CS 4786/5786: Machine Learning for Data Science
Spring 2020

News:

- *Placement exam can be found [here](). Its due on January 31st.*

Logistics:

- Course pre-requisite: Pass placement exam.
- Course grades:
  1. **Midterm: 25% of the grade**. On March 26th at URHG01, 7:30PM.
  2. **Finals: 25% of the grade**. May 16th, 2PM.
  3. **Assignments: 28% of the grade**. 4 Assignments worth 7% each. To be submitted via Vocareum.
  4. **Competition: 20% of the grade**. Will be hosted on In-class Kaggle
  5. **Survey: 2% of the grade**. Two surveys worth 1% each to get feedback on course.

Course Description

An introductory course in machine learning, with a focus on data modeling and related methods and learning algorithms for data sciences. Tentative topic list:

- *Dimensionality reduction*, such as principal component analysis (PCA) and the singular value decomposition (SVD), canonical correlation analysis (CCA), random projection, etc. (We expect to cover some, but probably not all, of these topics).
- *Clustering*, such as k-means, Gaussian mixture models, the expectation-maximization (EM) algorithm, link-based clustering. (We do not expect to cover hierarchical or spectral clustering.).
- *Probabilistic-modeling topics* such as graphical models, inference (e.g., belief propagation), parameter learning, Hidden Markov Model.
- *Socially responsible ML: Fairness in Machine Learning, Differential Privacy etc.*

Prerequisites: CS4780/5780 (Machine Learning for Intelligent Systems), probability theory (BTRY 3080, ECON 3130, MATH 4710, or strong performance in ENGRD 2700 or equivalent); linear algebra (MATH 2940 or equivalent); CS2110 or equivalent programming proficiency.