18339

BRAINWARE UNIVERSITY

Term End Examination 2025-2026
Programme – M.Tech.(CSE)-AIML-2025
Course Name – Probability and Statistics for Machine Learning
Course Code - MTA10401
(Semester I)

Full Marks : 60 **Time : 2:30 Hours**

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

Group-A
(Multiple Choice Type Question) 1 x 15=15

1. *Choose the correct alternative from the following :*

(i) Select the PMF of a fair coin toss
a) 0.5 for each side
b) 1 for each side
c) 0 for both sides
d) 2 for both sides

(ii) Recall the range of a CDF
a) 0 to 1
b) 1 to ∞
c) -1 to 1
d) -∞ to ∞

(iii) Classify the expected value as a measure
a) Central tendency
b) Dispersion
c) Spread
d) Skewness

(iv) Use bootstrapping for confidence interval estimation.
a) Generate multiple samples
b) Use full population
c) Fix confidence level
d) Eliminate bias

(v) Apply bootstrapping on a small dataset.
a) Resample with replacement
b) Take moving average
c) Compute sample mean
d) Use normal approximation

(vi) Assess use of bootstrapping for hypothesis testing.
a) Valid for unknown distributions
b) Invalid
c) Only parametric
d) Only for mean

(vii) Choose a regularization technique that reduces model complexity.
a) Ridge
b) KNN
c) Naive Bayes
d) Linear Regression

(viii) Apply Ridge regression to handle multicollinearity.
a) Shrinks coefficients
b) Increases variance
c) Removes variables
d) Adds noise

(ix) Examine residuals to validate regression model.

a) Normality
b) Skew
c) Noise
d) Mean

(x) Estimate the number of clusters using elbow method.

a) Optimize within-cluster sum
b) Maximize variance
c) Choose PCA axes
d) Estimate entropy

(xi) Represent covariance between features.

a) Symmetric matrix
b) Diagonal matrix
c) Binary matrix
d) Sparse matrix

(xii) Create cluster summary for customer groups.

a) Mean and count
b) Confusion matrix
c) Covariance
d) Entropy

(xiii) Identify role of hidden variables in EM.

a) Increase dimensionality
b) Represent noise
c) Capture latent structure
d) Reduce bias

(xiv) Differentiate E and M steps of EM.

a) E maximizes, M estimates
b) E predicts, M updates
c) E estimates hidden variables, M optimizes parameters
d) E trains network

(xv) Construct a posterior using likelihood and prior in probabilistic learning.

a) Prior only
b) Likelihood only
c) Posterior = Likelihood x Prior / Evidence
d) Posterior = Prior / Evidence

Group-B

(Short Answer Type Questions) 3 x 5=15

2. Design a clustering algorithm combining K-means and hierarchical clustering. (3)
3. List the key characteristics of Gaussian distribution. (3)
4. Use confidence interval estimation to report the mean score from a sample mean of 50, std dev 10, and n=25 at 95%. (3)
5. Apply ANOVA to assess whether three diets result in different weight loss results. (3)
6. Design a simple Bayesian network for a student's exam performance. (3)

OR

Construct a probabilistic model combining EM and Gaussian Mixture Models. (3)

Group-C

(Long Answer Type Questions) 5 x 6=30

7. Analyze impact of data imbalance on ROC-AUC and t-test reliability. (5)
8. Construct a workflow using SVD for topic modeling in document-term matrix. (5)
9. Derive Bayes' theorem from conditional probability. (5)
10. Apply bootstrapping to estimate the confidence interval for the median in a sample {12, 15, 20, 21, 25}. (5)
11. Analyze the usefulness of resampling methods in robust statistical estimation. (5)
12. Construct a generative model using Bayesian principles for image synthesis. (5)

OR

Create a probabilistic deep learning model using dropout as Bayesian approximation. (5)

****