

Machine Learning using Python

MODULES

KEY LEARNING OUTCOMES

WEEK 1

Introduction

- Introduction to data science
- Why Python for data science

Introduction to Spyder

- Setting working directory
- Creating and saving script files
- Executing pieces of code
- Commenting
- Clearing the environment and console
- Removing variables from environment
- Commenting script files
- Creating variables in Python and naming conventions
- Arithmetic operators
- Logical operators
- Data types and related conversions

Participants will be able to :-

- Set-up Python
- Run basic commands in Python
- Differentiate between data types
- To create and manipulate different data types

WEEK 2

Sequence data types and associated operations

- Strings
- Lists
- Arrays
- Tuples
- Dictionary
- Sets
- Range

NumPy

- ndArray

Participants will be able to :-

- To create and manipulate strings, lists, arrays, tuples, dictionary, sets, range

WEEK 3

Basics of Statistics

- Descriptive statistics
- Measures of central tendency
- Measures of spread
- Distribution of mean and variance
- Sampling basics
- Notion of probability

Participants will be able to :-

- Compute basic statistical measures from data
- Compute probabilities of events
- Compute confidence intervals of estimates
- Conduct exploratory data analysis on the data
- Draw inferences from their analysis
- Present preliminary results and inferences from their analysis

WEEK 4

Pandas dataframe and dataframe related operations

- Reading files
- Comma separated value files
- Tab-delimited files
- Excel files
- Exploratory data analysis
- Data preparation and pre- processing

Data visualization using matplotlib and seaborn libraries

- Scatter Plot
- Bar Plot
- Histogram
- Box plot
- Pair plot

Control structures

- If-else-if family
- For loop
- For loop with "if break"
- While loop

Functions

Participants will be able to :-

- Graphically visualize data to draw inferences
- Present results in a visually appealing manner
- Participants will be able to collect data from disparate sources and learn to organize large data sets

WEEK 5

Inferential Statistics

- Introduction to hypotheses testing
- Performance of hypotheses tests
- Test for mean (one sample)
- Test for differences in means (two sample test)
- Test for differences in variances (F test)

Participants will be able to :-

- Setup hypotheses testing protocols
- Interpret hypothesis test results

WEEK 6

Linear Algebra

- Eigen values & Eigen vectors
- Singular Value Decomposition
- Understanding independence of variables
- Understanding relationships between variables

Participants will be able to :-

- Identify relationships between variables in large datasets
- Identify information sufficiency in terms of both equations and variables

WEEK 7

Optimization

- Basics of optimization - objective function, constraints, decision variables
- Types of optimization problems
- Statement of first order KKT necessary conditions
- Basic concepts in multi-objective optimization
- Introduction to optimization viewpoint in predictive modelling and machine learning

Participants will be introduced to :-

- Optimization problems
- Participants will be able to identify how most of the data analytics problems can be viewed as an optimization problem

WEEK 8 & 9

Predictive Modelling

- Correlation
- Basics of regression
- Ordinary least squares
- Model building
- Model assessment and improvement
- Diagnostics
- Multiple linear regression (model building & assessment)
- Random forest & Decision tree

Participants will be able to :-

- Build simple and multiple linear regression models
- Use regression models for prediction
- Assess quality of regression models using different criteria
- Select best predictors among several variables

WEEK 10

Machine Learning

- Classification
- Logistic regression
- K nearest neighbours
- Clustering
- K means

Participants will be able to :-

- Solve supervised & unsupervised problems using classification & clustering algorithms

WEEK 11

Machine Learning

- Dimensionality reduction methods
- Principal component analysis and its variants

Participants will be able to :-

- Solve classification & function approximation problems using advanced machine learning techniques

WEEK 12

Machine Learning

- Support vector machine
- Neural networks

Participants will be able to :-

- Solve classification & function approximation problems using advanced machine learning techniques

