification problem involving 3 classes: professors, students, and staff members. There are 750 students, 150 staff members and 100 professors. All professors have blond hair, 50 staff members have blond hair, and 250 students have blond hair. Compute the information gain of the test "hair color = blond" that returns true or false. (6)

OR

- 14 a) Consider the training data in the following table where Play is a class attribute. (7)
In the table, the Humidity attribute has values "L" (for low) or "H" (for high), Sunny has values "Y" (for yes) or "N" (for no), Wind has values "S" (for strong) or "W" (for weak), and Play has values "Yes" or "No".

Humidity	Sunny	Wind	Play
L	N	S	No
H	N	W	Yes
H	Y	S	Yes
H	N	W	Yes
L	Y	S	No

What is class label for the following day (Humidity=L, Sunny=N, Wind=W), according to naïve Bayesian classification?

- b) Consider the hypothesis for the linear regression $h_0(x) = \Theta_0 + \Theta_1 x$, and the cost function $J(\Theta_0, \Theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_0(x^{(i)}) - y^{(i)})^2$ where m is the number of training examples. Given the following set of training examples. (7)

x	y
3	2
1	2
0	1
4	3

Answer the following questions :

- 1) Find the value of $h_0(2)$ if $\Theta_0 = 0$ and $\Theta_1 = 1.5$
- 2) Find the value of $J(0,1)$
- 3) Suppose the value of $J(\Theta_0, \Theta_1) = 0$. What can be inferred from this?

Module III

- 15 a) State the mathematical formulation of the SVM problem. Give an outline of the method for solving the problem (8)
- b) What is the basic idea of back propagation algorithm (6)

OR

- 16 a) Consider a neuron with four inputs, and weight of edge connecting the inputs are 1, 2, 3 and 4. Let the bias of the node is zero and inputs are 2, 3, 1, and 4. If the activation function is linear $f(x) = 2x$, compute the output of the neuron. (4)
- b) Illustrate the various steps in the back propagation algorithm, for a small network with two inputs, two outputs and one hidden layer (10)

Sample	Input1	Input2	Output1	Output2
1	0.05	0.10	0.01	0.99
2	0.25	0.18	0.23	0.79

Module IV

- 17 a) Show the final result of hierarchical clustering with complete linkage by drawing a dendrogram. (10)

	A	B	C	D	E	F
A	0					
B	0.12	0				
C	0.51	0.25	0			
D	0.84	0.16	0.14	0		
E	0.28	0.77	0.70	0.45	0	
F	0.34	0.61	0.93	0.20	0.67	0

- b) Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8): (4)
- (i) Compute the Euclidean distance between the two objects.
- (ii) Compute the Manhattan distance between the two objects.

OR

- 18 a) Suppose that we have the following data (one variable). Use single linkage Agglomerative clustering to identify the clusters. (8)
- Data: (2, 5, 9, 15, 16, 18, 25, 33, 33, 45).
- b) Illustrate the idea of PCA for a two dimensional data using suitable diagrams (6)

Module V

- 19 a) Assume you have a model with a high bias and a low variance. Interpret the (6)

characteristics of such a model?

- b) State ROC space and ROC curve in machine learning? In ROC space, which points correspond to perfect prediction, always positive prediction and always negative prediction? Why? (8)

OR

- 20 a) Suppose the dataset had 9700 cancer-free images from 10000 images from cancer patients. Find precision, recall and accuracy? Is it a good classifier? Justify. (6)

Actual Class Predicted class	cancer = yes	cancer = no	Total
cancer = yes	90	210	300
cancer = no	140	9560	9700
Total	230	9770	10000

- b) Suppose 10000 patients get tested for flu; out of them, 9000 are actually healthy and 1000 are actually sick. For the sick people, a test was positive for 620 and negative for 380. For the healthy people, the same test was positive for 180 and negative for 8820. Construct a confusion matrix for the data and compute the accuracy, precision and recall for the data. Also construct the ROC curve. (8)
