This webpage is for an old version of the course; content may be out of date!

CSE 255: Dahta Mining and Predictive Analytics

Winter 2015, Mondays 18:30-21:20, CENTR 222

For the current version of this class, see here

CSE 255 is a graduate-level course devoted to current methods for data mining and predictive analytics. No previous background in machine learning is required, but all participants should be comfortable with programming (all example code will be in Python), and with basic optimization and linear algebra.

The course meets once a week on Monday evening, starting January 5. There will be no classes on January 19 (MLK day) or on February 16 (President's day). Meetings are in CENTR 222.

There is no textbook for the course, though chapter references will be provided from Pattern Recognition and Machine Learning (Bishop), and from Charles Elkan's 2013 course notes.

Office hours: I'll hold office hours on **Friday 9-11am** in CSE 4102. Additional office hours will be held by Dongcai Shen on **Mondays 10-12** in CSE 4127. For other discussions see the course's Piazza page.

Part 1: Methods

Week	Topics	Files	References	Slides	Homework
1 (Jan 5)	**Supervised Learning: Regression** • Least-squares regression • Overfitting & regularization • Training, validation, and testing	50k beer reviews lecture1.py	Bishop ch.3 Elkan ch.3,6	introduction course outline lecture 1 case study: reddit	Homework 1 **due January 12**
2 (Jan 12)	**Supervised Learning: Classification** • Logistic regression • SVMs • Multiclass & multilabel classification • How to evaluate classifiers	50k book descriptions 5k book cover images lecture2.py homework2.py	Bishop ch.4 Elkan ch.5,8	lecture 2	Homework 2 Homework 3 **both due January 26**
3 (Jan 26)	**Dimensionality Reduction & Clustering** • Singular value decomposition & PCA • K-means & hierarchical clustering • Community detection	facebook ego network lecture3.py	Bishop ch.9 Elkan ch.13	lecture 3 assignment 1 case study: social circes	Homework 4 **due February 2** Assignment 1 **due February 23** reports
	Graphical Models & Interdependent Variables			lecture 4	Homework

4 (Feb 2)	• Directed and undirected models • Labeling via graph-cuts		Bishop ch.8	case study: image labeling	5 **due February 9**

Part 2: Applications

Week	Topics	Files	References	Slides	Homework
5 (Feb 9)	**Recommender Systems** • Latent-factor models • Collaborative filtering	homework 6/7 data assignment 2 data baselines.py	Elkan ch.11	lecture 5 assignment 2 case study: beer experts	Homework 6 Homework 7 **both due February 23** (or morning of February 25 outside 4102) Assignment 2 **due March 10**
6 (Feb 23)	**Text Mining** • Sentiment analysis • Bags-of-words • TFIDF • Stopwords, stemming, and topic models	lecture6.py	Elkan ch.12	lecture 6 case study: text and opinions	Homework 8 **due March 2**
7 (Mar 2)	**Network Analysis** • Power-laws and small-worlds • Random graph models • triads and weak ties • HITS and PageRank		Elkan ch.14 Easley & Kleinberg	lecture 7 case study: rich-clubs	Homework 9 **due March 9**
8 (Mar 9)	**Modeling Temporal and Sequence Data** • Sliding windows and autoregression • Hidden Markov Models • Temporal dynamics in recommender systems • Temporal dynamics in text and social networks	lecture8.py		lecture 8	no homework!