

HONOURS SUBJECT

PCS4D001 DATA ANALYTICS (4-0-0)

(I) Predictive Analytics

1.Linear Methods for Regression and Classification: Overview of supervised learning,Linear regression models and least squares, Multiple regression, Multiple outputs, Subset selection , Ridge regression, Lasso regression , Linear Discriminant Analysis , Logistic regression , Perceptron learning algorithm.

2.Neural Networks(NN) , Support Vector Machines(SVM),and K-nearest Neighbor: Fitting neural networks, Back propagation,Issues in training NN, SVM for classification, Reproducing Kernels, SVM for regression, K-nearest –Neighbour classifiers(Image Scene Classification)

3.Unsupervised Learning and Random forests: Association rules,Cluster analysis,Principal Components,Random forests and analysis.

(II) Inferential Statistics and Prescriptive analytics

4.Assessing Performance of a classification Algorithm(t-test,McNemar’s test,Paired t-test,paired F-test),Analysis of Variance, Creating data for analytics through designed experiments.

Introduction to big data and Challenges for big data analytics.

(III)Lab work

5. Implementation of following methods using R or Matlab (One of the class tests with a weightage of 15 marks be used to examine these implementations):

Simple and multiple linear regression,Logistic regression,Linear discriminant analysis,Ridge regression, Cross-validation and boot strap, Fitting classification and regression trees, K-nearest neighbours, Principal component analysis ,K-means clustering.

RecommendedTexts:

1.Trevor Hastie, Robert Tibshirani,Jerome Friedman , *The Elements of Statistical Learning-Data Mining, Inference,and Prediction*,Second Edition , Springer Verlag, 2009.

[chapters: 2,3(3.1-3.4,3.6),4(4.3-4.5),11(11.3-11.6),12(12.1-12.3),13.3,14(14.1-14.3.8,14.5.1),15]

2. (**For unit 5 only**) -G.James,D.Witten,T.Hastie,R.Tibshirani-*An introduction to statistical learning with applications in R*,Springer,2013.(2.3,3.6.1-3.6.3,4.6.1-4.6.3,5.3,6.6.1,8.3.1,8.3.2,10.4,10.5.1)

3 (**for unit 4 only**).E.Alpaydin, *Introduction to Machine Learning*, Prentice Hall Of India,2010,(Chapter-19)

References

1.C.M.Bishop –*Pattern Recognition and Machine Learning*,Springer,2006

2. L.Wasserman-*All of statistics*

Texts 1 and 2 and reference 2 are available on line.