What's up in Computer Science?

- **Lots of attention to MOOCs**
  - ➤ **Massively Open Online Courses**
  - ➤ **Many started by computer scientists, courses, why?**
  - ➤ **Leverage statistical/computational techniques for …**

- **Machine Learning, what is it?**
  - ➤ **Digital pathologist? Data-driven science**
  - ➤ **Statistical and computational techniques**
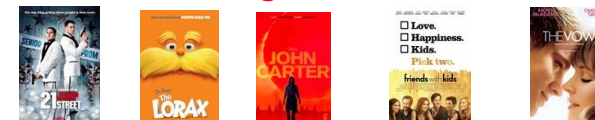 - ➤ **The new AI?**

Math, Engineering, Sociology

- **Netflix prize in 2009**
  - ➤ **Beat the system, win**
  - ➤ **http://nyti.ms/sPvR**

Collaborative Filtering

- **How does Amazon know what I want?**
  - ➤ **Lots of customers, lots of purchases**

- **How does Pandora know music like Kanye's?**
  - ➤ **This isn't really collaborative filtering, more content-based**

- **How doe Netflix recommend movies?**
  - ➤ **Why did they offer one million $$ to better their method?**

- **Students at Duke who like Compsci also like …**
  - ➤ **Could this system be built?**

From User Rating to Recommendations

21 Jump Street	Lorax	John Carter	Friends with Kids	The Vow
3	-3	5	-2	-3
2	2	3	2	3
4	4	-2	1	-1

- **What should I choose to see?**
  - ➤ **What does this depend on?**
- **Who is most like me?**
  - ➤ **How do we figure this out**

How do we store/use tabular data?

- **How do we access data in FriendScore APT?**
  - What does `friends[i][j]` mean?
    - What is `friends[i]`, what is `s[j]` where s is a string?

- **What is a matrix?**
  - Table of numbers? Entries? Two-Dimensional, m X n
  - Many arithmetic operations available
  - Convenient for representing correlated data

- **In Python what is a list of lists?**
  - What is an array? Multi-dimensional?

If we want to find the best match…

- **Assign a number to each possible match**
  - What's the best match among the numbers?
  - What about finding the nearest k and using them?

- **We need to sort the data to find closest matches**
  - Differences between lst.sort() and sorted(lst)
  - How do we change criteria used in sorting?
  - What sorting algorithm is used, do we need to know this?

- **How do assign a number to each possible match**
  - Rankings, ratings, questions, indirect measures, …

Recommender assignment

- **Read data into standard format**
  - Different types of data file, standard list and dictionary
  - Issues in reading multi-line data? Alternatives?

- **Understanding how to measure "closeness"**
  - Examples use math, closeness in metric space, what does this mean?
  - How do we weight measures to get better 'closeness'?

- **How do we know if we've done this correctly, lots of data**

Daphne Koller

Computers learn to diagnose breast cancer?

- **AI and Machine Learning**
  - http://www.youtube.com/watch?v=DX2_9ntOcQI
  - http://med.stanford.edu/ism/2011/november/computer.html

"But to practice problem-solving, a student must first master certain concepts. By providing a cost-effective solution for this first step, we can focus precious classroom time on more interactive problem-solving activities that achieve deeper understanding — and foster creativity."
Coursera Founder, NY Times, December 5, 2011