MODEL QUESTION PAPER

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

THIRD SEMESTER B. TECH DEGREE EXAMINATION, NOVEMBER/DECEMBER 2025

Course Code: GNEST305

Course Name: INTRODUCTION TO ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

Max. Marks: 60 | Duration: 2 hours 30 minutes

PART A

		Answer all questions. Each question carries 3 marks	CO	Marks
1		List the types of machine learning systems with one example each.	CO1	(3)
2		What is the difference between classification and regression?	CO1	(3)
3		What is the importance of linear algebra in data representation?	CO2	(3)
4		Define Singular Value Decomposition (SVD) with its component matrix and its relevance with reference to the rows and columns of the data matrix.	CO2	(3)
5		Define a random variable with an example.	CO3	(3)
6		Differentiate between correlation and regression in terms of their applicability in statistical modelling.	CO3	(3)
7		What are the main benefits of data science in modern industries?	CO4	(3)
8		What is meant by Big Data, and how is it related to data science?	CO4	(3)

PART B

Answer any one full question from each module. Each question carries 9 marks

Module 1

9	a)	In binary classification, a model outputs a probability of p = 0.8. If the decision boundary is 0.5, state the predicted class. Justify your reasoning.	CO1	(3)
	b)	Apply one step of K-means clustering to the dataset {2, 4, 8, 10} with k = 2 and initial centroids 2 and 8. Show new clusters.	CO1	(3)
	c)	A dataset has 100 samples. A classification model misclassifies 12 samples. Compute its accuracy.	CO1	(3)
10	a)	For input x = [1, 1], weights w = [0.5, 0.5], and bias b = 0, compute the perceptron output before applying activation.	CO1	(3)
	b)	Explain with an example how a Multi-Layer Perceptron (MLP) can solve a problem that a single perceptron cannot.	CO1	(3)
	c)	Logistic regression is used to predict whether a patient has a disease (Yes/No). Explain how the probability output can be used for decision-making.	CO1	(3)

Module 2

11	a)	Find the eigenvalues of the matrix $A=\begin{pmatrix}2 & 0\\ 0 & 3\end{pmatrix}$.	CO2	(3)
	b)	Apply spectral decomposition to the symmetric matrix $A=\begin{pmatrix}1 & 2\\ 2 & 4\end{pmatrix}$.	CO2	(3)
	c)	For the covariance matrix $A=\begin{pmatrix}4 & 2\\ 2 & 3\end{pmatrix}$, identify whether the eigenvectors will be orthogonal or not.	CO2	(3)
12	a)	A dataset has 100 features. After applying PCA, only 10 features are retained. What does this imply about dimensionality reduction and why might it be useful in Machine Learning?	CO2	(3)
	b)	A noisy image matrix is approximated using only the first two largest singular values in its SVD. Explain computationally why this reduces storage and how it helps in image compression.	CO2	(3)
	c)	For the matrix $A=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$, find $A^T A$ and use it to compute the singular values of A.	CO2	(3)
		Module 3		
13	a)	A coin is tossed 3 times. List the possible outcomes of the experiment. Also write the probability distribution for the event of getting at least two heads.	CO3	(3)
	b)	A bag contains 3 red and 2 blue balls. One ball is drawn at random. Find the probability that it is (i) red, (ii) blue.	CO3	(3)
	c)	A factory has 3 machines A, B, C producing 30%, 50%, and 20% of items, respectively. Their defect rates are 2%, 3%, and 4%. If a randomly chosen item is defective, find the probability it was produced by machine B.	CO3	(3)
14	a)	The marks of 5 students are: 50, 60, 70, 80, 90. Compute the mean and variance. Use these parameters to interpret the data distribution.	CO3	(3)
	b)	The data for X: {1, 2, 3, 4, 5}, Y: {2, 4, 6, 8, 10}. Compute the correlation coefficient. What does the result indicate?	CO3	(3)
	c)	A sample of size n is drawn from a normal distribution $N(\mu,\sigma)$ with known variance. Show that the sample mean is the MLE of μ.	CO3	(3)
		Module 4		
15	a)	A bank wants to predict whether a loan applicant will default (Yes/No). The dataset contains: Age, Income, Loan Amount, Previous Default (Yes/No). (a) Write the steps to build a classification model using machine learning. (b) Suggest which algorithm (Logistic Regression / Decision Tree) could be used and why.	CO4	(5)
	b)	Discuss any two applications of machine learning in data science.	CO4	(4)
16	a)	A university wants to analyze students' performance data - attendance, assignment scores, exam scores. Describe how the data science process	CO4	(5)

		would be applied here.		
	b)	A bank receives the following types of data daily: 500,000 transaction records (numerical), 50,000 customer feedback forms (text), ATM surveillance videos (image/video). (a) Identify the Big Data characteristics present. (b) Briefly explain why traditional methods fail and how data science techniques overcome this.	CO4	(4)
***				