

PS5841

Data Science in Finance & Insurance

Front Matter

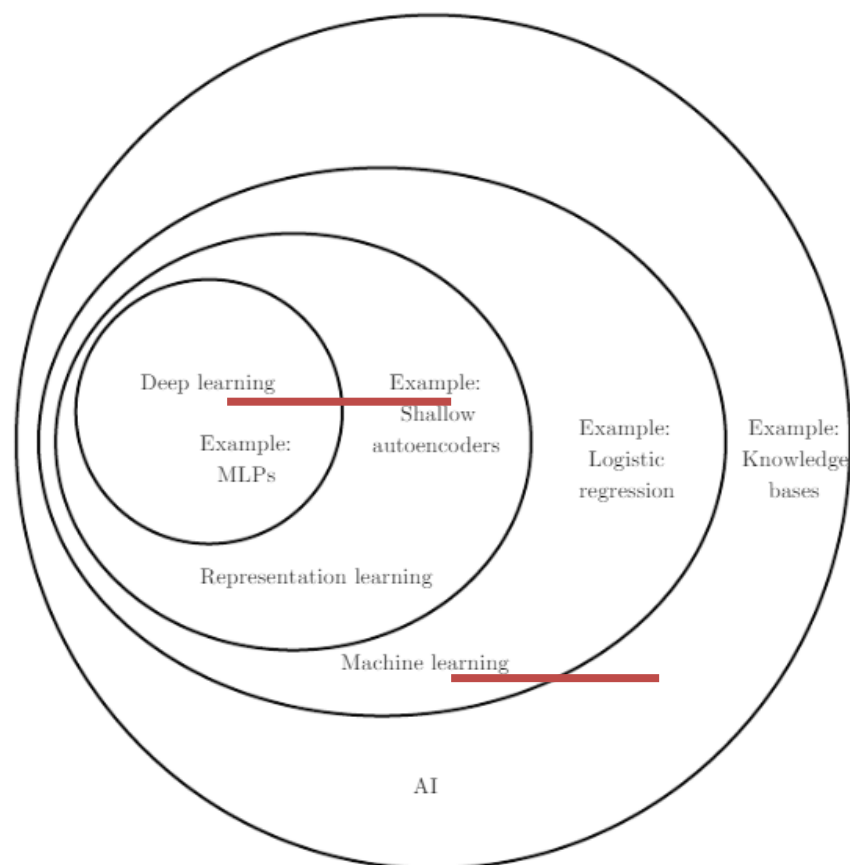
Yubo Wang

Spring 2022

Data Science

- Data Science Techniques
 - Extract info from data
 - Produce inputs for decision making
- Trendy labels
 - Machine learning
 - Deep learning
 - Artificial intelligence

ML & SL



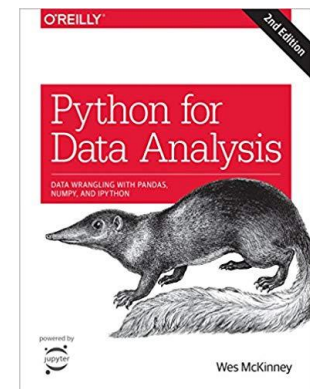
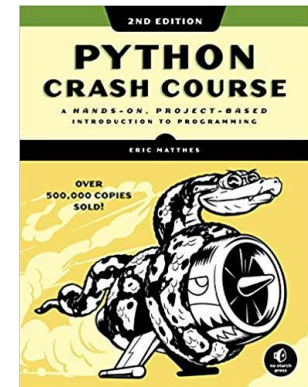
- 1800s linear regression
- 1930s LDA
- 1940s logistic regression
- 1970s GLM
- 1980s trees, GAM, NN
- 1990s SVM
- ...

This Course

- Course Goals
 - Coding & Algo
 - ML/SL Models
 - Portable Skills
 - Review & Highlights (overlap)
- Helpful preparations
 - Probability & Statistics, Calculus, Linear Algebra

Reference Materials - Coding

- “Official” Python Tutorial
- Matthes, *Python Crash Course, 2nd ed*, No Starch Press.
- McKinney, *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython, 2nd ed.*, O’Reilly Media.



Computing Environment

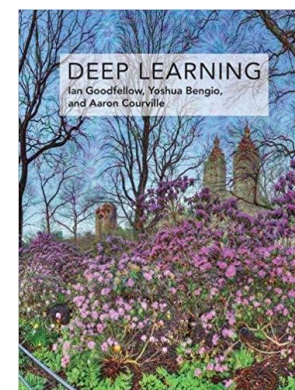
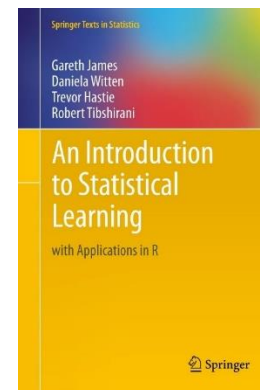
- Tools
 - Python, and virtual environments
 - R
 - Spreadsheets
- Modes
 - Terminal
 - Editor and/or IDLE(e.g. spyder)
 - Jupyter-notebook

Open Source Python Packages

- Numpy
- Pandas
- Matplotlib
- Scipy
- Sklearn
- Statsmodels
- Tensorflow/keras

Reference Materials – SL

- James, Witten, Hastie & Tibshirani, *An Introduction to Statistical Learning, with Applications in R*, Springer.
 - 2nd ed available
 - SOA: SRM, PA, CAS: MAS-I, MAS-II
- Goodfellow, Bengio and Courville, *Deep Learning*, MIT Press.



Reference Materials – more SL

- Select readings for other ACTU core courses
 - Frees [SOA: SRM]
 - Cowpertwait & Metcalfe [CAS, MAS-I]
 - Dobson & Barnett [CAS, MAS-I]
 - James et al [SOA: SRM, PA, CAS: MAS-I, MAS-II]

Learning From Data

- Supervised learning
 - Outcome measurements
 - Prediction and inference
 - Regression and classification
- Unsupervised learning
 - No outcome measurements
 - Data organization
- ML/SL methods
 - Regularization
 - Cross validation
 - Ensemble learning

Important Pieces

- Training Set
- Model Class
- (Fitted) Model
- Validation Set
- Test Set

Keep in mind

- No universally best approach
- Curse of dimensionality
 - Parametric vs non-parametric approaches
- Bias & variance tradeoff when predicting
 - Bias: how close is the model estimate on average
 - Variance: how variable is the model estimate when fitted with different training sets

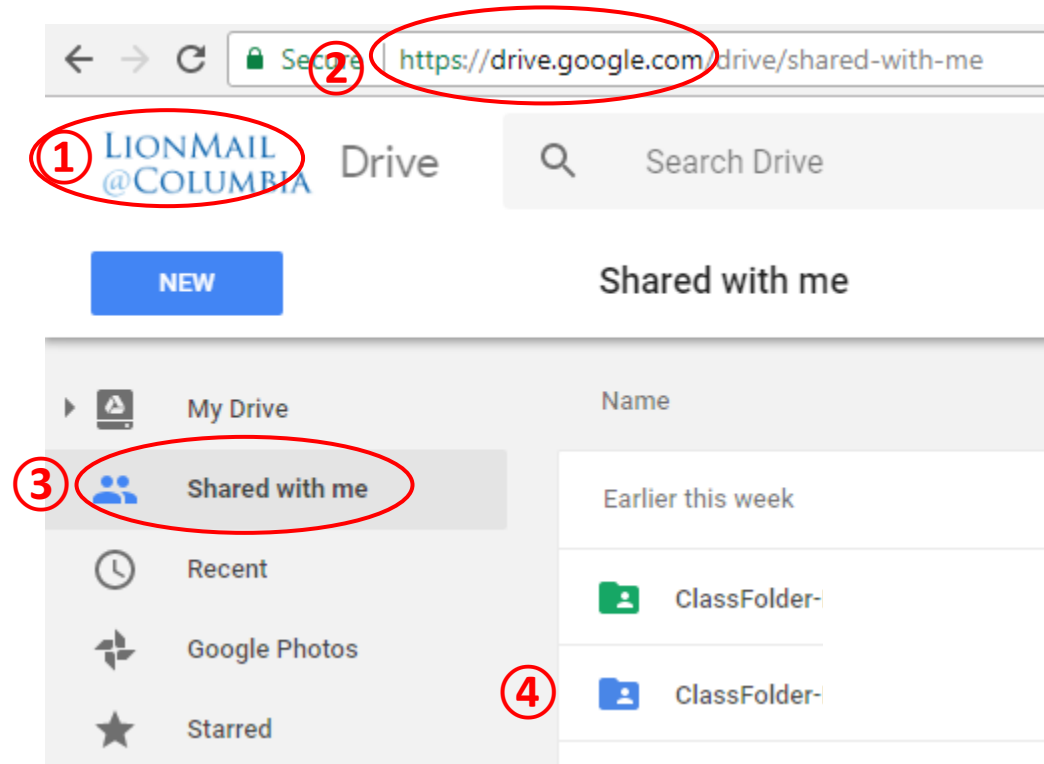
School Stuff

- Calendar
 - First class 1/18 (Thu)
 - Midterm 3/10 (Thu)
 - No classes 3/16 (Tue) 3/17 (Thu)
 - Project 4/26 (Tue), 4/28 (Thu)
 - Last class 4/28 (Thu)
 - Final 5/12 (Thu) – 9am-noon **FAY 301M**

Class Folder

- Class Folder

- ① Log into CU email with your UNI
- ② Go to drive.google.com
- ③ Go to “shared with me”
- ④ Go to ClassFolder-DataSci-Spr2022



- Class Folder

- ① Log into CU email with your UNI
- ② Then go to <https://tinyurl.com/ds2022spring>

Group Project (1)

- Who – minimum 3 and maximum 4 people per team
 - Get to know your peers
 - Build on each other's strengths
- What – issues in finance or insurance
- Why – justify its merit for you and your audience
- How –
 - Find/Construct the relevant data set
 - Apply the tools and approaches discussed in the course to appropriately analyze the data to shed light on your questions
 - Educate the class with your informative and lively presentation!
 - Writeup
- When – see the next page

Group Project (2)

- Keep the dates
 - Project proposal due week 8 (3/10)
 - Draft writeup due week 12 (4/14)
 - Project presentation week 14 (4/26, 4/28)
 - Final writeup due at Final

That was

