Suggested Teaching Guidelines for

Practical Machine LearningPG-DAI February 2025

Duration: 52 Classroom hours and 48 Lab hours

Objective: Practicing Machine Learning Algorithms

Prerequisites: Good knowledge of Python Programming and Statistics

Evaluation method: Theory exam– 40%
Lab Exam - 40%
Internal exam- 20%

List of Books / Other training material

Courseware:

Machine Learning using Python, Manaranjan Pradhan , U Dinesh Kumar , Wiley India, 1st Edition

Reference Book:

1. Introduction to Machine Learning with Python – A guide for Data Scientists byAndreasmuller, Sarah Guido.
2. Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2, 3rdEdition by Sebastian Raschka
3. Hands-On Machine Learning with Scikit-Learn, Keras and Tensor Flow: Concepts, Toolsand Techniques to Build Intelligent Systems by AurelienGeron
4. AI and Machine Learning for Coders: A Programmer's Guide to Artificial Intelligence byLaurence Moroney
5. Machine Learning for Big Data: Hands- On for Developer by Jasaon Bell , Wiley

Note:

- **Each session mentioned is of 2 hrs of Theory and 2 hours of Lab duration, unless indicated otherwise. T for Theory & L for Lab.**

Session 1: (2T)

- Machine Learning in Nutshell
- Supervised and Unsupervised Learning
- ML Applications
- Evaluating ML techniques

Session 2 & 3: (4T + 2L)

- Feature engineering and Data Pre-processing: Data Preparation, Feature creation,
- Data cleaning & transformation
- Data Validation &Modelling
- Feature selection Techniques
- Dimensionality reduction

Machine Learning Algorithms:

Session 4 & 5:

- Principal Component analysis (PCA)
- t-SNE
- K means
- Distance Measure and Data Preparation – Scaling & Weighting

Lab Assignments:

Download any dataset from kaggle.

(a) Form n no .of clusters according to your observation.
(b) Get WSS value for each cluster.
(c) Find best K value

Explore t-SNE

Session 6:

- Evaluation and Profiling of Clusters
- Hierarchical Clustering
- Clustering Case Study

Lab Assignments: Implement the model hierarchical clustering on any dataset loaded from kaggle.

Session 7:

- DBSCAN Clustering

Session 8 & 9:

- Linear Regression
- Logistic Regression
- Polynomial Regression
- Stepwise Regression
- Ridge Regression
- Lasso Regression
- Elastic Net Regression

Lab Assignments:

Download Dataset, perform linear, Ridge, Lasso, Logistic, Polynomial regression and check forMAE, MSE, RMSE and also check F1-score and explain with conclusion.

Session 10 & 11:

- Support vector Machines
- Basic classification principle of SVM
- Linear and Nonlinear classification (Polynomial and Radial)

Lab Assignments:
Download Air Quality Dataset from Kaggle Predict Air Quality Index using Linear regressionand classify it into five categories using SVM (i.e. Very good, good, moderate, poor, worst)

Session 12:

- Discriminant Analysis (Linear and Quadratic)
- K-Nearest Neighbors Algorithm

Lab Assignments:
Load any dataset from kaggle, preprocess the data and apply KNN Algorithm.

Session 13:

- Decision Trees
- Classification and Regression Trees

Lab Assignments:
Load any dataset from UCI Machines preprocess the data and apply Decision tree Algorithm.

Session 14:

- Bayesian analysis and Naïve Bayes classifier
- Assigning probabilities and calculating results

Lab Assignments:
Load any dataset preprocess the data and apply Naïve Bayes Algorithm.

Session 15 & 16:

- Concept of Model Ensembling
- Random forest, Gradient boosting Machines, Model Stacking

Lab Assignments:
Implement Random Forest and gradient boosting algorithm on any dataset.

Session 17 & 18:

- Association rules mining
- Apriori algorithms

Lab Assignments:
Generate Association Rules for shop using Apriori.

Session 19 & 20:

- Moving average, Exponential Smoothing
- Holt's Trend Methods
- Holt-Winters' Methods for seasonality

Lab Assignments:
Implement the smoothing models on the time series data.

Session 21 & 22:

- Auto-correlation (ACF & PACF)
- Auto-regression
- Auto-regressive Models
- Moving Average Models
- ARMA &ARIMA

Lab Assignments:
Download one data set and calculate Auto correlation. oExplain ARMA and ARIMA model intime series and Explain difference between them by applying on any dataset.

Session 23:

- ML in Real-Time
- Algorithm Performance Metrics
- ROC and AOC
- Confusion Metrix
- F1 Score
- MSE and MAE

Lab Assignments:
Calculate Confusion matrix, F1 score, MSE, RMSE, MAE scores on any model.
Load a dataset in Jupyter notebook and implement any classification algorithm and generateclassification report and accuracy score.

Session 24 & 25:

- Recommendation Systems
- Data Collection & Storage, Data Filtering
- Collaborative Filtering
- Factorization Methods
- Evaluation Metrics: Recall, Precision, RMSE, Mean Reciprocal Rank, MAP at K,NDCG

Lab Assignments:
Download Movie dataset from kaggle and implement the movie recommendation system.

Session 26:

- Anomaly detection
- Point, Contextual and Collective Anomaly
- Supervised and Unsupervised anomaly detection

Lab Assignments:

Apply Anomaly Detection on dataset using isolation forest algorithm and calculate the accuracy.

Case Studies:

- Usage of ML algorithms, Algorithm performance metrics (confusion matrix sensitivity,Specificity, ROC, AOC, F1score, Precision, Recall, MSE, MAE)
- Credit Card Fraud Analysis
- Intrusion Detection system