**Boston University MET College** **2024**

**Department of Computer Science**

**Data Science with Python (CS 677)**

****

**Syllabus**

**Time & Location**: Tue 6 PM - 8:45 PM, KCB 102 (565 Commonwealth Avenue)

**Instructor**: Avi Mohan (avimohan@bu.edu)

**Syllabus** (tentative)

The course is divided into 7 modules that will each last roughly 2 weeks.

|  |  |  |
| --- | --- | --- |
| Module | Topics | Readings |
| 1 | Introduction to machine learning. Review of matrix algebra, NumPy, and MatplotLib | [M12] Chapters 2, 3 and 4, Class notes (based on [NFI21] Chapters 1, 4, 5). |
| 2 | Review of statistics, Seaborn, and Pandas | [M12] 5, 9, Class notes (based on [NFI21] Chapter 3, 5). |
| 3 | Linear and polynomial regression. Regularization and Gradient Descent. Basics of feature engineering. | Class notes (based on [NFI21] chapter 9) |
| 4 | Logistic Regression and K nearest neighbors | Class notes (based on [NFI21] chapters 9, 10) |
| 5 | Random Trees, bagging and random forests | Class notes (based on [J24] chapter 8) |
| 6 | Support Vector Machines and Kernel SVMs | Class notes (based on [J24] chapter 9) |
| 7 | Advanced topics | Class notes. |

**Required Textbook**:

[M12] McKinney, Wes. ***Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*.** O'Reilly Media, Inc., 2012.

ISBN-13: 978-1098104030

**Additional Material**

1. [BBG20] Bruce, Peter, Andrew Bruce, and Peter Gedeck. Practical statistics for data scientists: 50+ essential concepts using R and Python. O'Reilly Media, 2020.

ISBN**-** 13: 978-1492072942

1. [J24] James, Gareth, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor. **An Introduction to Statistical Learning: With Applications in Python**. Springer Nature, 2024.

ISBN**-** 13: 978-3031387463

1. [NFI21] Navlani, Avinash, Armando Fandango, and Ivan Idris. ***Python Data Analysis: Perform data collection, data processing, wrangling, visualization, and model building using Python***. Packt Publishing Ltd, 2021.

ISBN**-** 13: 978-1787127487