MMF1922H: Data Science for Investment and Finance

Contents

1 Administrative

Course Coordinator: Nicholas Hoell
E-mail: nicholashoell@gmail.com
Office: PG200A Phone: 416-946-3771

Class Time: F: 6:00-9:00PM Class Location: MMF
Office Hours: By Appointment

2 Content of this Course

This is an introduction to mathematical and statistical methods in industrial applications., focussing on extracting information from large data sets. We cover techniques and principles of classical methods in statistics such as parameter estimation as well as more modern methods from machine learning which are becoming more ubiquitous in finance.

3 Grading

Students will receive a letter grade based on completion of homework assignments (some involving programming in Python), quizzes and 2 course projects. The projects are 50% of the course grade. Homework and Quizzes are 25% apiece.

4 Teaching Assistant

Your TA this term is Jonathan Mostovoy, jonathan.mostovoy@mail.utoronto.ca, who will hold some office hours throughout the semester. Details on time will be announced later in the term.

5 Textbook

No single textbook exists which fully meets the needs of a course like this so my suggestion is to stay close to the lectures and follow the material closely in the references. Beyond that, when you encounter an interesting idea you wish to study more, you will find a lot of help in the listed optional texts. As well, I have an ongoing (and, sadly, *incomplete*) course notes document, available at http://www.math.toronto.edu/nhoell/MMF1922/ds_course_notes.pdf which I am writing. It's a work in progress (and will only cover the first half of the course) but worth looking at.

5.1 Recommended Texts

We will be using material from the following books:

1. Taniguchi, Hirukawa, & Tamaki's *"Optimal Statistical Inference in Financial Engineering"* published by Chapman & Hall/CRC, ISBN: 1-58488-591-2. This is a very good collection of classical topics useful for finance. I've placed an order for this in the University of Toronto bookstore, though availability in Canada seems to be an issue.

2. Kaipio & Somersalo's *"Statistical and Computational Inverse Problems"* published by Springer, ISBN: 978-0-387-27132-3. This is a great book which focuses on applications important in biomedical research. It is in the traditional theorem/proof style you should be familiar with.

3. Christopher Bishop's *"Pattern Recognition and Machine Learning"* published by Springer, ISBN: 978-0-387-31073-2. This is a very readable, excellent textbook which covers a lot more material than we'll need but which will serve you well in the future. It's mathematical but written more in the style of statistical sciences.

4. Hastie, Tibshirani & Friedman's *"The Elements of Statistical Learning"*, published by Springer ISBN: 978-0-387-84587-0. This covers a lot of material which goes beyond what we can cover but is an excellent reference.

5. Hoel, Port, & Stone's *"Introduction to Statistical Theory"* published by Houghton Mifflin, ISBN: 0-395-04637-8. This is a lucid elementary presentation of some of the important classical topics we'll cover.

6. Goodfellow, Bengio, & Courville's *"Deep Learning"* published by the MIT Press, ISBN: 978-0-262-03561-3. This is a very modern, dense text which explains the current art of the field. We will not get through much of it, but it is an excellent overview of the field.

6 The Menu

We will keep, *very roughly*, to the following schedule.

- Weeks 1 & 2 . . . . . .
  - **Topics:** Overview of data science and the course. The Blob Factory. Probability, random variables, expected value, conditional probability, Bayes theorem. Entropy and information. Common distributions, characteristic functions and limit theorems. Risk functions and Bayes decision function.
- Weeks 3 & 4 . . . . . .
  - **Topics:** Estimators: efficiency, asymptotic efficiency,. The Cramer-Rao bound. Sufficient statistics and factorization. Exponential families. Maximum likelihood and method of moments. Hypothesis testing.
- Week 5 . . . . . .
  - **Topics:** Regression, classification and regularization. Gradient methods.
- Week 6 . . . . . .

- **Topics:** The curse of dimensionality. Dimension reduction with PCA, MDS, tSNE, and parametric tSNE. Introduction to learning methods via K-means and logistic regression.

- Weeks 7 & 8 . . . . . .

  - **Topics:** Neural nets and backpropagation. Regularization and introduction to deep learning and modern approaches.