Course Outline

Machine Learning Essentials

Course Summary

Description

This course introduces popular Machine Learning techniques.

This course is intended for data scientists and software engineers.
We assume no previous knowledge of Machine Learning.
We teach popular Machine Learning algorithms from scratch.

For each machine learning concept, we first discuss the foundations, its applicability and limitations. Then we explain the implementation and use, and specific use cases

Please note that this course does not cover in-depth coverage of Math / Stats is behind Machine Learning.

Topics

- Machine Learning (ML) Overview
- Machine Learning Environment
- Machine Learning Concepts
- Feature Engineering (FE)
- Linear regression
- Logistic Regression
- Classification: SVM (Supervised Vector Machines)
- Classification: Decision Trees & Random Forests
- Classification: Naive Bayes
- Clustering (K-Means)
- Principal Component Analysis (PCA)
- Recommendation (Collaborative filtering)
- Final workshop (time permitting)

Audience

This class is designed for Data Scientists and Software Engineers**.**

Prerequisites

- Working knowledge of either R, Python or Apache Spark
- Programming background
- No previous machine learning knowledge is assumed

Duration

Three Days

Course Outline

Machine Learning Essentials

Course Outline

I. Machine Learning (ML) Overview
A. Machine Learning landscape
B. Machine Learning applications
C. Understanding ML algorithms & models (supervised and unsupervised)

II. Machine Learning Environment
A. Introduction to Jupyter notebooks / R-Studio
B. Lab: Getting familiar with ML environment

III. Machine Learning Concepts
A. Statistics Primer
B. Covariance, Correlation, Covariance Matrix
C. Errors, Residuals
D. Overfitting / Underfitting
E. Cross validation, bootstrapping
F. Confusion Matrix
G. ROC curve, Area Under Curve (AUC)
H. Lab: Basic stats

IV. Feature Engineering (FE)
A. Preparing data for ML
B. Extracting features, enhancing data
C. Data cleanup
D. Visualizing Data
E. Lab: data cleanup
F. Lab: visualizing data

V. Linear Regression
A. Simple Linear Regression
B. Multiple Linear Regression
C. Running LR
D. Evaluating LR model performance
E. Lab
F. Use case: House price estimates

VI. Logistic Regression
A. Understanding Logistic Regression
B. Calculating Logistic Regression
C. Evaluating model performance
D. Lab
E. Use case: credit card application, college admission

VII. Classification: SVM (Supervised Vector Machines)
A. SVM concepts and theory
B. SVM with kernel
C. Lab
D. Use case: Customer churn data

VIII. Classification: Decision Trees & Random Forests
A. Theory behind trees
B. Classification and Regression Trees (CART)
C. Random Forest concepts
D. Labs
E. Use case: predicting loan defaults, estimating election contributions

IX. Classification: Naive Bayes
A. Theory behind Naive Bayes
B. Running NB algorithm
C. Evaluating NB model
D. Lab
E. Use case: spam filtering

X. Clustering (K-Means)
A. Theory behind K-Means
B. Running K-Means algorithm
C. Estimating the performance
D. Lab
E. Use case: grouping cars data, grouping shopping data

Machine Learning Essentials

Course Outline (cont.)

XI. Principal Component Analysis (PCA)

A. Understanding PCA concepts
B. PCA applications
C. Running a PCA algorithm
D. Evaluating results
E. Lab
F. Use case: analyzing retail shopping data

XII. Recommendation (Collaborative filtering)

A. Recommender systems overview
B. Collaborative Filtering concepts
C. Lab
D. Use case: movie recommendations, music recommendations

XIII. Final workshop (time permitting)

A. Students will analyze a couple of datasets and run ML algorithms. This is done as a group exercise. Each group will present their findings to the class.