**ST 295: Topic Course: Machine Learning in Data Science, Spring 2022**

Annie Qu, Chancellor’s Professor

Donald Bren Hall 2212

E-mail: aqu2@uci.edu

**Reader:** Qi Xu <qxu6@uci.edu>

**Lecture information**: MW 11am-12:20pm, classroom: DBH 1300 **(in person)**

**Office Hours for Instructor**: TBA or by appointment via emails.

**Course overview:** This course will train students to analyze large-scale complex-structure data using advanced statistical learning methods and algorithms. The main topics in the course include: data exploration and interpretation in data science; large data processing; optimizations in statistical learning, regularization methods; deep learning, recommender systems, and NLP and text mining. Students will gain practical skills of data mining and knowledge discovery in various applications such as business, political Science, biology and medicine.

**Objective**: After taking this course, the student should gain working knowledge of using machine learning and data mining techniques in the field of data science applications.

**Examples of Data:** The following data will be discussed and analyzed in the class or students’ projects:

Presidential election data, IRi marketing data, MovieLens data, Millions Song Dataset (MSD), Last.fm data, Trip Advisor data, Yelp data, UCI Machine Learning Repository data, US stock market data, wearable device data, hospital triage data, social network data, microbiome data

**We will add more data during the course. Please provide your input regarding the data.**

**Schedule**

• Week 1 - Syllabus, data exploration and domain application samplers in data science

• Week 2 - Regularization methods

• Week 3 - Large data processing using Python and R

• Week 4/5 – Introduction of deep learning and optimization tools

• Week 6/7- Recommender systems and individualized modeling

• Week 8/9 – Introduction of text mining and NLP

• Week 10 - Student presentations based on final projects

• Week 11 - Final project due

**Final Project (You can be either a solo or a two-person team for the final project):**

The final project is similar to journal paper which has sections of abstract, introduction and background, proposed method, computational algorithms and implementation, comparisons to existing methods, numerical studies based on simulations and real data application, and conclusion. The final project requires students to get started at the beginning of the second half of quarter. Each student/team will give a 10 (if solo) or 15 (if two-person team) minutes presentation for their project. Students are expected to prepare power point slides, and provide visualization of the data, and convincing arguments why their method and algorithm are the best for the data they analyze.

**Sample projects (papers and slides) will be provided.**

**General guidelines of the final project report:**

(1) Summary of your paper

(2) Introduction: motivations of your work, existing literature, overall goals of your approach, and advantages of your methods

(3) Main methods and algorithms

(4) Numerical studies: your data description, data analysis, simulation results if you have one, and numerical comparisons of different methods.

(5) Conclusion

(6) References

Your report needs to be self-explanatory, that is, you need to explain clearly for your notations, methods and packages you used. For existing methods, you can provide general descriptions of these methods, and rational of using these methods.

**Your report must be written by you or your team, and no sentences or paragraphs are copied from other sources.**

You can put your coding in the Appendix, if your file is too large to send, you can send your report only, and provide an either link of your Appendix files in Github or create a Dropbox folder shared with me.

**Software used in class**: R, Matlab and Python

**Prerequisites**: ST200A, ST210, ST210B/ST211

**Workload**: homework (40% of your grade), a final written project (45%) and a final presentation (15%)

**Academic Integrity:**

All students are expected to abide by the UCI Academic Integrity Policy and Academic Honesty Policy (http://senate.uci.edu/academic-integrity/).

Infractions of these rules will not be tolerated. Infractions include, but are not limited to, copying any part of other students’ homework or allowing another student to copy any part of your homework. Final project must be written by you, and no sentences or paragraphs are copied from other sources.