Queens College, Data Science/Machine Learning Club Syllabus

Bryan Lliguicota, Muhammad

Fall 2019

E-mail: qc.datascienceclub@gmail.com Web: https://datascience-ai-qc.github.io/

Club Objectives

Successful Members will have a grasp of:

1. Pandas
2. Data Visualization(Matplotlib, seaborn)
3. Machine Learning(SciKit-Learn)
4. Data Manipulation
5. Hyper-parameter optimization
6. An understanding of basic to intermediate Machine Learning Algorithms and Data Science techniques

Club Structure

Club Structure

The material will be broken down into three modules. The first consisting of data manipulation with pandas. The second module will focus on data visualization and the third will focus on machine learning techniques, how to properly tackle a problem and optimize parameters to minimize your loss function.

Aside from these learning objectives, we want to make the club as interactive as possible. We would like to bring in speakers, have kaggle competitions and once a month host tutorials on a real life problems.

Members are encouraged to bring forth there ideas on how to improve the club, or if they would like to present a topic they are more then welcome to.

Schedule and weekly learning goals

The schedule is tentative and subject to change. The learning goals below should be viewed as the key concepts you should grasp after each week, and also as a study guide before each meeting, and at the end of the semester. The lectures are not independent and therefore if you miss a meeting; we encourage you to reference the club github page: **https://github.com/datascience-ai-qc** for the meeting notes.

Week 01, 09/09 - 09/13: Getting Started

- Introductions and resource sharing
- Setting up environments and necessary packages

  Github, Anaconda, reference book, ...

Week 02, 09/16 - 09/20: Pandas

- Data Indexing and Selection
- Operations on Data in Pandas
- Handling Missing Data
- Hierarchical Indexing

Week 03, 09/23 - 09/27: Pandas

- Concat and Append
- Merge and Join
- Aggregation and Grouping

Week 04, 09/30 - 10/04: Pandas

- Pivot Tables
- Vectorized String Operations
- Working with Time Series
- High Performance Pandas

Week 05, 10/07 - 10/11: Data Visualization

- Simple Line and Scatter plots
- Visualization Errors
- Density and Contour Plots

Week 06, 10/14 - 10/18: Data Visualization

- Histogram, Binning and Density
- Customizing Plot Legends and Colorbars
- Multiple Subplots
- Text and Annotation

Week 07, 10/21 - 10/25: Data Visualization

- Customizing Ticks
- Three-Dimensional Plots
- Geographical Data
- Visualization with Seaborn

Week 08, 10/28 - 11/01: Machine Learning

- What Is Machine Learning?
- Introduction to SciKit-Learn

Week 09, 11/04 - 11/08: Machine Learning

- Hyper-parameters and Model Validation
- Feature Engineering

Week 10, 11/11 - 11/15: Machine Learning

- In-Depth Linear Regression
- In-Depth Naive Bayes Classifier

Week 11, 11/18 - 11/22: Machine Learning

- Continuation of Last Week
- In-Depth Support Vector Machine

Week 12, 11/25 - 11/29: Machine Learning

- In-Depth Decision Trees and Random Forests

Week 13, 12/02 - 12/06: Machine Learning

- Continuation of last week
- In-Depth Principle component analysis
- In-Depth Manifold Leaning

Week 14, 12/09 - 12/13: Machine Learning

- In-Depth K-mean Clustering
- In-Depth Gaussian Mixture Models
- In-Depth Kernel Density Estimator