CPSC 340:
Machine Learning and Data Mining

Conclusion

Admin

- Assignment 6:
  - Due Friday night.
- Final:
  - April 25 (8:30am, ESB 1013).
  - Covers Assignments 1-6.
  - Past exams posted on GitHub.
  - Closed-book, cheat sheet: 1-page double-sided (same as midterm).
- Office hours:
  - We'll have fewer office hours next week, and more leading up to the exam.
  - A review session is also in the works.
  - See calendar for details.

Big ideas of the course

- Supervised vs. Unsupervised learning
  - Supervised: regression, classification; focus on prediction
  - Unsupervised: find “structure” or “patterns” in the data; clustering; dimensionality reduction
- Fundamental tradeoff of ML, under/overfitting, (cross-)validation
- Golden rule: test data should not (significantly?) influence training
- No free lunch theorem: there is no “best” ML model (so we learn lots)
- The different decisions you need to make and their effects:
  - Collecting your data, preprocessing (standardize columns? What data?)
  - Choosing a model (including feature selection, regularization, basis?)
  - Choosing a loss (am I doing classification or regression? Robust fit?)
  - Choosing an optimization method (GD, SGD, SVD, normal eqn’s, …)

Individual topics covered

- Part 1: EDA, decision trees, NB, KNN, ensembles & random forests
- Part 2: k-means, DBSCAN, hierarchical clustering, outlier detection, association rules
- Part 3: linear regression, basis & other non-linear regression, regularization (L2, L1, L0), gradient descent & SGD, logistic regression, SVM & kernels, maximum likelihood & MAP
- Part 4: PCA & variants, NMF, recommender systems, nonlinear dimensionality reduction (MDS, ISOMAP, t-SNE)
- Part 5: Neural networks, CNNs, deep learning software.

- (blue indicates that this topic appeared on one of the assignments)

CPSC 340 vs. CPSC 540

- Goals of CPSC 340: practical machine learning.
  - Make accessible by avoiding some technical details/topics/models.
  - Present most of the fundamental ideas, sometimes in simplified ways.
  - Choose models that are widely-used in practice.
- Goals of CPSC 540: research-level machine learning.
  - Covers complicated details/topics/models that we avoided.
  - Targeted at people with algorithms/math/stats/sciComp background.
  - Goal is to be able to understand ICML/NIPS papers at the end of course.
  - More on optimization, density estimation, structured prediction & graphical models, Bayesian methods, recurrent neural networks,

Evaluations & surveys (15ish minutes)

- Science course evaluation
  - https://eval.ctlt.ubc.ca/science
  - Deadline is April 10 at 7:00am
- TA evaluations (paper)
  - Only for the tutorial TAs: Issam Laradji and Bita Nejat
  - Need 2 volunteers to take them to the CS main office (ICCS 201) after class
    - Please put all blank forms back in the envelope as well
  - The department ran out of pencils so please share any pencils you have
- GitHub survey
  - https://survey.ubc.ca/s/cpsc340gh/
  - Since this was the first run, your responses will inform us in a big way

Prizes

- I'd like to thank a few students who participated a lot in class.
  - This makes it a lot more meaningful for me.
  - I didn't announce this at the beginning because I didn't want people to participate "for the prize".
- Top Piazza answerers:
  - Winners: Syed I, Dominic K
  - Honourable mention to Vaastav A, Jay C, Bowen J
- Top in-class participants:
  - Winners: Saeid A, Vincent H, Aaron M
  - Honourable mention to Richard C, Amir H, Konrad I

Concluding remarks

- Keep in mind that we've covered a slice of machine learning
- We did not say much about…
  - Data collection/preparation
  - Causality
  - Confidence in our predictions, risks & AI safety
  - Large scale problems and/or distributed computing
  - Presenting results in an understandable way to non-experts
  - Many more ML methods
- Despite all this we covered a lot of ground.
  - You all should feel a sense of accomplishment!

(Unsolicited) General Life Advice

- Try to find an intersection of work you enjoy and careers with enough jobs
  - It's up to you to determine both of those things
  - Do your own research, make your own decisions
  - Don't let your parents influence you too much
- Don't obsess over grades (gamification)
  - You should know when/why you need good grades; they are not worth anything inherently
  - They do not reflect the skills needed for success (not even close)
- Make sure you're happy in the present moment
  - Don't sacrifice happiness because you're "working towards something"
  - You should enjoy university!
- Don't assume the system makes sense
  - Undergraduate achievement has little to do with graduate school or research achievement
  - Your education may not focus on the skills you need to succeed in 2017-2100 (we try but it's hard)
  - Older or more senior people (like me) **can be wrong**!
- You are lucky
  - UBC is one of the best schools in Canada
  - But the above applies even in the best places ☺

Thank you

- Thank you for being a great audience!