

Machine Learning Using Python

By Dr. Vishwanath Rao

Day 1

1. Python for Machine Learning

- In-Class Discussion
- Python Data Science-Centric Libraries
- NumPy
- NumPy Arrays
- Select NumPy Operations
- SciPy
- pandas
- Creating a pandas DataFrame
- Fetching and Sorting Data
- Scikit-learn
- Matplotlib
- Seaborn
- Python Dev Tools and REPLs
- IPython
- Jupyter
- Jupyter Operation Modes
- Jupyter Common Commands
- Anaconda
- Summary

2. Defining Data Science

- What is Data Science?
- Data Science, Machine Learning, AI?
- The Data-Related Roles
- The Data Science Ecosystem
- Tools of the Trade
- Who is a Data Scientist?
- Data Scientists at Work
- Examples of Data Science Projects
- An Example of a Data Product
- Applied Data Science at Google
- Data Science Gotchas
- Summary

3. Data Processing Phases

- Typical Data Processing Pipeline
- Data Discovery Phase
- Data Harvesting Phase
- Data Priming Phase
- Exploratory Data Analysis
- Model Planning Phase
- Model Building Phase
- Communicating the Results
- Production Roll-out
- Data Logistics and Data Governance
- Data Processing Workflow Engines
- Apache Airflow
- Data Lineage and Provenance
- Apache NiFi
- Summary

Day 2

4. Descriptive Statistics Computing Features in Python

- Descriptive Statistics
- Non-uniformity of a Probability Distribution
- Using NumPy for Calculating Descriptive Statistics Measures
- Finding Min and Max in NumPy
- Using pandas for Calculating Descriptive Statistics Measures
- Correlation
- Regression and Correlation
- Covariance
- Getting Pairwise Correlation and Covariance Measures
- Finding Min and Max in pandas DataFrame
- Summary

5. Repairing and Normalizing Data

- Repairing and Normalizing Data
- Dealing with the Missing Data
- Sample Data Set
- Getting Info on Null Data
- Dropping a Column
- Interpolating Missing Data in pandas
- Replacing the Missing Values with the Mean Value

- Scaling (Normalizing) the Data
- Data Preprocessing with scikit-learn
- Scaling with the scale() Function
- The MinMaxScaler Object
- Summary

6. Data Visualization in Python

- Data Visualization
- Data Visualization in Python
- Matplotlib
- Getting Started with matplotlib
- The matplotlib.pyplot.plot() Function
- The matplotlib.pyplot.bar() Function
- The matplotlib.pyplot.pie () Function
- Subplots
- Using the matplotlib.gridspec.GridSpec Object
- The matplotlib.pyplot.subplot() Function
- Figures
- Saving Figures to a File
- Seaborn
- Getting Started with seaborn
- Histograms and KDE
- Plotting Bivariate Distributions
- Scatter plots in seaborn
- Pair plots in seaborn
- Heatmaps
- ggplot
- Summary

Day 3

7. Data Science and ML Algorithms in scikit-learn

- In-Class Discussion
- Types of Machine Learning
- Terminology: Features and Observations
- Representing Observations
- Terminology: Labels
- Terminology: Continuous and Categorical Features

- Continuous Features
- Categorical Features
- Common Distance Metrics
- The Euclidean Distance
- What is a Model
- Supervised vs Unsupervised Machine Learning
- Supervised Machine Learning Algorithms
- Unsupervised Machine Learning Algorithms
- Choosing the Right Algorithm
- The scikit-learn Package
- scikit-learn Estimators, Models, and Predictors
- Model Evaluation
- The Error Rate
- Confusion Matrix
- The Binary Classification Confusion Matrix
- Multi-class Classification Confusion Matrix Example
- ROC Curve
- Example of an ROC Curve
- The AUC Metric
- Feature Engineering
- Scaling of the Features
- Feature Blending (Creating Synthetic Features)
- The 'One-Hot' Encoding Scheme
- Example of 'One-Hot' Encoding Scheme
- Bias-Variance (Underfitting vs Overfitting) Trade-off
- The Modeling Error Factors
- One Way to Visualize Bias and Variance
- Underfitting vs Overfitting Visualization
- Balancing Off the Bias-Variance Ratio
- Regularization in scikit-learn
- Regularization, Take Two
- Dimensionality Reduction
- PCA and isomap
- The Advantages of Dimensionality Reduction
- The LIBSVM format
- Life-cycles of Machine Learning Development
- Data Splitting into Training and Test Datasets
- ML Model Tuning Visually
- Data Splitting in scikit-learn
- Cross-Validation Technique
- Hands-on Exercise

- Classification (Supervised ML) Examples
- Classifying with k-Nearest Neighbors
- k-Nearest Neighbors Algorithm
- k-Nearest Neighbors Algorithm
- Hands-on Exercise
- Regression Analysis
- Regression vs Correlation
- Regression vs Classification
- Simple Linear Regression Model
- Linear Regression Illustration
- Least-Squares Method (LSM)
- Gradient Descent Optimization
- Multiple Regression Analysis
- Evaluating Regression Model Accuracy
- The R²

Day 4

- Model Score
- The MSE Model Score
- Logistic Regression (Logit)
- Interpreting Linear Logistic Regression Results
- Decision Trees
- Decision Tree Terminology
- Properties of Decision Trees
- Decision Tree Classification in the Context of Information Theory
- The Simplified Decision Tree Algorithm
- Using Decision Trees
- Random Forests
- Hands-On Exercise
- Hands-on Exercise
- Support Vector Machines (SVMs)
- Naive Bayes Classifier (SL)
- Naive Bayesian Probabilistic Model in a Nutshell
- Bayes Formula
- Classification of Documents with Naive Bayes
- Unsupervised Learning Type: Clustering
- Clustering Examples
- k-Means Clustering (UL)

- k-Means Clustering in a Nutshell
- k-Means Characteristics
- Global vs Local Minimum Explained
- Hands-On Exercise
- XGBoost
- Gradient Boosting
- Hands-On Exercise
- A Better Algorithm or More Data?
- Summary

8. AI Systems and Platforms Overview

- Heuristics and Expert Systems
- What is AI?
- AI, Machine Learning (ML), and Deep Learning
- Neural Networks in AI
- Deep Learning Neural Networks
- TensorFlow
- Keras
- Colab Notebooks
- PyTorch
- ML at Scale: Python on Spark - PySpark
- AWS IoT Service
- AWS ML Services
- SageMaker
- DeepLens
- The DeepLens Device (an IoT Device)
- A DeepLens Use Case
- Rekognition
- Rekognition's Object and Scene Detection Demo
- AWS ML Algorithm Marketplace

Day 5

9. Text Mining and NLP Overview

- What is Text Mining?
- The Common Text Mining Tasks
- What is Natural Language Processing (NLP)?
- Some of the NLP Use Cases
- Machine Learning in Text Mining and NLP

- Machine Learning in NLP
- TF-IDF
- The Feature Hashing Trick
- Stemming
- Example of Stemming
- Stop Words
- Popular Text Mining and NLP Libraries and Packages
- Google Natural Language Cloud Service
- Trying it Out
- How Google NL Service Works
- Google Translate Service
- Comprehend
- How Comprehend Works
- Comprehend in the AWS Management Console
- Use Cases for Comprehend
- Lex
- Polly
- Polly's Text-to-Speech Dashboard
- Example of Using Polly's AWS CLI
- Transcribe
- Translate