Part I

Administrative Stuff
Course Information

• Instructor:
  – Kilian Q. Weinberger

• Homepage:
  – https://courses.cis.cornell.edu/cs4780/2017sp/
  – Vocareum: http://vocareum.com
  – Piazza: https://piazza.com/class/iyang4nk2rsxsv

• TAs:
  – multiple

• Office Hours / Recitations:
  • TA Office Hours: Every day. (Details will be posted on Piazza.)
  • My Office Hours: Mondays 9am-10am Gates 410

• Questions:
  • Post all questions on Piazza (you can make them private)
  • Do not email me directly (except in an emergency)

• (Optional) Course Books
  • Machine Learning a Probabilistic Perspective (K.P. Murphy)
  • The Elements of Statistical Learning (Hastie, Tibshirani, Friedman)
Course Breakdown

- **50% Theory: Midterm + Final (1/4 each)**
  - Closed book
  - Optional collective cheat sheets
  - No personal notes

- **50% Programming Assignments**
  - Up to 2 members in each team
  - 2 days extension *per team per project*
  - Autograder (unlimited resubmissions)
  - *Extra credit if you beat my own submission*
  - *Extra credit if you win contests*

- **(0% or 100%) Homeworks**
  - Up to 5 members in each team
  - Binary grades {0,100}
  - Preparation for exam
  - Must submit a *faithful attempt*
Placement Exam

- Due Monday!!!
- Label with your NETID on front!!!
- Get started early!!
- It is there for your own protection!
- This is how you get a Vocareum invite!
Study Groups (2-5 people)

- You **must** join a study group by next week.
- Find people on Piazza
- This course will cover non-trivial material, learning in a group makes it less hard and more fun!
MLunch

- Lunch with Kilian every Monday
- Up to 8 students (online sign up sheet)
- (will post link on Piazza)
- Details will follow …
Pass placement exam -> get account
There will be (roughly) 8 projects
You have (roughly) 2 weeks for each project
Unlimited submits until deadline
Important notes:

- `language.txt` must be filled with one word:
  - `julia`
  - `python`
- Setup a (secret) `screen name` for your team for the leaderboard
  - (top right corner - click on your login)
- Only text with `# <GRADED> and # </GRADED>` will be graded
- !!!!You MUST form teams before you get started!!!!!
Course Topics

* We will cover:
  * Parametric / Non-parametric learning
  * Empirical Risk Minimization
  * Bias/Variance Trade-off
  * Boosting
  * Support Vector Machines
  * Deep Learning
  * (Recommender systems)

* We will not cover:
  * Graphical Models
  * Reinforcement Learning
  * Unsupervised Learning
  * Genetic Programming
Prerequisites
Prerequisites

- Three pillars of ML:
  - Statistics / Probability
  - Linear Algebra
  - Multivariate Calculus

- Should be confident in at least 1/3, ideally 2/3.

- TAs might be able to give recitations on some topics if needed. (But don’t rely on it.)
About this course

- Take this course if …
  - you are interested in Machine Learning
  - you are comfortable with a decent amount of mathematics
  - you are not scared of programming
- Don’t take this course if …
  - matrices scare you
  - you don’t remember how to take derivatives
  - you want an easy A
- You cannot take this course if you fail the placement exam.
  - In that case, take appropriate prerequisites and come back next year.
Student comments

• “[…] Requires a good knowledge in math and derivatives.”

• “A TON of work, but mostly worth it for a very valuable skill.”

• “great course, but prepare to work your butt off.”

• “The topics were pretty complicated and difficult to understand quickly. I would have preferred a slightly slower pace.”

• “It's mostly a math class”
Academic Integrity

• Zero Tolerance policy
  • All occurrences will be reported
• I actively look for academic conduct violations
• The autograder checks for plagiarism daily
• Past outcomes
  1. Student cries in my office
  2. Student may ruin transcript or (much) worse
  3. (If on visa, student can lose permission to stay in U.S.)
Academic Integrity

Examples:

✦ Most common: Students steal from same source

✦ Students post to RentACoder.com or other page

✦ Students post solutions on the web

✦ Students use solutions from last year’s course
Part II: What is MACHINE LEARNING?
Traditional Computer Science

Traditional CS:

Data → Program → Computer → Output
Machine Learning

Traditional CS:

Data → Program → Output → Computer

Machine Learning:

Data → Output → Program → Computer
Machine Learning

Machine Learning:

Data → Computer → Program → Output

Traditional CS:

Data → Computer → Output
Machine Learning

Training:

Data → Computer → Program → Output

Testing:

Data → Computer → Output
What is Machine Learning?

* Formally: (Mitchell 1997): A computer program $A$ is said to learn from experience $E$ with respect to some class of tasks $T$ and performance measure $P$, if its performance at tasks in $T$, as measured by $P$, improves with experience $E$.

* Informally: *Algorithms that improve on some task with experience.*
A (very brief) History of ML
Samuel’s Checker Player (1952)

- Basically Shannon’s Minimax Algorithm (traditional AI)
- Included simple learning algorithm to improve board evaluation.
- Player improved over time!!
Perceptron 1957
(Frank Rosenblatt @ Cornell)

- Provable convergence properties
- Eventually leads to **multilayer perceptron** = Artificial Neural Networks = Deep Learning
AI Winter

- (1969) Minsky & Papert “killed” AI
- Burst huge expectation bubble
- Funding for AI research collapsed for decades
Rebirth as Machine Learning

- Machine Learning:
  - Originally: Mostly a name game to get funding.

- Profound difference:
  - ML: Bottom up, AI: Top down
  - ML: More practical smaller goals
  - Based on Statistics and Optimization, not Logic
TD-Gammon (1994)

- Gerry Tesauro (IBM) teaches a neural network to play Backgammon. The net plays 100K+ games against itself and beats world champion [Neurocomputation 1994]

- Algorithm teaches itself how to play so well!!!
Deep Blue (1997)

- IBM’s Deep Blue wins against Kasparov in chess. Crucial winning move is made due to Machine Learning (G. Tesauro).
Example: Websearch
Example: Spam Filter

From: Leap 2 Upgrades <upgrade.noreply@kagi.com>
Subject: [Bulk] Leap 2 Upgrades: Kilian Weinberger, save over 50% when you upgrade to Leap 2
Date: September 21, 2010 6:33:55 PM CDT
To: Kilian Weinberger <kilianw@seas.upenn.edu>

Limited Time Special Upgrade Offer

Upgrade to Yep 2 or Leap 2 for only $19 or... buy the 4-product bundle for only $29
Top Stories

Lindsay Lohan »
Lindsay Lohan Released From Jail Hours After Getting a 30 Day Sentence
Actress Archives - Fabiana Santana - 2 hours ago
Lindsay Lohan has been freed from jail, just hours after receiving a 30 jail sentence by Superior Court Judge Elden Fox. Lohan posted $30000 bail and was released according to a Los Angeles County Sheriff's Department spokesperson.

Sydney Mom...
Video: At Least 28 Days in Jail for Lindsay Lohan
The Associated Press

It Could Take Ten Years for Lindsay Lohan's Career to Recover, Experts Say
FOXNews
Los Angeles Times - The Hollywood Gossip - PopEater - CNN International
all 2,876 news articles »

Carmelo Anthony »
Possible four-team Carmelo Anthony trade has "legs"
Denver Post - Benjamin Hochman - 7 hours ago
The Nuggets know they can't get a Melo for Melo. But they can get younger, bigger and free up salary cap space. That appears to be the direction they are heading, according to multiple NBA sources who say a four-team trade that ...

CBC.ca
Nets move closer to landing Anthony
Yahoo! Sports

NBA Saturday: Will Anthony Sign Off?
HoopsWorld
all 661 news articles »

GLBT »
Judge orders military to reinstate gay nurse
Washington Post - Brian Concannon

Spotlight

Japan turns down China's demand for an apology over boat row
CNN - Wen Jiabao - 3 hours ago
Oracle Growth Plans Worry Rivals and Customers
Oracle spreadsheet

kilian@gmail.com | Web History | Settings | Sign out
Very soon: Autonomous Cars
When will it stop?

* The human brain is one big learning machine
  * We know that we can still do a lot better!

* However, it is hard. Very few people can design new ML algorithms.

* But many people can use them!
What types of ML are there?

As far as this course is concerned:

* **supervised learning:** Given labeled examples, find the right prediction of an unlabeled example. (e.g. *Given annotated images learn to detect faces.*)

* **unsupervised learning:** Given data try to discover similar patterns, structure, sub-spaces (e.g. *automatically cluster news articles by topic*)

* **reinforcement learning:** Try to learn from delayed feedback (e.g. *robot learns to walk, fly, play chess*)
#1 Skill employers wish for

12 IT skills that employers can't say no to

Job hunters with these IT skills are assured of employment, now and in the future

1) Machine learning

As companies work to build software such as collaborative filtering, spam filtering and fraud-detection applications that seek patterns in jumbo-size data sets, some observers are seeing a rapid increase in the need for people with machine-learning knowledge, or the ability to design and develop algorithms and techniques to improve computers' performance, Scott says.

"It's not just the case for Google," he says. "There are lots of applications that have big, big, big data sizes, which creates a fundamental problem of how you organize the data and present it to users."

Demand for these applications is expanding the need for data mining, statistical modeling and data structure skills, among others, Scott says. "You can't just wave your hand at some of these problems -- there are subtle differences in how the data structures or algorithms you choose impacts whether you get a reasonable solution or not," he explains.

You can acquire machine-learning knowledge either through job experience or advanced undergraduate or graduate coursework, Scott says. But no matter how you do it, "companies are snapping up these skills as fast as
A few quotes ...

- “A breakthrough in machine learning would be worth ten Microsots” (Bill Gates, Chairman, Microsoft)

- “Machine learning is the next Internet” (Tony Tether, form. Director, DARPA)

- “Machine learning is the hot new thing” (John Hennessy, President, Stanford)

- “Web rankings today are mostly a matter of machine learning” (Prabhakar Raghavan, Director Research, Yahoo)

- “Machine learning is going to result in a real revolution” (Greg Papadopoulos, CTO, Sun)

- “Machine learning is today’s discontinuity” (Jerry Yang, CEO, Yahoo)
Now let’s get crack’n
To-do action items:

- Join Piazza!
- Finish Placement Exam!
- Find study group
- Autograder Setup:
  - Find Teammate
  - Start Project 0