Fixel Algorithms - Royi Avital & Or Yair

Machine Learning - Advanced Methods

2024 (v0.1.000)

Motivation

Some problems can not be solved using basic Classification / Regression:

- **Forecasting** the demand of a product.
- **Ranking** the options given to users by the most likely to be chosen to the least.
- Using the **probability measure of a classifier** for decision making.
- Adjusting the classification loss function to handle **imbalanced data** and real world problems.
- **Feature Engineering** beyond the correlation with the target.
- Optimizing hyper parameters with a limited **time budget**.

There are many real world problems not covered in regular Machine Learning courses.
This course is built as a second course in Machine Learning to present practical and advanced methods in Machine Learning.

Overview

This is a 8 days (64 Hours) course which introduces them to advanced methods, algorithms, tools, frameworks, concepts and strategies in Classic Machine Learning.
The course:

- Covers advanced concepts in: Classification, Regression, Feature Engineering and Hyper Parameter Optimization.
- Introduces the concept of supervised learning: Ranking.
- Introduces Time Series and the task of Forecasting.
- Demonstrates through Hand On exercises how to formulate and solve real world problems.
- The course is accompanied by more than 30 notebooks given in a dedicated GitHub repository.

Main Topics

- Classification: Calibration, Custom Loss Function, Cost Sensitive.
- Gaussian Processes for Regression.
- Feature Engineering: Beyond the cross correlation, Auto ML.
- Hyper Parameter Optimization: Bayesian methods.
- Supervised Learning: Ranking.
- Time Series: Feature engineering, Forecasting.
- Interpretability / Explainability for Models.

Goals

- The participants will be able to match the proper approach to a given problem.
- The participants will be able to implement, adjust, fine-tune and benchmark the chosen method.
- The participants will be able to build a pipeline with Auto ML, Hyper Parameter optimization and explainability module.
- The participants will be able to calibrate a classifier.
- The participants will be able to create a custom loss function for a classifier.
- The participants will be able to train and use a Forecasting / Ranking model.
- The participants will be able to train and use a Gaussian Process based model.

Audience

Experienced developers who use Machine Learning: Algorithm Engineers, Data Engineers, Data Scientists.

Prerequisites

- Mathematical: Linear Algebra (Basic), Probability (Basic), Statistics (Basic).
- Machine Learning: Classification, Regression.

The course will be using Python as the programming language.

The Machine Learning - Advanced Methods course page holds in depth information and up to date syllabus.

Syllabus

Day I: Advanced Classification

1. Course Overview
    - Motivation.
    - Agenda.
2. Classification Recap: The Geometric Problem, The Decision Function, Probabilistic Classification, Score vs. Loss.
3. Classifier Recap: Linear Classifier, SVM, Kernel SVM, Logistic Regression, Decision Tree.
4. Ensemble Methods: Stacking, Random Forest, AdaBoost, Gradient Boosting, XGBoost, LightGBM.
5. Ordinal Classification: Use Case, Limitations of the Classifier, Definition, Loss Function.
6. Workbooks: Ensemble Methods, Ordinal Classification.

Day II: Advanced Classification

7. Cost Sensitive Classifier: Imbalanced Data & Scores, Loss Matrix Model, Cost Sensitive Classifier, Weights Adjustment.
8. Custom Loss Function: Test Case, Implementation in Python, Using XGBoost.
9. Classifier Calibration: Decision based on probability, Uncalibrated model, Calibration process.
10. Workbooks: Loss Sensitive Classifier, Custom Classification Loss 001, Custom Classification Loss 002, Classifier Calibration.

Day III: Advanced Regression & Gaussian Processes

11. Random Process Recap: Probability, The Gaussian Distribution, Random Process, Stationarity, Auto Correlation Function.
12. Local Regressors: Kernel Regression, Weighted Local Kernel Regression.
13. Gaussian Process: The Model, The Parameters, Fitting, Gaussian Process Regressor, Gaussian Process Classifier.
14. Isotonic Regression: Use Cases, The Model, The Loss Function, Fitting.
15. Workbooks: Local Regression, GPR, GPC, Isotonic Regression.

Day IV: Feature Engineering

16. Challenges: Why correlation fails, Discrete Features, Issues with One Hot Encoding, Ordinal Features, Cyclic Features.
17. Feature Transforms: Cyclic Features, Cyclic Objectives, Unsupervised Methods & LDA.
18. Feature Imputer: Statistics Based, Feature Based, Model Based.
19. Auto ML: Concept, Frameworks, Pipeline.
20. Feature Selection: Univariate Methods, Multivariate Methods, Sparsity, Predictive Score.
21. Workbooks: Feature Engineering 001, Feature Engineering 002, Regression with Cyclic Objective, Feature Selection 001, Feature Selection 002.

Day V: Hyper Parameter Optimization

22. Grid Methods: Cross Validation, Uniform Search, Random Search.
23. Bayesian Methods: Concept, Conversion from Discrete to Pseudo Smooth, Optimization.
24. Workbooks: Hyper Parameter Optimization 001, Hyper Parameter Optimization 002, Hyper Parameter Optimization 003.

Day VI: Ranking

25. Motivation: Use Cases, Ranking vs. Classification.
26. Ranking Models: Pointwise, Pairwise, Listwise.
27. Learning to Rank: Data Preparation, XGBoost, LightGBM.
28. Workbooks: Ranking Basics, Product Recommendation (Recommendation System).

Day VII: Time Series & Forecasting

29. Time Series: Stationarity, Seasonality, Noise, Forecasting Loss.
30. Generation Models: MA, AR, ARMA, ARIMA, SARIMA.
31. Concepts: Differencing, Exponential Smoothing.
32. Time Series Forecasting: Statistical Models, Kalman Filter, Learning Models.
33. Time Series Regression
    - Feature Engineering: Auto, Rocket.
    - Models.
34. Workbooks: Forecasting with Stat Models, Forecasting with Learning Model, Forecasting with Regression (Feature Engineering).

Day VIII: Explainability & Interpretability

35. Motivation: Explain decision, Analysis of results, Investigation of results.
36. Models: LIME Model, SHAP Model.
37. Pipeline: Integration into a Pipeline.
38. Workbooks: LIME, SHAP.

Remarks

- Some of the notebooks / workbooks are guided exercises.
- The days are a crude partitioning. In practice some subjects will take more than a day and some less.