Machine Learning

Introduction to Machine Learning

- Introduction to Data Science, Types of Analytics - Descriptive, predictive, prescriptive, diagnostic.
  Overview of data analysis pipeline
- Introduction to Python: Understanding the tool - Jupyter notebook, basics of Python language, input/output, built-in functions

Programming in Python for Data Science

- Introduction to Python: Operators, Control structures, looping structures, user defined functions
- Introduction to Python: Python Data Structures, Strings, Lists, Tuples, Dictionaries, File objects
- Introduction to Python: Lambdas, list comprehension, working with flat files
- Object Oriented Programming using Python: Methods, Constructors, __init__(), self etc. Creating objects, calling methods
- Virtual Environment Setup. Pip command, IDE for python – Pycharm, jupyter notebook, VS code
  Introduction to NumPy: NumPy arrays, array operations, reshape etc.
- Introduction to Pandas: Pandas series and dataframe, basic operations on dataframe, reading csv files, extracting rows and columns from the data frame
- Working with tabular data (Basics): Doing SQL like tasks using pandas: sorting data, group by aggregations, map and apply constructs
- Working with tabular data (Misc.): Doing joins, working with dates and strings, doing miscellaneous transformations
- Working with JSON files, PDF files and Word docs using PyPDF2 packages

Exploratory Data Analysis and Data Visualization using Python

- Data Visualization using Python: Creating univariate plots such as barplots, histograms, boxplots.
- Creating bivariate plots such as scatter plots, conditional scatter plots
- Data Exploration and Preparation: Doing sanity checks, treating missing values, finding outliers and treating them.
- Pandas Profiling
- Identify potential predictors, extract useful features from date data, string data, bin continuous variables
- Data Exploration: Outlier analysis, data normalization: Z-score, min-max normalizations, Data transformations/encoding like one-hot encoding, label encoding etc.
- Dimensionality Reduction, Feature Engineering using techniques like PCA

Machine Learning Algorithms

- Introduction to Machine Learning: What is Machine Learning, ML vs. Traditional programs, Supervised Learning Definition, Unsupervised Learning Definition, Applications of ML
- Introduction to Optimization Problems: Need for Optimization. Building an optmization problem. Define objective function, constraints etc.
-

- Demonstration of optimization problem with the examples like Travelling salesman problem, shortest path algorithms, Assignment problem, Transportation Problem etc.
- Linear Regression: Understanding regression problem, interpreting beta coefficients, validating model fit using holdout set
- Linear Regression: Checking OLS assumptions, finalizing model. Drawing insights from the final model
- Logistic Regression: Implementing logistic regression on a classification problem. Finalizing the model, drawing insights from the final model
- Evaluation Metrics for classification problems: Confusion Matrix, Accuracy, Precision, Recall, ROC, AUC
  Class imbalanced problem, SMOTE
- Naïve Bayes Algorithm for Classification, Need for Laplace smoothing in Naïve Bayes
  SVM Algorithm, Kernals, Hyperparameters
- Instance Based Learning: KNN Algorithm.
  Tree Based Models: Classification trees, Gini, entropy.
- Decision Trees: Implementation, variable importance, and feature engineering, Hyperparameter tuning using GridSearch
- Tree based ensembles: Bagged trees, random forests and boosted trees. Hyper parameter tuning for tree based ensembles
- Clustering: Understanding the meaning of unsupervised learning, using kmeans to create clusters, find optimal number of clusters
- Clustering: Creating cluster profiles, using agglomerative clustering, doing silhouette analysis.

Overview of Neural Networks

- Need for Neural Networks over ML algorithms. Linear v/s non-linear separability of the classes, interaction features and a simple demo on how NN differs from ML
- Concepts of Loss functions, Activation Functions, Weight updates, Gradient Descent, Learning Rate etc
  Basics of Feed Forward Neural Networks and back propagation