

Logistics and Overview

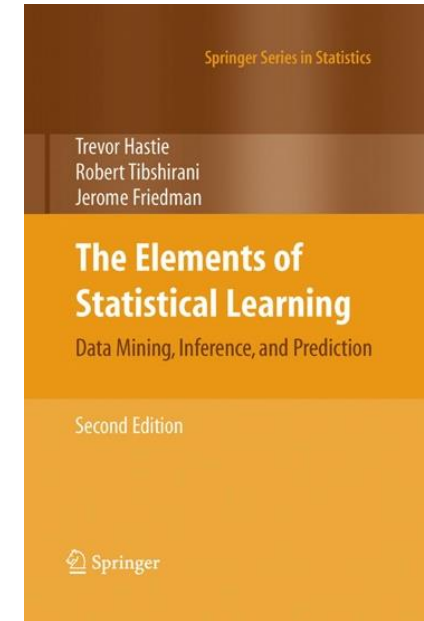
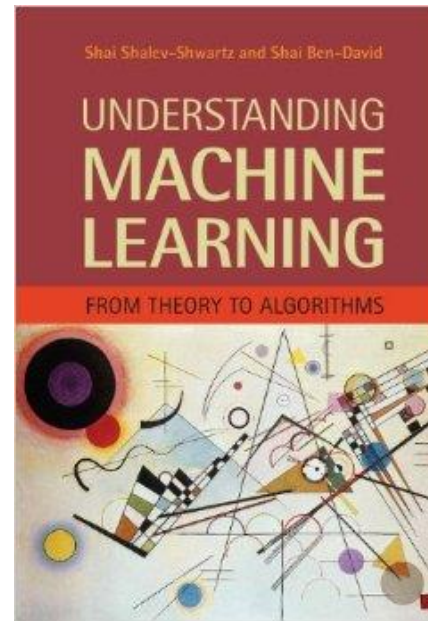
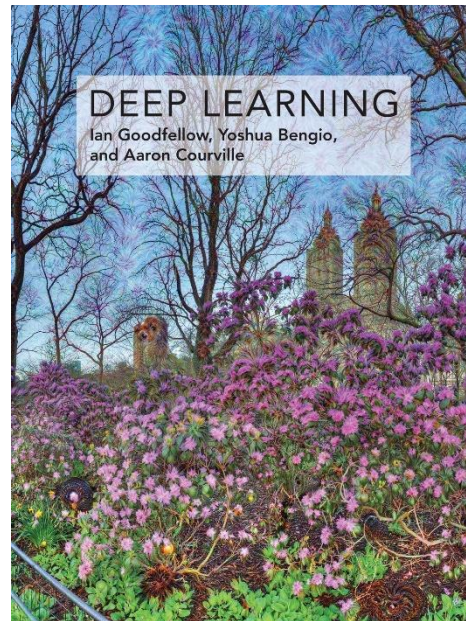
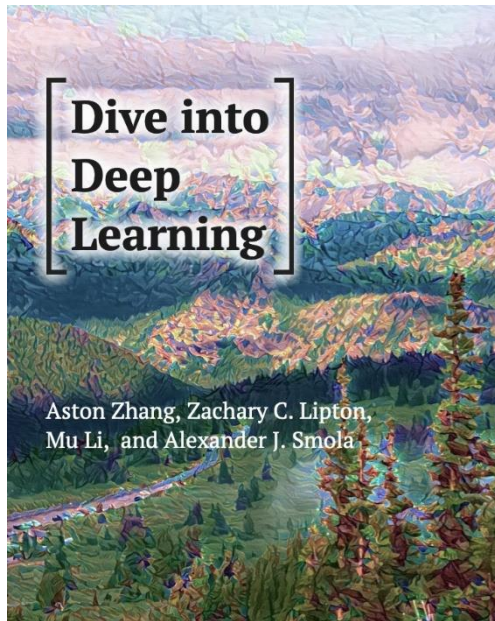
Gautam Kamath

Course Info

- Instructor: Gautam Kamath (gckamath@uwaterloo.ca)
 - Office Hours: Tuesday and Thursday at noon Waterloo time (or by appt)
- TAs: Lasantha Fernando, Yuxiang Huang, Kamyar Setayesh Ghajar
- Website: <http://www.gautamkamath.com/courses/CS480-sp2021.html>
 - Slides, notes, assignments, logistics
- YouTube
 - Lecture videos
- Piazza: <http://www.piazza.com/uwaterloo.ca/spring2021/cs480680/>
 - Announcements, questions and answers
- Learn and CrowdMark
 - Submissions and grades
- Teams
 - Office hours

Textbook

- No required textbook
- Materials will be linked from the website
- Some good reference books also listed



Course Prerequisites

- Officially:
 - Algorithms (CS 341 or SE 240)
 - Statistics (STAT 206 or STAT 231 or STAT 241)
- In reality:
 - Probability, statistics, linear algebra, calculus, programming (Python)
 - Mathematical maturity

Workload

- Between 20-25 lectures, each ~1 hour
- Five assignments, roughly every other week
 - Lowest assignment score dropped
 - Each one worth 20%, $4 \times 20\% = 80\%$ total
- Submit on CrowdMark (writeup) and LEARN (code)
 - LaTeX typesetting is recommended

Policies

- Do your own work!
 - High-level discussions: Great!
 - Copying/sharing code/solutions: Bad
 - Acknowledge sources
- Two 48-hour extensions for assignments. Read rules on website.
- Further extensions require justification + documentation (illness, etc.)
- Regrade requests: within 1 week of grades released

Final Project

- Remaining 20% of grade
- CS 480: Kaggle Competition
- CS 680: Research Project
 - Ideal: A novel and interesting project submittable to an ML conference
 - Project proposal (2-4 pages) worth 5%, Report (8 pages) worth 15%

What is Machine Learning, and why should I care?

- Given data, learn something about it!
 - Find some structure that exists in the data
 - Learn something about the distribution that generated the data
 - Figure out the relationship between features and labels of datapoints
- Machine learning is ubiquitous!
 - Gmail, Netflix, Facebook, Apple, Amazon, ...
- Essential topic for a student of Computer Science

Types of Learning (roughly)

- Supervised Learning
- Reinforcement Learning
- Unsupervised Learning

Supervised Learning

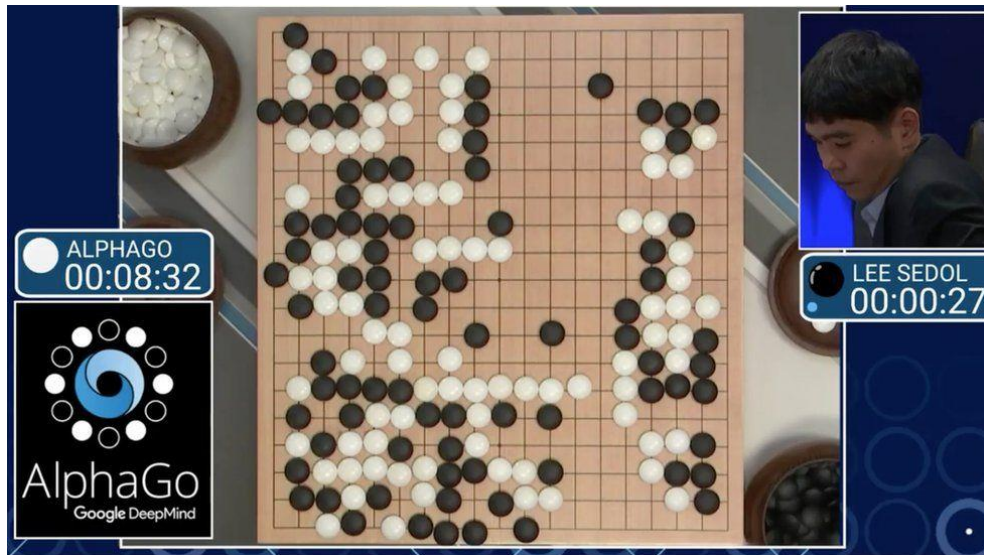
- Classification, Regression, Ranking



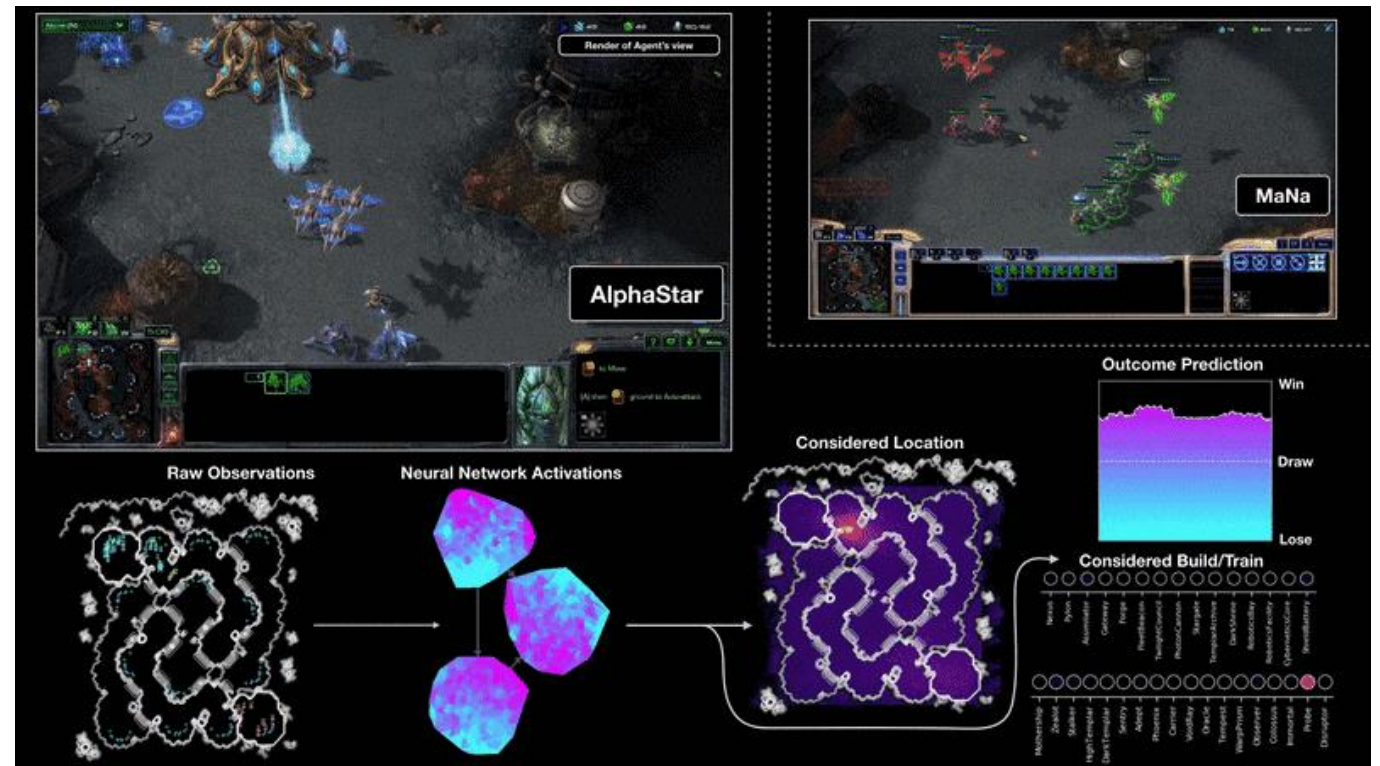
cat

Reinforcement Learning

- Control, Games



AlphaGo - The Movie



AlphaStar: Mastering the Real-Time Strategy Game StarCraft II

Unsupervised Learning

- Clustering, Visualization, Representation, Density Estimation

SYSTEM PROMPT (HUMAN-WRITTEN) *In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.*

MODEL COMPLETION (MACHINE-WRITTEN, 10 TRIES)

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

Pérez and the others then ventured further into the valley. "By the time we reached the top of one peak, the water looked blue, with some crystals on top," said Pérez.

Pérez and his friends were astonished to see the unicorn herd. These creatures could be seen from the air without having to move too much to see them - they were so close they could touch their horns.

Better Language Models and Their Implications



<https://thispersondoesnotexist.com/>

Course Topics

- Classic Machine Learning
 - Perceptron, Linear Regression, k-Nearest Neighbours, Logistic Regression, Support Vector Machines, Kernel Methods, Decision Trees, Boosting
- Neural Networks
 - Multilayer Perceptron, Deep Networks, Convolutional Neural Networks, Recurrent Neural Networks, Graph Neural Networks, Optimizers
- Generative Models
 - Mixture Models, GANs, VAEs
- Modern Topics in Neural Networks
 - Attention, Robustness, Privacy, Fairness