L	T	P	C
3	0	0	3

Course Code: CSE214

COMPUTATIONAL STATISTICS

Course Objective: To provide learners with in-depth knowledge on various computational statistical techniques. The student will able to design and implement machine learning algorithms for data analysis, clustering, regression and classification**.**

UNIT - I — 11 Periods

Multivariate Normal Distribution: Multivariate Normal Distribution Functions - Conditional Distribution and its relation to regression model - Estimation of parameters. **Multiple Linear Regression Model:** Standard multiple regression models with emphasis on detection of collinearity - outliers - non-normality and autocorrelation - Validation of model assumptions.

UNIT - II — 11 Periods

Multivariate Regression: Assumptions of Multivariate Regression Models - Parameter estimation - Multivariate Analysis of variance and covariance. **Discriminant Analysis:** Statistical background - Linear discriminant function analysis - Estimating linear discriminant functions and their properties.

UNIT - III — 12 Periods

Principal Component Analysis: Principal components - Algorithm for conducting principal component analysis - Deciding on how many principal components to retain H-plot. **Factor Analysis:** Factor analysis model - Extracting common factors - Determining number of factors- Transformation of factor analysis solutions - Factor scores.

UNIT - IV — 11 Periods

Cluster Analysis: Introduction, Types of clustering, Correlations and distances, Clustering by partitioning methods, Hierarchical clustering, Overlapping clustering, K-Means Clustering - Profiling and Interpreting Clusters.

TEXT BOOKS

1. Jobson, J. Dave, *Applied multivariate data analysis*, Vol, I & II, 2012, Springer-Verlag New York Inc.
2. Anderson TW, *An introduction to multivariate statistical analysis*, 3rd Edition, 2009, Wiley Publications.

REFERENCES

1. Mark Lutz, *Programming Python*, 4th Edition, 2010, O'Reilly Media.
2. Tim Hall and J-P Stacey, *Python 3 for Absolute Beginners*, 3rd Edition, 2009, Apress
3. Magnus Lie Hetland, *Beginning Python: From Novice to Professional*, 3rd Edition, 2005, Apress

UNITWISE LEARNING OUTCOMES

Unit I	The learner will have an understanding of the concepts of multivariate normal distribution and multiple linear regression models.
Unit II	The learner will have an understanding of the multivariate regression and discriminant analysis.
Unit III	The learner will have an understanding of the principal component and factor analysis.
Unit IV	The learner will be able to implement various clustering algorithms

COURSE OUTCOMES

Upon successful completion of this course, the learner will be able to

- Understand the multivariate normal distribution and apply multiple linear regression models
- Interpret multivariate regression and discriminant analysis
- Apply principal component and factor analysis for preprocessing
- Implement various clustering algorithms for different applications