2018

Advanced Data Analysis

Jun Li

Welcome!

- Subject contents and preparation
- What's to expect
- Assignments

Learning DA

- Theory and the state-of-the-art techniques — we want to have the cake and eat it!
- In class, we mostly learn theory
- In tutorials, we practice the learned theory and skills
- Preview learning will be provided to enable you get hands dirty in the learning (mostly labs)
- Guest lectures about the bleeding-edge techniques (they are likely to change *before* you reach employment market! So keep a cool mind and sharp your basic skills to *keep* learning!)

6 Basic Lectures

- Overall view on learning from data and linear “Perceptron”
- Neural Networks
- Measuring information — revisiting decision trees
- Resampling and bagging methods
- Data representation, unsupervised learning
- (?) Bayes school of learning / thinking

Tool Choice: Python + Packages

- We need to do some programming, yes, keyboard-striking programming, not clicking-and-drawing programming
- Keep cool — the language is
  - i) natural, say, when you want x to be 1 if y > 3, otherwise x to be 2, you write something like: “x=1 if y>3 else 2”
  - ii) widely used: as much used as javascript / R / java / C. I.e. it is mainstream and unlikely to become obsolete in your career time
  - iii) it talks to almost all other languages
  - iv) the de facto standard language of cutting-edge AI framework (DeepMind, Google’s tensorflow, Facebook’s pytorch, Caffe) — yes you can write AlphaGo in Python

Tool Choice: Cont

- You can also have your own preference.
- **WE DO NOT REQUIRE ASSIGNMENTS TO BE DONE IN ANY PARTICULAR TOOL. THE CRITERIA IS FUNCTIONAL NOT IMPLEMENTATION DETIAL.**
- Once the algorithm is learned, all modern computation software should allow to implement. E.g. to take the mean value from a data set X:
  - X -= X.mean() # in Python
  - X = X-mean(X(:)) % in Matlab
  - X <- X-mean(X) # in R

Wt	UG		PG		Wt	Due
30%		Literature Understanding — **Review**			30%	W5
40%	Movie	Practical Data Analytics Project — **Report**			30%	W10
30%		Exam: An Open Problem — **Report**				W11
			Pitch and Plan — **Report**	Movie	40%	

UG-1 PG-1

Assignment (30%)

- Understand the literature
- Individual work
- 10 page report - your own critic comments
- Week 1 Out / Week 5 In

UG-2

Assignment (40%)

- Develop an algorithm or apply DA to a dataset of your choice
- Individual work or in pairs
- 20 page report + presentation (recorded in Movie)
- Choose something that you are passionate about
- Week 5 Out / Week 10 In / Selected presentations will be played in week 11.

UG-3

Final Exam (30%)

- Take home!
- Open questions
- Week 10 - Out / Week 11 - In
- Policy and Criteria

PG-2

Assignment (30%)

- Practical data analytics project derived from your workplace (or related industry)
- Individual work or in pairs
- 20-30 page report
- Choose something that you are passionate about
- Week 5 Out / Week 10 In

PG-3

Pitch and Plan (40%)

- 10 Page plane for a data analytics project plus an "elevator pitch" (recorded in movie) — to sell the idea to a potential investor
- Individual work
- Choose something that you are passionate about — and people will be interested in
- Week 11 In