Machine Learning in Finance (SF)

Version 2019-05

|  |  |
| --- | --- |
| 09:00 - 10:30 | Mentimeter   * What do you want to learn today? * How familiar are you with Python? * Have you work with AI / machine learning before?   Setting the scene and PowerPoint slides to cover   * Types of Analytics * What is Data analysis * What is Data Science * Artificial Intelligence, Machine learning   + Supervised, unsupervised * Use cases * ML Algorithms |
| 10:30 - 11:00 | Break |
| 11:00 - 12:30 | Practical   * Lesson 1 - Getting Started.ipynb * Lesson 2 - Data Preprocessing.ipynb * Lesson 3 - Scikit Learn API.ipynb * Lesson 4 - EDA.ipynb * Lesson 5 - Linear Regression.ipynb * Lesson 6 - Variance-Bias Trade Off.ipynb * Lesson 7 - Cross Validation.ipynb * Lesson 8 - Classification (Logistic Regression).ipynb * Lesson 9 - Classification.ipynb * Lesson 10 - Clustering.ipynb   Work with the following datasets:  Regression: housing.data, RetailMart.xlsx  Classification: titanic\_train.csv |
| 12:30 - 13:30 | Lunch |
| 13:30 - 15:00 | Applications:   1. Supervised (classification) – churn, credit card, credit score, fraud detection 2. Unsupervised - segmentation   Datasets:   1. **Churn (Attrition): WA\_Fn-UseC\_-HR-Employee-Attrition.csv** 2. **Market Segmentation: CC GENERAL.csv** 3. Credit Card Default: default of credit card clients.xls 4. Credit Score: mini-lending-club-data.xlsx 5. Fraud Detection: creditcard.xlsx |
| 15:00 - 15:30 | Break |
| 15:30 - 17:00 |  |

Lesson 1 - Getting Started.ipynb

1. Anaconda
2. IDE. Jupyter Notebook
3. Input / output
4. Python Scientific Stack – Pandas, Seaborn, Numpy
5. EDA
6. Visualisation

Lesson 2 - Data Preprocessing.ipynb

1. Motivational Example
2. Various Pre-processing Methods
   1. Standardisation / Mean removal
   2. Min-Max
   3. Normalisation
   4. Binarisation
3. Encoding Categorical Features
   1. LabelEncoder
   2. One Hot Encoder

Lesson 3 - Scikit Learn API.ipynb

1. Introduction
2. General Principles
3. Basic Steps of Using Scikit-Learn API
4. Simple example
5. Modelling with Scikit Learn
6. Visualise

Lesson 4 - EDA.ipynb

1. Data
2. Data Type
3. Dimension / Shape
4. Descriptive Analytics
5. Correlation Analysis & Feature Selection

Lesson 5 - Linear Regression.ipynb

1. Introduction
2. Cross Validation Metrics
3. K-fold
4. Stratified k-fold
5. Pipeline
6. Model Performance Metrics

Lesson 6 - Variance-Bias Trade Off.ipynb

1. Validation curve
2. Learning curve

Lesson 7 - Cross Validation.ipynb

1. Holdout method
2. Holdout with validation
3. K-Fold CV
4. Stratified K-Fold CV

Lesson 8 - Classification (Logistic Regression).ipynb

1. Logistic Regression
2. Scikit Learn

Lesson 9 - Classification.ipynb

1. Introduction
2. MNIST
3. Visualisation
4. Split Train and Test
5. Training a binary classifier
6. Prediction
7. Performance Measures
   1. Stratified K-fold
   2. cross\_val\_score
8. Danger of blindly applying evaluator
9. Confusion Matrix
   1. Precision
   2. Recall
   3. F1 score
   4. Precision/Recall Tradeoff
   5. The ROC Curve
10. Model Comparison

Lesson 10 - Clustering.ipynb

1. Introduction
2. Types
3. Example from MLXTEND
4. Ward’s Agglomerative Hierachical Clustering
   1. Dendrogram
   2. Truncating Dendrogram
   3. Retrieve the clusters
   4. By distance
   5. By clusters
5. K-means clustering
   1. Elbow method
   2. Silhouette analysis
6. Mean shift