

ECE 697DA: Data Analytics

Fall 2017

Syllabus

- For up-to-date information about office hours, homework, test coverage, class announcements, and other administrative matters, please see the course moodle page.
- Texts:
 - Charu Aggarwal, *Data Mining: The Textbook*, Springer-Verlag, 2015.
 - J. Leskovic, A. Rajaraman, J. Ullman, *Mining of Massive Datasets*, Cambridge University Press, 2014;
<http://infolab.stanford.edu/~ullman/mmds/book.pdf>
 - J. Hopcroft and R. Kannan, *Foundations of Data Science*, 2014;
<https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/book-No-Solutions-Aug-21-2014.pdf> .
 - S. Ross, *Simulation*, Academic Press, 2013.
 - G. Strang, *Linear Algebra and Its Applications*, any edition.

These books are referred to below as CA, LRU, HK, Ross, and Strang, respectively. CA and Strang have been placed on reserve in the library. The authors of LRU and HK appear to have made their book pdfs available online. The university library has Ross's book available online.

- Background: Probability theory (ideally ECE 603 or something close to it), algorithms (ideally ECE 665 or something close to it) and linear algebra (usually covered as part of an undergraduate course). We assume a level of mathematical maturity consistent with the graduate engineering level. Note that this is essentially an applied mathematics course. This course is probably best taken in the second year of a student's MS program.
- Grading:
 - Two in-term tests, 25% each.
 - Final examination, 35%
 - Homework, 15%.

All tests will be closed-book, closed-notes.

Detailed (Tentative) Coverage List

(Not necessarily in the order of coverage)

- Introduction
 - CA Chapter 1.
 - LRU: (Read by yourself: no lecture coverage): Section 1.3 for background material.
 - Issues associated with high-dimensional space: HK Section 1
- Principal Component Analysis and Singular Value Decomposition
 - LRU Sections 11.1 to 11.3. Students should read Section 11.3.4 by themselves.
 - Strang Sections 3.1 to 3.4.
 - HK Section 3.
- Principles of statistical inference
 - Central Limit Theorem: Ross, Section 2.9.
 - Confidence Intervals: Ross, Sections 8.1 and 8.2.
 - Bayesian Analysis: <http://www.uv.es/bernardo/BayesStat.pdf>
- Distance Measures
 - LRU Sections 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.8.
 - CA Sections 3.1, 3.2 (skip 3.2.1.7 onwards), 3.3, 3.4.
 - Applications: Document Analysis
- Pattern Analysis
 - CA Sections 4.4.2, 4.4.3, 5.4.
 - LRU Sections 6.1, 6.2, 6.3, 6.4.
- Clustering Techniques
 - CA Sections 6.1, 6.2.1.
 - LRU Sections 7.1, 7.2, 7.3, 7.4, 7.5.
- Outlier Analysis
 - CA Sections 8.2, 8.3, 8.5, 8.8.
 - Paper by Chandola, et al. (will be posted)
- Data Classification
 - CA Sections 10.1, 10.2, 10.3, 10.5
 - LRU Section 12.3
- Recommendation Systems
 - LRU Chapter 9
- Pagerank Algorithm
 - LRU Chapter 5