

# Fundamentals of Artificial Intelligence (AI) and Machine Learning (ML)

## Getting Started

- Installing a Python Data Science Environment
- Using and understanding iPython (Jupyter) Notebooks
- Python basics: Part 1
- Understanding Python code
- Importing modules
- Python basics: Part 2
- Running Python scripts

## Statistics and Probability Refresher and Python Practice

- Types of data
- Mean, median, and mode
- Using mean, median, and mode in Python
- Standard deviation and variance
- Probability density function and probability mass function
- Types of data distributions
- Percentiles and moments

## Matplotlib and Advanced Probability Concepts

- A crash course in Matplotlib
- Covariance and correlation
- Conditional probability
- Bayes' theorem

## Algorithm Overview

- Data Prep
- Linear Algorithms
- Non-Linear Algorithms
- Ensembles

## Predictive Models

- Linear regression
- Polynomial regression

- Multivariate regression and predicting car prices
- Multi-level models

### **Applied Machine Learning with Python**

- Machine learning and train/test
- Using train/test to prevent overfitting of a polynomial regression
- Bayesian methods: Concepts
- Implementing a spam classifier with Naïve Bayes
- K-Means clustering

### **Recommender Systems**

- What are recommender systems?
- Item-based collaborative filtering
- How item-based collaborative filtering works?
- Finding movie similarities
- Improving the results of movie similarities
- Making movie recommendations to people
- Improving the recommendation results

### **More Applied Machine Learning Techniques**

- K-nearest neighbors - concepts
- Using KNN to predict a rating for a movie
- Dimensionality reduction and principal component analysis
- A PCA example with the Iris dataset
- Data warehousing overview
- Reinforcement learning

### **Dealing with Data in the Real World**

- Bias/variance trade-off
- K-fold cross-validation to avoid overfitting
- Data cleaning and normalization
- Cleaning web log data
- Normalizing numerical data
- Detecting outliers

### **Apache Spark: Machine Learning on Big Data**

- Installing Spark
- Spark introduction
- Spark and Resilient Distributed Datasets (RDD)
- Introducing MLlib
- Decision Trees in Spark with MLlib
- K-Means Clustering in Spark
- TF-IDF
- Searching wikipedia with Spark MLlib
- Using the Spark 2.0 DataFrame API for MLlib

### **Testing and Experimental Design**

- A/B testing concepts
- T-test and p-value
- Measuring t-statistics and p-values using Python
- Determining how long to run an experiment for
- A/B test gotchas

### **GUIs and REST**

- Build a UI for your Models
- Build a REST API for your Models