Lecture 27 – Course summary

DSC 40A, Fall 2022 @ UC San Diego

Dr. Truong Son Hy, with help from **many others**

Announcements

- Final exam is coming soon!
- Review the solutions to previous homeworks and groupworks.
- Identify which concepts are still iffy. Re-watch lecture and ask questions (now!).
- Look at the past exams at `https://dsc40a.com/resources`.
- Study in groups.
- Make a “cheat sheet”.
- Bring a calculator.
- Remember to submit The Course and Professor Evaluations (CAPE) – deadline December 2. If everyone submits CAPE, everyone will get a bonus percentage!

Final schedule

O	40A	Theor Fndtns of Data Sci I (4 Units)						Prerequisites | Resources | Evaluations		
		LE	A00	MWF	3:00p-3:50p	PCYNH	122	Hy, Truong Son		
	88107	DI	A01	M	5:00p-5:50p	PCYNH	122	Hy, Truong Son	9	115
		FI	12/03/2022	S	7:00p-9:59p	CSB	001			
O	40A	Theor Fndtns of Data Sci I (4 Units)						Prerequisites | Resources | Evaluations		
		LE	B00	MWF	4:00p-4:50p	PCYNH	122	Soleymani, Mahdi		
	88109	DI	B01	M	6:00p-6:50p	PCYNH	122	Soleymani, Mahdi	19	115
		FI	12/03/2022	S	7:00p-9:59p	CSB	002			

Time: December 3rd, 2022 – 7:00pm to 10:00pm (3 hours)
Location: CSB building – room 001 (for my section)

https://act.ucsd.edu/scheduleOfClasses/scheduleOfClassesStudentResult.htm

Please double-check!

Agenda

- Acknowledgements
- High-level summary of the course.

Acknowledgements

Acknowledgements

Special thanks to:

- Mahdi Soleymani, the other instructor of the course.
- Pushkar Bhuse, the teaching assistant, and Weiyue (Larry) Li, Karthikeya Manchala, Aryaman Sinha, Yujia (Joy) Wang, Yuxin Guo, Vivian Lin, Shiv Sakthivel, Jessica Song, the tutors of the course.
- Justin Eldridge, Janine Tiefenbruck and Suraj Rampure, the instructors of the past courses for their helps.
- And to all of you, the students who attended, worked hard and gave us feedback to improve the course further.

What was this course about?

Part 1: Supervised learning

The "learning from data" recipe to make predictions:

1. Choose a **prediction rule**. We've seen a few:
   - Constant: $H(x) = h$.
   - Simple linear: $H(x) = w_0 + w_1 x$.
   - Multiple linear: $H(x) = w_0 + w_1 x^{(1)} + w_2 x^{(2)} + \ldots + w_d x^{(d)}$.
2. Choose a **loss function**.
   - Absolute loss: $L(h, y) = |y - h|$.
   - Squared loss: $L(h, y) = (y - h)^2$.
   - 0-1 loss, UCSD loss, etc.
3. Minimize **empirical risk** to find optimal parameters.
   - Algebraic arguments.
   - Calculus (including vector calculus).
   - Gradient descent.

Part 1: Unsupervised learning

- When learning how to fit prediction rules, we were performing **supervised machine learning**.
- Then, we discussed **k-Means Clustering**, an **unsupervised machine learning** method.
  - Supervised learning: there is a “right answer” that we are trying to predict.
  - Unsupervised learning: there is no right answer, instead we’re trying to find patterns in the structure of the data.

Part 2: Probability fundamentals

- If all outcomes in the **sample space** S are equally likely, then $P(A) = \frac{|A|}{|S|}$.
- $\bar{A}$ is the **complement** of event A. $P(\bar{A}) = 1 - P(A)$.
- Two events A, B are **mutually exclusive** if they share no outcomes, i.e. they don't overlap. In this case, the probability that A happens or B happens is $P(A \cup B) = P(A) + P(B)$.
- More generally, for any two events, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- The probability that events A and B both happen is $P(A \cap B) = P(A)P(B|A)$.
  - $P(B|A)$ is the probability that B happens given that you know A happened.
  - Through re-arranging, we see that $P(B|A) = \frac{P(A \cap B)}{P(A)}$.

Part 2: Combinatorics

- A **sequence** is obtained by selecting k elements from a group of n possible elements with replacement, such that order matters.
  - Number of sequences: n^k.

- A **permutation** is obtained by selecting k elements from a group of n possible elements without replacement, such that order matters.
  - Number of permutations: $P(n,k) = \frac{n!}{(n-k)!}$.

- A **combination** is obtained by selecting k elements from a group of n possible elements without replacement, such that order does not matter.
  - Number of combinations: $\binom{n}{k} = \frac{n!}{(n-k)!k!}$.

Part 2: The law of total probability and Bayes' theorem

- A set of events $E_1, E_2, ..., E_k$ is a **partition** of S if each outcome in S is in exactly one E_i.
- The **law of total probability** states that if A is an event and $E_1, E_2, ..., E_k$ is a partition of S, then

$$\begin{aligned} P(A) &= P(E_1) \cdot P(A|E_1) + P(E_2) \cdot P(A|E_2) + ... + P(E_k) \cdot P(A|E_k) \\ &= \sum_{i=1}^{k} P(E_i) \cdot P(A|E_i) \end{aligned}$$

- **Bayes' theorem** states that

$$P(B|A) = \frac{P(B) \cdot P(A|B)}{P(A)}$$

- We often re-write the denominator $P(A)$ in Bayes' theorem using the law of total probability.

Part 2: Independence and conditional independence

- Two events A and B are **independent** when knowledge of one event does not change the probability of the other event.
  - Equivalent conditions: $P(B|A) = P(B)$, $P(A|B) = P(A)$, $P(A \cap B) = P(A) \cdot P(B)$.

- Two events A and B are **conditionally independent** if they are independent given knowledge of a third event, C.
  - Condition: $P((A \cap B)|C) = P(A|C) \cdot P(B|C)$.

- In general, there is no relationship between independence and conditional independence.

- See pinned post on Campuswire for clarification.

Part 2: Naive Bayes

- In classification, our goal is to predict a discrete category, called a **class**, given some features.

- The **Naive Bayes** classifier works by estimating the numerator of $P(\text{class}|\text{features})$ for all possible classes.

- It uses Bayes' theorem:

$$P(\text{class}|\text{features}) = \frac{P(\text{class}) \cdot P(\text{features}|\text{class})}{P(\text{features})}$$

- It also uses a "naive" simplifying assumption, that **features are conditionally independent given a class**:

$$P(\text{features}|\text{class}) = P(\text{feature}_1|\text{class}) \cdot P(\text{feature}_2|\text{class}) \cdot \ldots$$

Summary

Learning objectives

At the start of the quarter, we told you that by the end of DSC 40A, you'll...

- understand the basic principles underlying almost every machine learning and data science method.
- be better prepared for the math in upper division: vector calculus, linear algebra, and probability.
- be able to tackle problems such as:
    - How do we know if an avocado is going to be ripe before we eat it?
    - How do we teach a computer to read handwritten text?
    - How do we predict a future data scientist's salary?

What's next?

In DSC 40A, we just scratched the surface of the theory behind data science. In future courses, you'll build upon your knowledge from DSC 40A, and will learn:

- More supervised learning.
  - Logistic regression, decision trees, neural networks, etc.
- More unsupervised learning.
  - Other clustering techniques, PCA, etc.
- More probability.
  - Random variables, distributions, etc.
- More connections between all of these areas.
  - For instance, you'll learn how probability is related to linear regression.
- More practical tools.