



Machine Learning Training

1. Day one first half

a. Python basics - Part 1

- i. Understanding Python code
- ii. Importing modules
- iii. Data structures
- iv. Experimenting with lists
- v. Pre-colon
- vi. Post colon
- vii. Negative syntax
- viii. Adding list to list
- ix. The append function
- x. Complex data structures
- xi. Dereferencing a single element
- xii. The sort function
- xiii. Reverse sort
- xiv. Tuples Dereferencing an element
- xv. List of tuples
- xvi. Dictionarie
- xvii. Iterating through entries

b. Python basics - Part 2

- i. Functions in Python
- ii. Lambda functions - functional programming
- iii. Understanding Boolean expressions
- iv. The if statements
- v. The if-else loop
- vi. Looping
- vii. The while loop
- viii. Exploring activity
- ix. Running Python scripts
- x. More options than just the IPython/Jupyter Notebook

2. Day 1 second half

a. Value and Variables

- i. Statistics and Probability Refresher, and Python Practice
- ii. Types of data
- iii. Numerical data
- iv. Discrete data
- v. Continuous data
- vi. Categorical data
- vii. Ordinal data
- viii. Mean, median, and mode
- ix. Mean

- x. Median
- xi. The factor of outliers
- xii. Mode
- xiii. Using mean, median, and mode in Python
- xiv. Calculating mean using the NumPy package
- xv. Visualizing data using matplotlib
- xvi. Calculating median using the NumPy package
- xvii. Analysing the effect of outliers Calculating mode using the SciPy package
- xviii. Some exercises

b. Data Metrics

- i. Standard deviation and variance
- ii. Variance
- iii. Measuring variance
- iv. Standard deviation
- v. Identifying outliers with standard deviation
- vi. Population variance versus sample variance
- vii. The Mathematical explanation
- viii. Analysing standard deviation and variance on a histogram
- ix. Using Python to compute standard deviation and variance
- x. Try it yourself
- xi. Probability density function and probability mass function
- xii. The probability density function and probability mass functions
- xiii. Probability density functions
- xiv. Probability mass functions

3. Day 2 first half

a. Data distributions

- i. Uniform distribution
- ii. Normal or Gaussian distribution
- iii. The exponential probability distribution or Power law
- iv. Binomial probability mass function
- v. Poisson probability mass function
- vi. Percentiles and moments
- vii. Percentiles
- viii. Quartiles
- ix. Computing percentiles in Python

b. Data Visualization

- i. Matplotlib and Advanced Probability Concepts
- ii. A crash course in Matplotlib
- iii. Generating multiple plots on one graph
- iv. Saving graphs as images
- v. Adjusting the axes
- vi. Adding a grid Changing line types and colours

- vii. Labelling axes and adding a legend
- viii. A fun example
- ix. Generating pie charts
- x. Generating bar charts
- xi. Generating scatter plots
- xii. Generating histograms
- xiii. Generating box-and-whisker plots
- xiv. Try it yourself
- xv. Covariance and correlation
- xvi. Defining the concepts
- xvii. Measuring covariance

4. Day 2 Second Half

a. Correlation

- i. Computing covariance and correlation in Python
- ii. Computing correlation – The hard way
- iii. Computing correlation – The NumPy way
- iv. Correlation activity
- v. Conditional probability
- vi. Conditional probability exercises in Python
- vii. Conditional probability assignment
- viii. My assignment solution
- ix. Bayes' theorem

5. Day 3 first half

a. Predictive Models

- i. Linear regression
- ii. The ordinary least squares technique
- iii. The gradient descent technique
- iv. The co-efficient of determination or r-squared
- v. Computing r-squared
- vi. Interpreting r-squared
- vii. Computing linear regression and r-squared using Python
- viii. Activity for linear regression Polynomial regression
- ix. Implementing polynomial regression using NumPy
- x. Computing the r-squared error
- xi. Activity for polynomial regression
- xii. Multivariate regression and predicting car prices
- xiii. Multivariate regression using Python
- xiv. Activity for multivariate regression
- xv. Multi-level models

6. Day 3 second half

a. Machine Learning with Python

- i. Machine learning and train/test
- ii. Unsupervised learning
- iii. Supervised learning
- iv. Evaluating supervised learning
- v. K-fold cross validation
- vi. Using train/test to prevent overfitting of a polynomial regression
- vii. Activity
- viii. Bayesian methods - Concepts
- ix. Implementing a spam classifier with Naïve Bayes
- x. Activity
- xi. K-Means clustering
- xii. Limitations to k-means clustering
- xiii. Clustering people based on income and age
- xiv. Activity
- xv. Measuring entropy
- xvi. Decision trees - Concepts
- xvii. Decision tree example
- xviii. Walking through a decision tree
- xix. Random forests technique
- xx. Decision trees - Predicting hiring decisions using Python

7. Day 4 first half

a. Ensemble learning

- i. Activity
- ii. Ensemble learning Using a random forest
- iii. Support vector machine overview
- iv. Using SVM to cluster people by using scikit-learn Activity
- v. Recommender Systems
- vi. What are recommender systems?
- vii. User-based collaborative filtering
- viii. Limitations of user-based collaborative filtering
- ix. Item-based collaborative filtering
- x. Understanding item-based collaborative filtering
- xi. How item-based collaborative filtering works?
- xii. Collaborative filtering using Python

8. Day 4 Second half

a. Data Mining and Machine Learning Techniques

- i. K-nearest neighbours - concepts
- ii. Using KNN to predict a rating for a movie
- iii. Activity
- iv. Dimensionality reduction and principal component analysis

- v. Dimensionality reduction
- vi. Principal component analysis
- vii. A PCA example with the Iris dataset
- viii. Activity

9. Day 5 First half

a. Advanced

- i. XGBOOST
- ii. CATBOOST
- iii. Ensemble methods
- iv. bagging
- v. boosting
- vi. ADABOOST
- vii. Voting

b. Case Study

- i. part 1 (mini case) source Kaggle

10. Day 5 Second half

a. Case Study

- i. part 2 (full case) source Kaggle