RAMAKRISHNA MISSION VIDYAMANDIRA

CBCS Syllabus B.Sc. Computer Science Honours

Semester-VI

Credit: 6
Course Type: Discipline Specific Elective

Course Outcome:

i) Master the concepts of supervised and unsupervised learning, recommendation engine, and time series modeling
ii) Gain practical mastery over principles, algorithms, and applications of Machine Learning through a hands-on approach.
iii) Acquire thorough knowledge of the statistical and heuristic aspects of Machine Learning
iv) Implement support vector machines, kernel SVM, Naive Bayes, decision tree classifier, random forest classifier, logistic regression, K-means clustering, and more in Python
v) Validate Machine Learning models and decode various accuracy metrics; improve the final models using another set of optimization algorithm, which include boosting and bagging techniques
vi) Comprehend the theoretical concepts and how they relate to the practical aspects of Machine Learning.

CMSA DSE T: Machine Learning

Credit: 4 **Marks: 50**

Foundations of Machine Learning (ML): ML Techniques overview, Validation Techniques (Cross-Validations), Feature Reduction/Dimensionality reduction, Principal components analysis (Eigen values, Eigen vectors, Orthogonality) [5 L]

Data Management:

Data Acquisition: Gather information from different sources, Internal systems and External systems, Web APIs, Open Data Sources, Data APIs, Web Scrapping, Relational Database access (queries) to process/access data.

Data Pre-processing and Preparation: Data Munging, Wrangling, Plyr packages, Cast/Melt.

Data Quality and Transformation: Data imputation, Data Transformation (minmax, log transform, z-score transform etc.), Binning, Classing and Standardization, Outlier/Noise& Anomalies.

Handling Text Data: Bag-of-words, Regular Expressions, Sentence Splitting and Tokenization, Punctuations and Stop words, Incorrect spellings, Properties of words and Word cloud, Lemmatization and Term-Document T×D computation, Sentiment Analysis (Case Study).

Introduction to Big Data: Introduction to Big Data, Challenges of processing Big Data (Volume, Velocity and Variety perspective), Use Cases. [10 L]

Data Visualization: Design of data collection formats with illustration, Principles of data visualization - different methods of presenting data in business analytics, Concepts of Size, Shape, Color, Various Visualization types, Bubble charts, Geo-maps (Chlorpeths), Gauge charts, Tree map, Heat map, Motion charts, Force Directed Charts etc. [5 L]

Predictive Analytics:

Linear Regression: Regression basics: Relationship between attributes using Covariance and Correlation, Relationship between multiple variables: Regression (Linear, Multivariate) in prediction. Residual Analysis, identifying significant features, feature reduction using AIC, multi-collinearity, Non-normality and Heteroscedasticity, Hypothesis testing of Regression Model, Confidence intervals of Slope, R-square and goodness of fit, Influential Observations – Leverage. [8 L]

Multiple Linear Regressions: Polynomial Regression, Regularization methods, Lasso, Ridge and Elastic nets, Categorical Variables in Regression.

Non-Linear Regression: Logit function and interpretation, Types of error measures (ROCR), Logistic Regression in classification.

Forecasting models: Trend analysis, Cyclical and Seasonal analysis, Smoothing; Moving averages; Box-Jenkins, Holt-winters, Auto-correlation; ARIMA, Examples: Applications of Time Series in financial markets. [7 L]

Machine Learning Techniques:

Supervised Learning (Regression/Classification): Basic methods: Distance-based methods, Nearest-Neighbours, Decision Trees, Naive Bayes; Linear models: Linear Regression, Logistic Regression, Generalized Linear Models, Support Vector Machines, Nonlinearity and Kernel Methods, Beyond Binary Classification: Multi-class/Structured Outputs, Ranking. Ensembles methods:Bagging & boosting and its impact on bias and variance, C5.0 boosting, Random forest, Gradient Boosting Machines and XGBoost, Association Rule mining:The applications of Association Rule Mining: Market Basket, Recommendation Engines, etc., A mathematical model for association analysis; Large item sets; Association Rules, Apriori: Constructs large item sets with mini sup by iterations; Interestingness of discovered association rules; Application examples; Association analysis vs. classification, FP-trees. [8 L]

Unsupervised Learning (Clustering): Distance measures, Different clustering methods (Distance, Density, Hierarchical) , Iterative distance-based clustering, Dealing with

continuous, categorical values in K-Means, Constructing a hierarchical cluster, K-Medoids, k-Mode and density-based clustering, Measures of quality of clustering, Dimensionality Reduction: PCA and kernel PCA, Matrix Factorization and Matrix Completion, Generative Models (mixture models and latent factor models). [7 L]

Evaluating Machine Learning algorithms and Model Selection: Introduction to Statistical Learning Theory, Ensemble Methods (Boosting, Bagging, Random Forests) Sparse Modelling and Estimation, Modelling Sequence/Time-Series Data, Deep Learning and Feature Representation Learning**.** [5 L]

Scalable Machine Learning (Online and Distributed Learning): A selection from some other advanced topics, e.g., Semi-supervised Learning, Active Learning, Reinforcement Learning, Inference in Graphical Models, Introduction to Bayesian Learning and Inference Simulation Tool for Machine Learning, Hands on with recent tools WEKA, R, MATLAB Recent trends in various learning techniques of machine learning and classification methods for IOT applications. Various models for IOT applications. [5 L]

CMSA DSE P: Machine Learning Laboratory

Credit: 2 **Marks: 25**

- Introduction to various machine learning packages in python like NumPy, SciPy, Pandas, Seaborn, Scikit-Learn, Matplotlib, Natural Language Toolkit (NLTK), TensorFlow, Pytorch, Keras etc. [10 L]

- Introduction to data preparation, pre-processing, and wrangling using Plyr packages and other tools, Sentiment Analysis (Case Study), Intrduction to Bubble charts, Geo-maps (Chlorpeths), Gauge charts, Tree map, Heatmap, Motion charts, Force Directed Charts for data visualization. [15 L]

- Face Recognition using PCA; Practical applications of clustering; Experiments on supervised classification using MLP, RBF ANN, SVM and Decision Trees; Application of Classifiers Ensembles; Sequence classification using HMM; Applications of CNN and RNN; Path planning with Reinforcement Learning. [15 L]

Recommended Books and online resources:

1. Kevin Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012.
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning, Springer 2009 (freely available online).
3. Christopher Bishop, Pattern Recognition and Machine Learning, Springer, 2007.
4. Shai Shalev-Shwartz, Shai Ben-David, Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press; 1 edition (May 19, 2014)
5. https://keras.io/getting_started/learning_resources/
6. https://www.tensorflow.org/resources/learn-ml/basics-of-machine-learning