CPSC 340 and 532M: Machine Learning and Data Mining

Mark Schmidt
University of British Columbia, Fall 2018
www.ugrad.cs.ubc.ca/~cs340

Some images from this lecture are taken from Google Image Search, contact me if you want the reference
Big Data Phenomenon

• We are collecting and storing data at an unprecedented rate.

• Examples:
  – YouTube, Facebook, MOOCs, news sites.
  – Credit cards transactions and Amazon purchases.
  – Transportation data (Google Maps, Waze, Uber)
  – Gene expression data and protein interaction assays.
  – Maps and satellite data.
  – Large hadron collider and surveying the sky.
  – Phone call records and speech recognition results.
  – Video game worlds and user actions.
Big Data Phenomenon

• What do you do with all this data?
  – Too much data to search through it manually.

• But there is valuable information in the data.
  – How can we use it for fun, profit, and/or the greater good?

• Data mining and machine learning are key tools we use to make sense of large datasets.
Data Mining

• Automatically extract useful knowledge from large datasets.

• Usually, to help with human decision making.
Machine Learning

• Using computer to automatically detect patterns in data and use these to make predictions or decisions.

• Most useful when:
  – We want to automate something a human can do.
  – We want to do things a human can’t do (look at 1 TB of data).
Data Mining vs. Machine Learning

• Data mining and machine learning are very similar:
  – Data mining often viewed as closer to databases.
  – Machine learning often viewed as closer AI.

• Both are similar to statistics, but more emphasis on:
  – Large datasets and computation.
  – Predictions (instead of descriptions).
  – Flexible models (that work on many problems).
Deep Learning vs. Machine Learning vs. AI

• Traditional we’ve viewed **ML as a subset of AI**.
  – And “deep learning” as a subset of **ML**.
Applications

• Spam filtering:

• Credit card fraud detection:

• Product recommendation:
Applications

• Motion capture:

• Optical character recognition and machine translation:

• Speech recognition:
Applications

• Face detection:

• Object detection:

• Sports analytics:
Applications

• Personal Assistants:

• Medical imaging:

• Self-driving cars:
Applications

• Scene completion:

• Image annotation:
Applications

• Discovering new cancer subtypes:

• Automated Statistician:

2.4 Component 4: An approximately periodic function with a period of 10.8 years. This function applies until 1643 and from 1716 onwards.

This component is approximately periodic with a period of 10.8 years. Across periods the shape of this function varies smoothly with a typical lengthscale of 36.9 years. The shape of this function within each period is very smooth and resembles a sinusoid. This component applies until 1643 and from 1716 onwards.
Applications

• Mimicking artistic styles and inceptionism:
Applications

• “Deep dream”: 
Applications

• Fast physics-based animation:

• Mimicking art style in video.
• Recent work on generating text/music/voice/poetry/dance.
Applications

• Beating human Go masters:
Summary:
- There is a lot you can do with a bit of statistics and a lot data/computation.

But it is important to know the limitations of what you are doing.
- “The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.” – John Tukey
- A huge number of people applying ML are just “overfitting”.
  - Or don’t understand the assumptions needed for them to work.

We are in exciting times.
- Major recent progress in fields like speech recognition and computer vision.
- Things are changing a lot on the timescale of 3-5 years.
- A bubble in ML investments.
Course Outline

• Next class discusses “exploratory data analysis”.

• After that, the remaining lectures focus on five topics:
  1) Supervised Learning.
  2) Unsupervised learning.
  3) Linear prediction.
  4) Latent-factor models.
  5) Deep learning.
• I will include a lot of “bonus slides”.
  – May mention advanced variations of methods from lecture.
  – May overview big topics that we don’t have time for.
  – May go over technical details that would derail class.

• You are **not expected to learn** the material on these slides.
  – But they’re useful if you want to take 540 or work in this area.

• I’ll use this colour of background on bonus slides.
Photo I took in the UK on the way home from the “Optimization and Big Data” workshop:
Supervised Learning

• **Classification:**
  – Given an object, assign it to predefined ‘classes’.

• **Examples:**
  – Spam filtering.
  – Body part recognition.
Unsupervised Learning

• **Clustering:**
  – Find groups of `similar’ items in data.

• **Examples:**
  – Are there subtypes of tumors?
  – Are there high-crime hotspots?

• **Outlier detection:**
  – Finding data that doesn’t belong.

• **Association rules:**
  – Finding items frequently ‘bought together’.
Linear Prediction

- **Regression:**
  - Predicting continuous-valued outputs.
- Working with very **high-dimensional** data.
Latent-Factor Models

• Principal component analysis and friends:
  – Low-dimensional representations.
  – Decomposing objects into “parts”.
  – Visualizing high-dimensional data.

• Collaborative filtering:
  – Predicting user ratings of items.
Deep Learning

- **Neural networks**: Brain-inspired ML when you have a lot of data/computation but don’t know what is relevant.