PAPER ID-311543

Roll No: | | | | | | | | | | | | | |

BTECH
(SEM V) THEORY EXAMINATION 2025-26
MATHEMATICAL FOUNDATION AI, ML AND DATA SCIENCE

TIME: 3 HRS **M.MARKS: 70**

Note: Attempt all Sections. In case of any missing data; choose suitably.

SECTION A

1. Attempt *all* questions in brief. **02 x 7 = 14**

Q no.	Question	CO	Level
a.	Define the expectation $\boldsymbol{E(X)}$ of a random variable and state one of its properties.	CO1	K2
b.	What is skewness? Why is it important in data analysis?	CO1	K2
c.	Define hypothesis testing.	CO3	K2
d.	What is resampling in statistics?	CO2	K2
e.	Define a vector space.	CO4	K2
f.	What is Principal Component Analysis (PCA)?	CO4	K2
g.	State the Markov property.	CO2	K2

SECTION B

2. Attempt any *three* of the following: **07 x 3 = 21**

a.	The following data represents the heights (cm) and weights (kg) of **10 individuals**: Heights: 150, 155, 160, 162, 165, 168, 170, 172, 175, 178 Weights: 52, 55, 58, 60, 62, 64, 66, 68, 70, 72 (i) Compute the **Pearson correlation coefficient** $(\boldsymbol{r})$ using $\boldsymbol{r = \Sigma[(x - \bar{x})(y - \bar{y})] / \sqrt{[\Sigma(x - \bar{x})^2 \Sigma(y - \bar{y})^2]}}$ (ii) Test the significance of correlation at $\boldsymbol{\alpha = 0.05}$. (iii) Interpret the result for predictive modeling.	CO1	K4
b.	A discrete random variable $\boldsymbol{X}$ has the following probability distribution: X: 0, 1, 2, 3 $P(X)$: 0.1, 0.3, 0.4, 0.2 (i) Verify that the given distribution is valid. (ii) Find the expected value $\boldsymbol{E(X)}$ and variance $\boldsymbol{Var(X)}$ using $\boldsymbol{E(X) = \Sigma x \cdot P(X)}$ **and** $\boldsymbol{Var(X) = E(X^2) - [E(X)]^2}$. (iii) Verify Chebyshev's inequality for $\boldsymbol{k = 2}$ and comment on its significance when the underlying distribution is unknown.	CO1	K4
c.	A random sample of $\boldsymbol{n = 12}$ observations have sample mean $\boldsymbol{\bar{x} = 68}$ and standard deviation **s = 4.5**. (i) Construct a **95% confidence interval** for the population mean using $\boldsymbol{\bar{x} \pm t_{0.025,11} (s / \sqrt{n})}$. (ii) Explain how confidence level affects the width of the interval.	CO3	K5
d.	The following contingency table shows the classification outcome of a machine learning model:	CO3	K5
e.	Explain the principle of **Inverse Transform Sampling** and show how an exponential random variable can be generated using $\boldsymbol{X = -(1/\lambda) ln(1 - U)}$, where $\boldsymbol{U \sim Uniform(0,1)}$.	CO2	K4

SECTION C

3. Attempt any *one* part of the following: **07 x 1 = 07**

a.	Using **Monte Carlo simulation**, estimate $\int_0^1 \frac{1}{(1+x^2)}$ (i) Describe the random sampling procedure.	CO2	K6

PAPER ID-311543

Roll No: | | | | | | | | | | | | | |

BTECH
(SEM V) THEORY EXAMINATION 2025-26
MATHEMATICAL FOUNDATION AI, ML AND DATA SCIENCE

TIME: 3 HRS **M.MARKS: 70**

	(ii) Construct the Monte Carlo estimator. (iii) Discuss variance and convergence behavior.		
b.	Using the Gram–Schmidt orthogonalization process, obtain an orthonormal basis for the vectors: **(1, 1, 0), (1, 0, 1), (0, 1, 1)**. (i) Show all intermediate steps. (ii) Explain the importance of orthonormal bases in numerical stability.	CO4	K5

4. Attempt any *one* part of the following: 07 x 1 = 07

a.	Explain the concept of Analysis of Variance (ANOVA). Discuss the underlying assumptions of ANOVA, the logic of partitioning total variation into between-group and within-group variation, and explain how ANOVA is used to test the significance of differences among multiple population means in data science and machine learning applications.	CO3	K3
b.	Explain Singular Value Decomposition (SVD) of a matrix defined as $\boldsymbol{A = U\Sigma V^{T}}$. (i) Explain the significance of matrices $\boldsymbol{U}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{V}$. (ii) Analyze how SVD is used for dimensionality reduction and noise removal in machine learning and data science applications.	CO5	K4

5. Attempt any *one* part of the following: 07 x 1 = 07

a.	Explain the Metropolis–Hastings algorithm used in Markov Chain Monte Carlo (MCMC) methods. (i) Describe the role of proposal distribution $\boldsymbol{q(x' \mid x)}$. (ii) Explain the acceptance probability given by $\boldsymbol{\alpha = min(1, \pi(x')q(x \mid x') / [\pi(x)q(x' \mid x)])}$. (iii) Discuss convergence issues and challenges in high-dimensional probability spaces.	CO2	K6
b.	Discuss how numerical summaries (mean, variance, correlation) combined with graphical summaries (histograms, boxplots, scatter plots) help in interpreting and validating results of AI and ML models.	CO5	K3

6. Attempt any *one* part of the following: 07 x 1 = 07

a.	Perform a one-sample t-test to test whether the population mean is $\boldsymbol{\mu = 50}$, given the sample data: 48, 52, 55, 49, 51, 54, 50, 53, 52, 51. (i) State hypotheses. (ii) Compute the test statistic. (iii) Draw conclusion at $\boldsymbol{\alpha = 0.05}$.	CO3	K5
b.	Explain Type-I and Type-II errors with reference to hypothesis testing in machine learning model evaluation.	CO3	K4

7. Attempt any *one* part of the following: 07 x 1 = 07

a.	Explain change of basis in vector spaces and illustrate how it improves numerical stability in computations.	CO4	K5
b.	Discuss the importance of customized numerical and graphical summaries in presenting results of AI/ML analyses.	CO5	K3