DISCIPLINE SPECIFIC CORE COURSE – 11: Machine Learning (INDSC4B)

CREDIT DISTRIBUTION, ELIGIBILITY AND PRE-REQUISITES OF THE COURSE

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course (if any)
		Lecture	Tutorial	Practical/ Practice		
Machine Learning (INDSC4B)	**04**	**02**	**-**	**02**	**Class XII passed with Physics + Mathematics /Applied Mathematics + Chemistry/ Computer Science/Informatics Practices**	**Understanding of Mathematics & programming language**

Learning Objectives

The Learning Objectives of this course are as follows:

- Students have an understanding of issues and challenges of Machine Learning.
- Students should be able to select data, model selection, model complexity etc.
- Understanding of the strengths and weaknesses of many popular machine learning approaches.

Learning outcomes

The Learning Outcomes of this course are as follows:

- Identify the characteristics of datasets and compare the trivial data and big data for various applications.
- Understand machine learning techniques and computing environments that are suitable for the applications under consideration .
- Solve problems associated with batch learning and online learning, and the big data characteristics such as high dimensionality, dynamically growing data and in particular scalability issues.
- Develop scaling up machine learning techniques and associated computing techniques and technologies for various applications.
- Implement various ways of selecting suitable model parameters for different machine learning techniques.
- Integrate machine learning libraries, and mathematical and statistical tools with modern

- technologies like hadoop distributed file system and mapreduce programming model
- Familiarize with Simple Linear Regression and Logistic Regression.
- Appreciate the various nuances of Multiple Regressions and Model Building.
- Identify and apply the Classification algorithms.
- Apply the Clustering algorithms for developing applications

SYLLABUS OF DSC-11

UNIT – 1 **(8 hours)**

Introduction to Machine Learning: varieties of machine learning, Supervised Learning, Unsupervised Learning, Reinforcement Learning. Dimensionality Reduction, Subset Selection, Shrinkage Methods, Principal Components Regression: Linear Classification, Logistic Regression, Linear Discriminant Analysis, Optimization, Classification-Separating Hyperplanes Classification.

UNIT – 2 **(8 hours)**

Learning input/output functions, sample application. Boolean functions and their classes, CNF, DNF, decision lists and Bias – Variance, Version spaces for learning, version graphs, learning search of a version space, candidate elimination methods.

UNIT – 3 **(8 hours)**

Artificial Neural Networks (Early models, Back Propagation, Initialization, Training & Validation) Parameter Estimation (Maximum Likelihood Estimation, Bayesian Parameter Estimation) Decision Trees: ID4, C4.5, CART, Evaluation Measures, Hypothesis Testing.

UNIT – 4 **(6 hours)**

Clustering, Gaussian Mixture Models, Spectral Clustering, Ensemble Methods Learning Theory, Graphical Models.

K-Nearest Neighbors: Computational geometry; Voronoi Diagrams; Delaunay Triangulations K-Nearest Neighbor algorithm; Wilson editing and triangulations. Aspects to consider while designing K-Nearest Neighbor, Support Vector Machines and its classifications. Linear learning machines and Kernel space, Making Kernels and working in feature space.

Practical component: **(60 hour)**

Hardware requirement: i5 Processor, 8GB RAM, Internet Connection

Software Environment: IDE recommended PYCHARM (Recommended), JUPYTER, VISUAL STUDIO

1. Introduction to pandas and NumPy
2. Prediction based on different dataset: Vegetable Quality Prediction, Housing Price Prediction, Air Quality Prediction, Car Price Prediction

3. Prediction of diseases e.g. Liver Disease Prediction, Heart Disease Prediction, Crop disease.
4. Credit Default Prediction, Airline Passengers Prediction, Stock Price Prediction.
5. Bank Marketing, Media Content Problem, Online Retail Case Study
6. Energy Efficiency Analysis, Movie Sentiment Analysis, Car Evaluation
7. Program to demonstrate Simple Linear Regression
8. Program to demonstrate Logistic Regression using SCIKIT learn
9. Program to demonstrate Logistic Regression
10. Program to demonstrate k-Nearest Neighbor flowers classification
11. Program to demonstrate Decision Tree – ID3 Algorithm
12. Program to demonstrate Naïve- Bayes Classifier
13. Program to demonstrate Back-Propagation Algorithm
14. Program to demonstrate k-means clustering algorithm
15. Program to demonstrate K-Means Clustering Algorithm on Handwritten Dataset
16. Program to demonstrate K-Medoid clustering algorithm
17. Program to demonstrate DBSCAN clustering algorithm
18. Program to demonstrate SVM based classification
19. Program to demonstrate PCA on face recognition
20. Program to demonstrate PCA and LDA on Iris dataset
21. Mini Project works shall be given with a batch of four students considering different datasets such as digit dataset, face dataset, flower dataset and micro-array dataset.

Essential/recommended readings

1. Introduction to Machine learning, Nils J.Nilsson
2. Pattern Recognition and Machine Learning. Christopher Bishop. First Edition, Springer, 2006.
3. Pattern Classification. Richard Duda, Peter Hart and David Stock. Second Edition, Wiley-Interscience, 2000.
4. Machine Learning. Tom Mitchell. First Edition, McGraw-Hill, 1997.
5. T. Hastie, R. Tibshirani, J. Friedman. The Elements of Statistical Learning, 2e, 2008.

Suggestive readings

1. Christopher Bishop. Pattern Recognition and Machine Learning. 2e.
2. Tom M. Mitchell, “Machine Learning”, McGraw-Hill, 2010
3. Bishop, Christopher. Neural Networks for Pattern Recognition. New York, NY: Oxford University Press, 1995.

Note: Examination scheme and mode shall be as prescribed by the Examination Branch, University of Delhi, from time to time.