CSE 255
Data Mining and Predictive Analytics

Course outline
The course webpage is available here: http://cseweb.ucsd.edu/~jmcauley/cse255/

This page will include data, code, slides, homework and assignments
This course is in two parts:

1. **Methods** (lectures 1-4):
   - Regression
   - Classification
   - Unsupervised learning and dimensionality reduction
   - Graphical models and structured prediction

2. **Applications** (lectures 5-8):
   - Recommender systems,
   - Text mining
   - Social network analysis
   - Mining behavioral and sequence data
1. Today: supervised learning

- Linear regression and least-squares
- (a little bit of) feature design
- Overfitting and regularization
- Gradient descent
- Training, validation, and testing
2. Classification

• Logistic regression
• Support Vector Machines
• Multiclass and multilabel classification
3. Unsupervised learning

- Dimensionality reduction
- Principal component analysis
- Matrix factorization
- K-means
- Graph clustering and community detection
4. Graphical models

- Dealing with interdependent variables
- Labeling problems on graphs
- Hidden Markov Models and sequential data
• Latent factor models and matrix factorization (e.g. to predict star-ratings)
• Collaborative filtering (e.g. predicting and ranking likely purchases)
6. Applications – Text mining

- Sentiment analysis
- Bag-of-words representations
- TF-IDF
- Stopwords, stemming, and (maybe) topic models
7. Applications – Network analysis

• Measuring importance and influence of nodes (e.g. pagerank)
• Link prediction and recommendation (e.g. “people you may know”)

8. Temporal & sequence data

- Markov Models
- Sliding windows

- Wrap-up and (probably) some examples from academic literature
There is **no textbook** for this class

- I will give chapter references from *Bishop: Pattern Recognition and Machine Learning*
- I will also give references from Charles Elkan’s notes ([http://cseweb.ucsd.edu/~jmcauley/cse255/files/elkan_dm.pdf](http://cseweb.ucsd.edu/~jmcauley/cse255/files/elkan_dm.pdf))
Evaluation

• There will be **weekly** homework worth 5% each. Your **lowest grade** will be dropped, so that 8 homework assignments = 35%

• There will be two assignments
  1. On recommender systems (after lecture 5), worth 35%
  2. Open-ended, worth 30%
Evaluation

• Homework will be due at the beginning of the following lecture
• If you can’t attend the next week’s lecture drop off homework outside my office (CSE 4102) before the lecture
Questions?