

Machine Learning

1. Introduction

- Definition
- History
- Limitations and Considerations
- Key Elements of Machine Learning
 - Representation
 - Evaluation
 - Optimization
- Understanding the importance and impact of
 - Algorithm (Low)
 - Feature Selection (Medium)
 - Real World Data (High)
- Categories of Machine Learning
 - Supervised
 - Unsupervised
 - Semi-Supervised
 - Reinforcement
- Machine Learning Techniques
 - Classification
 - Regression
 - Linear Regression
 - Logistic Regression
 - Clustering
 - K-Means Clustering
 - Dimensionality Reduction
 - Classification Vs. Regression

2. Representing and Manipulating Data with Numpy and Pandas

- Gathering Data
 - Quantifying real world data

- Representing Data
- What is,
 - Training Data
 - Testing Data
 - Validation Data
- Splitting a Dataset
- Normalizing Data
 - Max Scaling
 - Min Max Scaling
- Overfitting and regularization
- Introduction to Numpy
 - Understanding datatypes in Python and Numpy
 - The Basics of NumPy Arrays
 - Computation on NumPy Arrays: Universal Functions
 - Aggregations: Min, Max, and Everything In Between
 - Computation on Arrays: Broadcasting
 - Comparisons, Masks, and Boolean Logic
 - Fancy Indexing
 - Sorting Arrays
 - Structured Data: NumPy's Structured Arrays
- Introduction to Pandas
 - Introducing Pandas Objects
 - Data Indexing and Selection
 - Operating on Data in Pandas
 - Handling Missing Data
 - Hierarchical Indexing
 - Combining Datasets: Concat and Append
 - Combining Datasets: Merge and Join
 - Aggregation and Grouping
 - Pivot Tables
 - Vectorized String Operations

- Working with Time Series
- High-Performance Pandas: `eval()` and `query()`

3. Visualizing Data with Matplotlib

- Simple Line Plots
- Simple Scatter Plots
- Visualizing Errors
- Density and Contour Plots
- Histograms, Binnings, and Density
- Customizing Plot Legends
- Customizing Colorbars
- Multiple Subplots
- Text and Annotation
- Customizing Ticks
- Customizing Matplotlib: Configurations and Stylesheets
- Three-Dimensional Plotting in Matplotlib
- Geographic Data with Basemap

4. Machine Learning using SciKit Learn

- Introduction to SciKit
 - Problem Approach with SciKit
 - Fit and Transform Methods
 - Predict Method
 - Importing Models
 - Choosing Hyperparameters and Model Validation
 - Cross Validation
 - Bias Variance Tradeoff
 - Validation Curves
 - Learning Curves
 - Feature Engineering
 - Categorical Features
 - Text Features
 - Image Features

- Derived Features
- Imputation of Missing Data
- Constructing Feature Pipelines
- ASSESSMENT: Min Max Scaling
- Naive Bayes Classification
 - Bayes Theorem
 - Gaussian Bayes
 - Multinomial Bayes
 - Visualizing with Confusion Matrix
 - LAB: Predict Onset of Diabetes using Bayes Rule
 - LAB: Modelling Coin flip patterns using Bayes Rule
- Linear Regression
- Support Vector Machines
 - CASE STUDY: Face Recognition
- Decision Trees and Random Forests
 - ASSESSMENT: Predicting Employee Retention
- Principal Component Analysis
- Manifold Learning
- k-Means Clustering
 - Choosing K
 - Visualizing with Pairplots
 - ASSIGNMENT: Identifying MNIST Data
- Gaussian Mixture Models
- Kernel Density Estimation