PCS4D001 Honours (CP) 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.Model Assesment and Selection : Bias,Variance,and model complexity,Bias-variance trade off, Optimisim of the training error rate ,Esimate of In-sample prediction error,Effective number of parameters, Bayesian approach and BIC, Cross- validation ,Boot strap methods, conditional or expected test error.

3.Additive Models,Trees,and Boosting: Generalized additive models**,** Regression and classification trees , Boosting methods-exponential loss and AdaBoost, Numerical Optimization via gradient boosting ,Examples (Spam data, California housing , NewZealand fish, Demographic data)

4.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)

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

(II) Inferential Statistics and Prescriptive analytics

6.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

7. 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 discreminant analysis, Ridge regression, Cross-validation and boot strap, Fitting classification and regression trees, K-nearest neighours, Principal component analysis ,K-means clustering.

Recommended Texts:

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),7(excluding 7.8 and 7.9),9(9.1,9.2),(10.1-10.5,10.8,10.10,10.14),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 7 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 6 only**).E.Alpaydin, *Introduction to Machine Learning*, Prentice Hall Of India,2010,(Chapter-19)

Refeerences

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.

Formal Language & Automata Theory Lab

Implementation of following concept of Theory of computation using C-program:

1. DFAs for some regular languages
2. **ϵ**-NFA to DFA conversion
3. NFA to DFA conversion
4. Program for DFA minimization
5. PDAs for some Context free languages
6. CYK parsing algorithm for some specific Context free grammars
7. Turing machine for some Recursively Languages

PRACTICE LIST OF EXPERIMENTS

1. (a) Identification of different components of a PC.
(b) Assembling & disassembling of a PC.

2. Study of different troubleshooting of a dot matrix printer using LX 1050+ Printer Trainer Module.

3. Study of the functions of SMPS using SMPS Trainer Kit.
(a) Study of SMPS with Single Output under Line Regulation.
(b) Study of SMPS with Multi Output under Line Regulation.
(c) Study of SMPS with Single Output under Load Regulation.

4. Study of different troubleshooting of CPU using CPU Trainer Module.

5. Familiarization of different types of byte addressing instruction using 8085 simulator.

6. Study of assembly Language program in PC using 8086 architecture.

7. Design of digital circuits (H/A, F/A, Decoder & Encoder) in VHDL using Active VHDL.

8. Design of digital circuits (MUX, DEMUX & ALU) in VHDL using Active VHDL.

9. Write a C/C++ program to perform signed bit multiplication using Booth's algorithm.

10. Write a C/C++ program for IEEE-754 floating point representation and perform Addition/Subtraction.