

CPSC 340: Machine Learning and Data Mining

Conclusion

Fall 2021

Admin

- Last class today!
- A6 due tonight (late-late deadline Wednesday).
- Final exam updates
 - Lots of office hours between now and the exam.
 - Detailed exam instructions coming over the next few days at <https://piazza.com/class/ksums2w1qd91se?cid=624> – please check it!
 - Academic misconduct: it's not worth it.
 - These things often end badly...

CPSC 340: Overview

1. **Intro to supervised learning** (using counting and distances).
 - Training vs. testing, parametric vs. non-parametric, ensemble methods.
 - Fundamental trade-off, no free lunch, universal consistency.
2. **Intro to unsupervised learning** (using counting and distances).
 - Clustering, outlier detection.
3. **Linear models and gradient descent** (for supervised learning)
 - Loss functions, change of basis, regularization, feature selection.
 - Gradient descent and stochastic gradient.
4. **Latent-factor models** (for unsupervised learning)
 - Typically using linear models and gradient descent.
5. **Neural networks** (for supervised and multi-layer latent-factor models).

Topics from Previous Years

- Slides for other topics that were covered in previous years:
 - [Finding similar items](#): “you may also like” recommendations.
 - [Ranking](#): finding “highest ranked” training examples (Google PageRank).
 - [Multi-dimensional scaling](#): more nonlinear unsupervised models.
 - [Semi-supervised](#): using unlabeled data to help supervised learning.
 - [Sequence mining](#): approximate matching of patterns in large sequences.
- And some bonus lectures from previous years:
 - [Automatic differentiation](#): how to build PyTorch, to do backprop for you.
 - [# of gradient descent iterations](#): theoretical analysis of how long it’ll take.

CPSC 330 vs. 340

- CPSC 330: Applied Machine Learning.
 - Not intended as a sequel or prequel to 340.
- There is **some overlap** in content, but **focus is different**:
 - More emphasis on the **other steps of the data processing pipeline**:
 - Data cleaning, feature extraction, reproducible workflows, communicating results.
 - More emphasis of “**how to apply ML**”, less on “how ML works”.
- Also some new topics:
 - Time-series data, word embeddings, data preprocessing.
 - Lecture videos available at https://www.youtube.com/playlist?list=PLWmXHcz_53Q2BXsWviGgEqdISHmfsjSzC

CPSC 330 vs. 440

- CPSC 440 is now an undergrad course (used to only be 540).
 - Intended as a direct sequel to 340.
 - Basically starting with CNNs and going from there.
- Main focuses:
 - What if y_i is a sentence or an image or a protein?
 - Giving you the background to understand the latest advances.
- Prerequisites:
 - Expected that you know everything in CPSC 340 and CPSC 320.

CPSC 440/540 Topics

- Much more on deep learning
- Generative models
 - How do I make “more” samples from this distribution?
- Latent-variable models
 - (Much) fancier versions of PCA-type unsupervised models
- Markov models
 - Modeling processes that are happening over time
- Probabilistic graphical models
 - Building any big complicated model you want
- Bayesian methods
 - How do I incorporate uncertainty every step of the way?

Other ML-Related Courses

- CPSC 406: Numerical optimization algorithms (e.g. gradient descent).
- CPSC 422: Time series, reinforcement learning (and more).
- CPSC 436N: Natural language processing (340 is a prereq).
- STAT 406: Similar/complementary topics.
- STAT 460/461: Advanced statistical issues ('n' goes to ∞)
- CPSC 532J: Never-ending reinforcement learning.
- CPSC 532S[utherland]: Theory of ML: when will our models work?
- CPSC 532S[igal]: Multimodal deep learning, e.g. vision + sound.
- CPSC 532V: Advanced natural language models (using common-sense reasoning).
- CPSC 532W: Probabilistic programming (better to take 440 first).
- CPSC 532P(?) / 538L: Differential privacy (main approach to keeping user data private in ML).
- CPSC 533R, 533V, 533Y: Deep learning for computer graphics and robotics.
- EECE 592: Deep learning and reinforcement learning.
- MATH 605D: Causality+graphical models (sometimes) / tensor decompositions (other times).
- STAT 520A: Bayesian analysis (incorporating uncertainty).

[CAIDA events](#)

[ML Reading Group](#)

Also: CPSC and STAT
are hiring AI/ML
faculty this year; go
to their talks!

Also: more AI courses at <https://caida.ubc.ca/index.php/teaching>

(pause)

Course evaluations

- We'll now take 10 min for you to fill out the course evaluations.
- Link: https://canvas.ubc.ca/courses/78047/external_tools/4732
- If you haven't already, please consider checking out this Piazza post on the evaluations:
<https://piazza.com/class/ksums2w1qd91se?cid=644>

(pause)

Grad School Advice / FAQs

- Don't do it!
 - Or, if you insist...
- Get work experience: co-op, full-time after undergrad, grad school internships
 - Build skills, confidence, awareness of what's out there, and *boredom* (or curiosity)
- Try for some research experience before grad school, if you can
 - Very helpful for getting in, but also vital for *knowing if you want to do it*
- Research Master's vs. Professional Master's
 - Research Master's (pays you): reading/writing papers, inventing things, leads to PhD
 - Straight to PhD (pays you): more-or-less default in US, unusual in Canada/Europe
 - Professional Master's (you pay): practical skills to get you a job, leads to industry
- Canada is a world leader in ML (but maybe think about US/Europe/... too)
- If you go, remember: the system is terrible, some advisors are a bad fit (or just bad), you're not an imposter, and leaving/taking breaks is not failing

Grad School Applications Advice

- Do interesting side-projects and post them on your GitHub profile (also great for industry)
- Make a personal website and/or blog (also great for industry)
- For a research Master's, prior ML research experience is a *huge* boost to your application
 - But it's hard to come by, because so many undergrads want to do ML research
 - Possible path: excel in ML course(s), become a TA, excel as a TA, do a summer research internship
- Pick your referees carefully
 - Knows you well in an academic/professional context >> is famous
 - Try to gauge if the person is enthusiastic about writing the letter
- Consider (also) applying to “ML-adjacent” programs/advisors
- If there's something that needs explaining (e.g., low grades), explain it!
- Personal contact with potential advisors can help
 - But we get *flooded* with generic emails, so only do it if you put the effort in
- Next steps at UBC:
 - Take CPSC 440 and other courses mentioned earlier
 - TA for ML courses (CPSC 340, 330, 440, etc)
 - Get work experience, do projects

Unsolicited General Life Advice - Now

- 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 *unnecessarily* obsess over grades (gamification)
 - You should know *why* you need good grades; they are not worth anything inherently
 - They do not reflect the skills needed for success (not even close)
 - Don't measure peoples' worth by grades: having a C+ vs. A+ GPA doesn't make you "inferior" or "superior"
- Make sure you're happy in the present moment
 - Don't sacrifice current happiness / mental health because you're "working towards something"
 - You should enjoy university!
 - If there's a Big Thing™ you're not letting yourself think about, maybe try thinking about it. (Don't be like Danica.)
- Don't assume the system makes sense
 - For example, 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 2021-2100 (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, and Canada is a great place to be
 - But all of the above applies no matter where you are!

Unsolicited General Life Advice - Future

- Some (many? most?) of you will find yourselves in positions of power one day
 - As parents
 - As supervisors/managers at work
 - As educators (like me)
- Try to remember the feeling of being on the powerless end of power dynamic
 - It can be easy to forget sometimes
 - Though sometimes, it can be hard to forget...
- When the time comes, try to be compassionate and humble
 - Try to remember that feeling!
 - A little thing you say might leave someone in a panic or reading into your wording for days
- It is complicated to assign blame
 - If someone is “lazy” or “flaky” is that their fault? What is their story?
 - Our culture of “meritocracy” is generally not very sympathetic...
- Likewise, it is complicated to assign credit
 - How did I get into this position of power as a prof? Did I “earn” it? That’s a complicated question.
- As UBC students, you are probably likely to hold positions of power
 - We will each have our own philosophy on the above issues
 - But hopefully we can agree on the value of compassion



Talia Ringer
@TaliaRinger

kind mediocrity > cruel genius

The End

- This is the last slide of the course!
- Good luck with the studying and take care, everyone.

