CS5824: Advanced Machine Learning

Dawei Zhou

CS, Virginia Tech

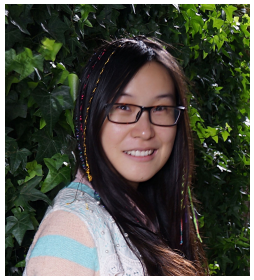
Please keep your face covering on!

COMPUTER SCIENCE
VIRGINIA TECH

Course Information

Basic Information

- Instructor: Dr. Dawei Zhou
  - Office hours: T 3:30pm – 5:30pm
  - Location: https://virginiatech.zoom.us/j/5543732127
  - Email: zhoud@vt.edu
- Lecture
  - Time: TR 2:00pm – 3:15pm
  - Location: SURGE 104A & Zoom
- TA
  - Tong Xiang (xtong@vt.edu)
    - Office hours: W/TH, 3:30 - 4:30 pm
    - Location: https://virginiatech.zoom.us/my/tongxiang
  - Longfeng Wu (longfengwu@vt.edu)
    - Office hours: M, 12:30 - 1:30 pm/ F, 10:00 - 11:00 am
    - Location: https://virginiatech.zoom.us/j/85444737403



Objective and Topics

- Objective
  - An in-depth understanding of machine learning and statistical pattern recognition techniques and their applications.
- Topics
  - Machine learning basics: Probability distributions, MLE, Bias-Variance
  - Machine learning models: regression, classification, decision tree, boosting, kernel methods, clustering, mixture models, graphical models, deep learning
  - Special topics: anomaly detection, text mining, self-supervised learning, fairness learning, and meta learning.

Prerequisites

- VERY IMPORTANT!
- Basics of linear algebra, probability, algorithm design and analysis.
- Proficient in programming in one of the following languages (Python, C++ or Java)
- This is not a programming class!
- Screen-quiz: later TODAY!

Tentative Class Schedule

Machine Learning Basics

Machine Learning Models

Special Topics

Week	Date	Lecture	Homework	Project
1	1/18	Introduction & Screen Quiz (30 min)		
	1/20	Probability Basics		
2	1/25	Probability Basics	HW1 out	
	1/27	Machine Learning from the Data Perspective		
3	2/1	Workshop #1: Data Preprocessing & Machine Learning Toolboxes		Project Proposal Due
	2/3	MLE		
4	2/8	Linear Regression		
	2/10	Bias-Variance Tradeoff		
5	2/15	Classification: Basic Concepts	HW1 due, HW2 out	
	2/17	Classification: Basic Concepts		
6	2/22	Classification: Basic Concepts		
	2/24	Classification: Advanced Method		
7	3/1	Classification: Advanced Method		
	3/3	Boosting, Instance-based Learning		
8	**Spring Break**			
9	3/15	Midterm #1 (75 min)		
	3/17	Cluster Analysis	HW2 due, HW3 out	
10	3/22	Cluster Analysis		
	3/24	Graphical Models		
11	3/29	Hidden Markov Models		
	3/31	Deep Learning		
12	4/5	Deep Learning		
	4/7	Workshop #2: Deep Learning Platforms (Pytoch, Tensorflow)		
13	4/12	Midterm #2 (75 min)	HW3 due	
	4/14	Special Topics: Self-Supervised Taxonomy Enrichment		
14	4/19	Special Topics: Rare Category Analysis		
	4/21	Special Topics: Fairness Machine Learning		
15	4/26	Special Topics: Out-of-Distribution Generalization		
	4/28	Special Topics: Graph Generative Model		
16	5/3	Poster Session for Course Project		
17	5/9	Final Week (No Class)		Final Report Due

Tentative Class Schedule

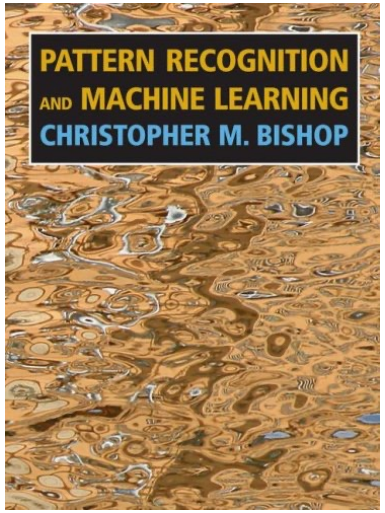
Online

In-Person

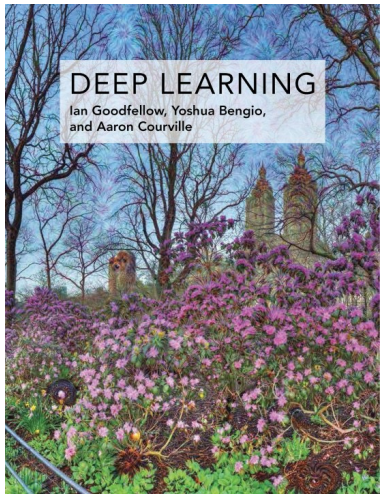
Week	Date	Lecture	Homework	Project
1	1/18	Introduction & Screen Quiz (30 min)		
	1/20	Probability Basics		
2	1/25	Probability Basics	HW1 out	
	1/27	Machine Learning from the Data Perspective		
3	2/1	Workshop #1: Data Preprocessing & Machine Learning Toolboxes		Project Proposal Due
	2/3	MLE		
4	2/8	Linear Regression		
	2/10	Bias-Variance Tradeoff		
5	2/15	Classification: Basic Concepts	HW1 due, HW2 out	
	2/17	Classification: Basic Concepts		
6	2/22	Classification: Basic Concepts		
	2/24	Classification: Advanced Method		
7	3/1	Classification: Advanced Method		
	3/3	Boosting, Instance-based Learning		
8	**Spring Break**			
9	3/15	Midterm #1 (75 min)		
	3/17	Cluster Analysis	HW2 due, HW3 out	
10	3/22	Cluster Analysis		
	3/24	Graphical Models		
11	3/29	Hidden Markov Models		
	3/31	Deep Learning		
12	4/5	Deep Learning		
	4/7	Workshop #2: Deep Learning Platforms (Pytoch, Tensorflow)		
13	4/12	Midterm #2 (75 min)	HW3 due	
	4/14	Special Topics: Self-Supervised Taxonomy Enrichment		
14	4/19	Special Topics: Rare Category Analysis		
	4/21	Special Topics: Fairness Machine Learning		
15	4/26	Special Topics: Out-of-Distribution Generalization		
	4/28	Special Topics: Graph Generative Model		
16	5/3	Poster Session for Course Project		
17	5/9	Final Week (No Class)		Final Report Due

Textbook

- Pattern Recognition and Machine Learning, Christopher M. Bishop, 2006.

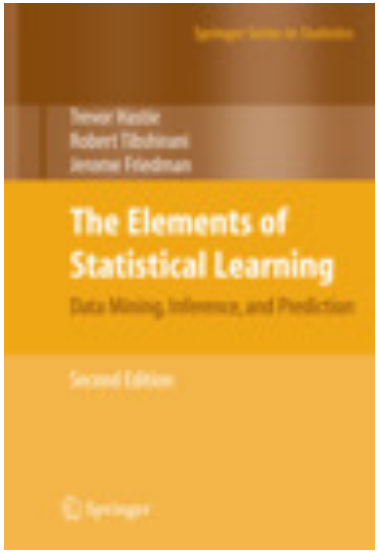


- Deep Learning, Book by Aaron Courville, Ian Goodfellow, and Yoshua Bengio, 2015



Reference Books

- The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Second Edition) by Trevor Hastie, Robert Tibshirani and Jerome Friedman (2009)



http://www-stat.stanford.edu/~hastie/Papers/ESLII.pdf

Grading

- Project: 25%
  - Each group up to 8 students
  - Project proposal (5%)
  - Github (5%)
  - Group presentation (5%)
  - Final project report (10%)
- Homework (3): 30%
  - Equal weights
  - Each student should finish ***independently***
- Exam (2): 40%
  - March 3 (20%), April 7 (20%)
  - Closed book for all exams: Cheat sheet allowed
- Class participation: 5%
  - Students are required to attend the lectures and participate in the class discussion
- A+; A: 90—100, A-: 85—89, B+: 80—84, B: 70—79, C: 60—70

Class Project

- Project proposal (1 page) due on **2/1/22** at 11:59pm
  1) Project title
  2) Team members: roles of each member
  3) Description of the problem you try to address
  4) Preliminary plan (milestones)
  5) Paper list
  6) Template (https://www.overleaf.com/project/61e2e86166cd5f51b36d8156)
- Final project report (10-15 pages) due on **5/9/22** at 11:59pm
  1) Introduction: a summary of the problem, previous work, methods, and results
  2) Problem description: a detailed description of the problem you try to address
  3) Methodology: a detailed description of methods used
  4) Results: a detailed description of your observations from the experiments
  5) Conclusions and future work: a brief summary of the main contributions of the project and the lessons you learn from the project, as well as a list of some potential future work.
  6) Template (https://www.overleaf.com/project/61e2e86166cd5f51b36d8156)

Start to form your group today!

Academic Integrity

- Violations of the University Academic Integrity policy will not be tolerated. Penalties include reduced or 0 credit for submitted work, a failing grade in the class, a note on your official transcript that shows you were punished for cheating, suspension, expulsion and revocation of already awarded degrees. The university requires that should I implement any of these penalties, I must report the matter to the Dean's office. The university academic integrity policy can be found at https://honorsystem.vt.edu/.

Title IX and VT Policy

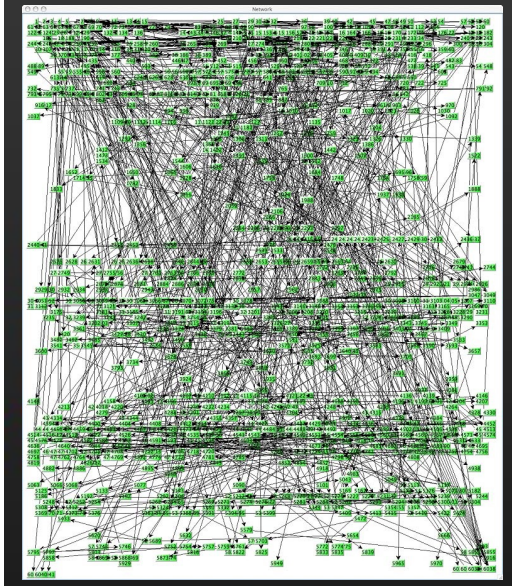
- Title IX is a federal law that provides that no person be excluded on the basis of sex from participation in, be denied benefits of, or be subjected to discrimination under any education program or activity. Both Title IX and university policy make clear that sexual violence and harassment based on sex is prohibited. An individual who believes they have been subjected to sexual violence or harassed on the basis of sex can seek support, including counseling and academic support, from the university. If you or someone you know has been harassed on the basis of sex or sexually assaulted, you can find information and resources at https://oea.vt.edu/harassment-discrimination.html.

Disclaimer

- Some lecture notes are modified based on the slides made by Dr. Jingrui He (with permission) from University of Illinois at Urbana-Champaign.

What is ***Machine Learning***?

Machine Learning



what society thinks I do

what my friends think I do

what my parents think I do

$$L_p = \tfrac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l} \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{w} + b) + \sum_{i=1}^{l} \alpha_i$$

$$\alpha_i \geq 0, \forall i$$

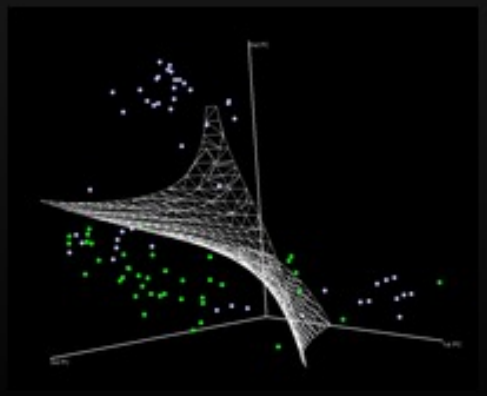
$$\mathbf{w} = \sum_{i=1}^{l} \alpha_i y_i \mathbf{x}_i, \sum_{i=1}^{l} \alpha_i y_i = 0$$

$$\nabla \hat{g}(\theta_t) = \frac{1}{n} \sum_{i=1}^{n} \nabla \ell(x_i, y_i; \theta_t) + \nabla r(\theta_t).$$

$$\theta_{t+1} = \theta_t - \eta_t \nabla \ell(x_{i(t)}, y_{i(t)}; \theta_t) - \eta_t \cdot \nabla r(\theta_t)$$

$$\mathbb{E}_{i(t)}[\ell(x_{i(t)}, y_{i(t)}; \theta_t)] = \tfrac{1}{n} \textstyle\sum_i \ell(x_i, y_i; \theta_t).$$

what other programmers think I do



what I think I do

```
>>> from scipy import SVM
```

what I really do

Machine Learning

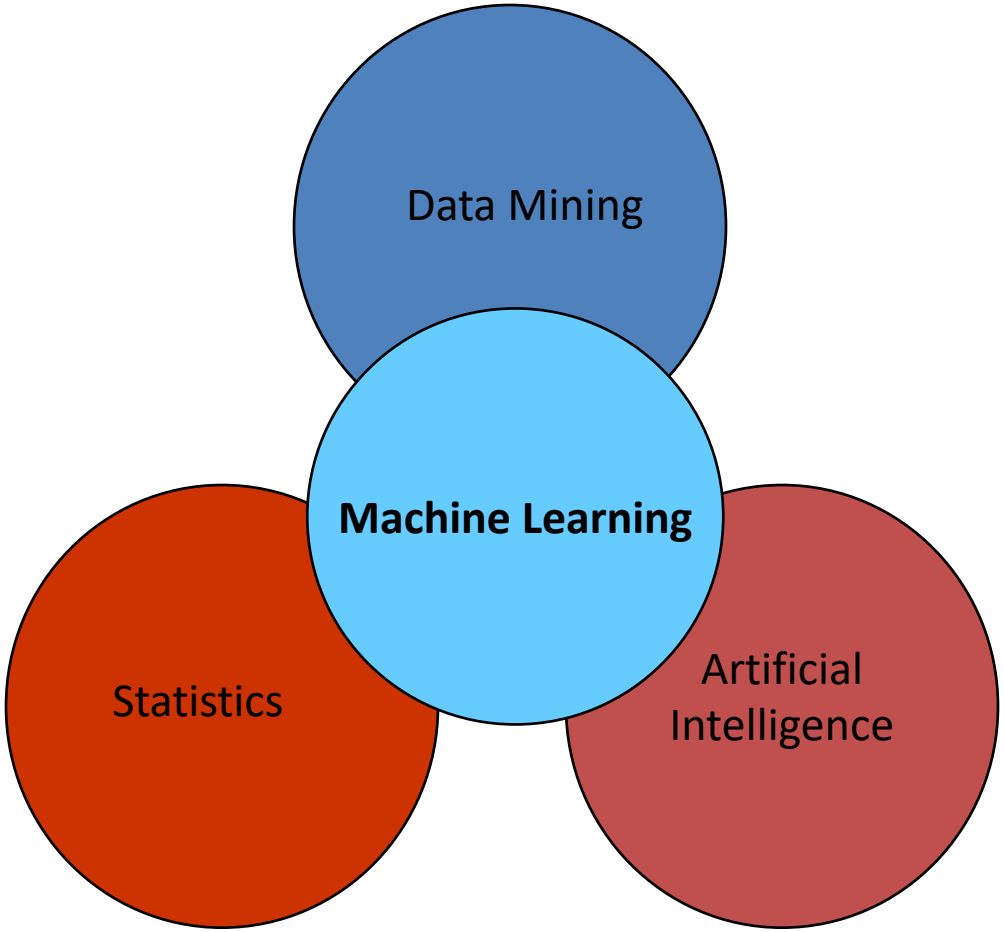
- Prof. Tom Mitchell@CMU

  'Study of algorithms that improve their performance, at some task, with experience'

- Prof. Andrew Ng@Stanford

  'Machine learning is the science of getting computers to act without being explicitly programmed'

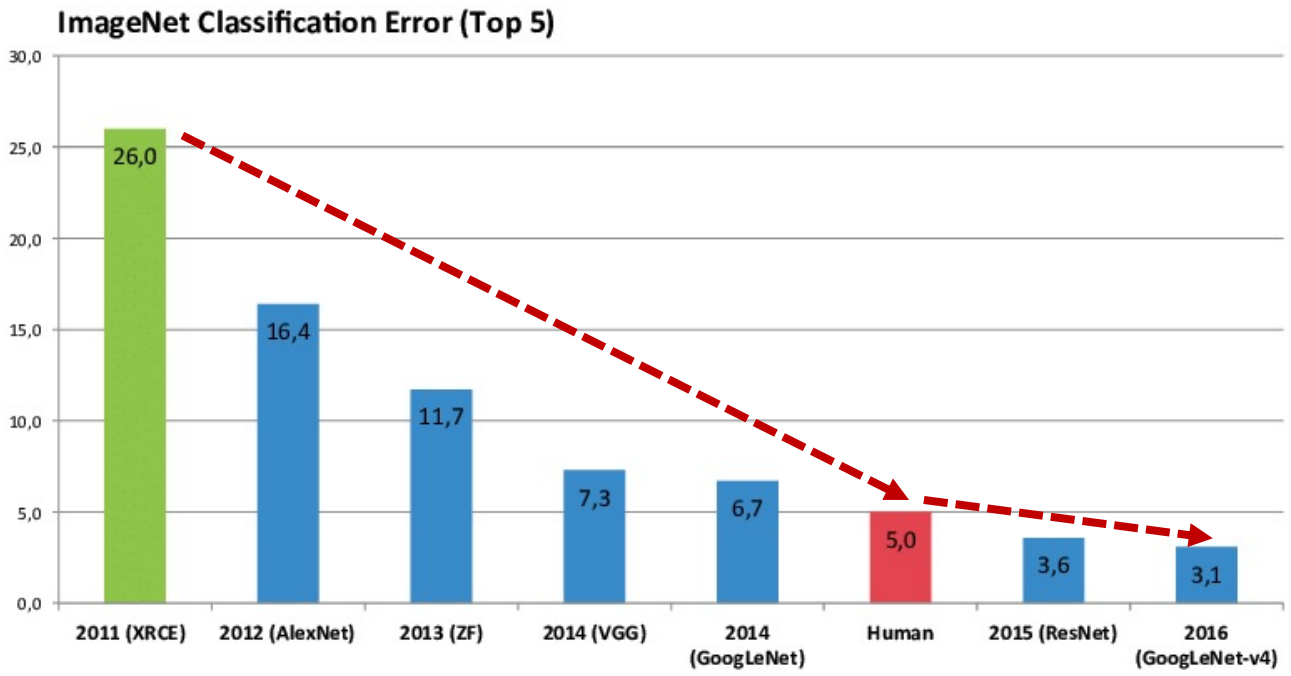
Related Fields



Big Progress in the Past Decades

- The success of large-scaled machine learning in computer vision:

1 million images with 1K categories



[Deng et al. 09]

Big Progress in the Past Decades

- The success of large-scaled machine learning in acoustic speech processing

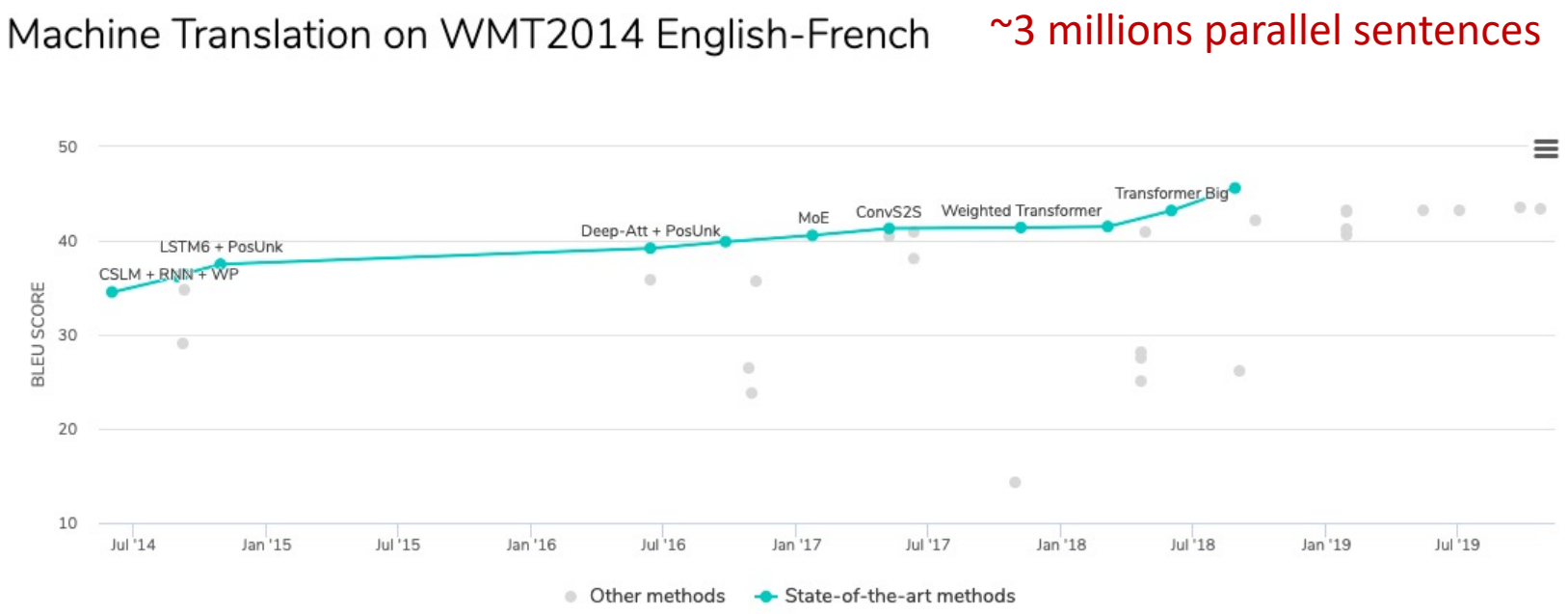
[TABLE 1] COMPARISONS AMONG THE REPORTED SPEAKER-INDEPENDENT (SI) PHONETIC RECOGNITION ACCURACY RESULTS ON TIMIT CORE TEST SET WITH 192 SENTENCES.

METHOD	PER
CD-HMM [26]	27.3%
AUGMENTED CONDITIONAL RANDOM FIELDS [26]	26.6%
RANDOMLY INITIALIZED RECURRENT NEURAL NETS [27]	26.1%
BAYESIAN TRIPHONE GMM-HMM [28]	25.6%
MONOPHONE HTMS [29]	24.8%
HETEROGENEOUS CLASSIFIERS [30]	24.4%
MONOPHONE RANDOMLY INITIALIZED DNNs (SIX LAYERS) [13]	23.4%
MONOPHONE DBN-DNNs (SIX LAYERS) [13]	22.4%
MONOPHONE DBN-DNNs WITH MMI TRAINING [31]	22.1%
TRIPHONE GMM-HMMs DT W/ BMMI [32]	21.7%
MONOPHONE DBN-DNNs ON FBANK (EIGHT LAYERS) [13]	20.7%
MONOPHONE MCRBM-DBN-DNNs ON FBANK (FIVE LAYERS) [33]	20.5%
MONOPHONE CONVOLUTIONAL DNNs ON FBANK (THREE LAYERS) [34]	20.0%

[Hinton et al. 12]

Big Progress in the Past Decades

- The success of large-scaled machine learning in natural language processing:



[Cho et al. 2014; Devlin et al. 2014]

Useful Resources

- The discipline of machine learning: http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf
- Coursera: https://www.coursera.org/course/ml
- Alex Smola@CMU's machine learning lectures: https://www.youtube.com/playlist?list=PLZSO_6-bSqHQmMKwWVvYwKreGu4b4kMU9
- Ben Taskar@UW's tutorial: https://alliance.seas.upenn.edu/~cis520/wiki/index.php?n=Recitations.MatlabTutorial
- Probability review by David Blei@Princeton: http://www.cs.princeton.edu/courses/archive/spring07/cos424/scribe_notes/0208.pdf

Seriously

- Do a great job in CS 5824!
- Read many many many … many papers
- Publish many many many … many papers
  - ICML: https://icml.cc/Conferences/2021
  - NeurIPS: https://nips.cc/
  - ICLR: https://openreview.net/group?id=ICLR.cc/2021/Conference
  - IJCAI: https://ijcai-21.org/
  - AAAI: https://aaai.org/Conferences/AAAI-21/
  - ACM KDD: https://kdd.org/kdd2021/
  - ICDM: http:/https://icdm2021.auckland.ac.nz
  - WWW: https://www2021.thewebconf.org/
  - Journal of Machine Learning Research: http://jmlr.org/
  - IEEE Transactions on Knowledge and Data Engineering: http://www.computer.org/portal/web/tkde

Who Wants Machine Learning People?

- IT
  - Outlier/fraud detection
  - Web image search
  - Recommendation
  - Information filtering
  - Community detection
  - Ad placement
  - Sentiment analysis
  - …

- Companies
  - Meta, Google, LinkedIn, Twitter, Microsoft, IBM, AT&T, Apple, Amazon, Siemens, Foursquare, Yelp, Walmart Lab, NEC, Generic Electric, Baidu, Samsung, …

Who Wants Machine Learning People?

- Finance
  - Stock market prediction
  - Algorithmic trading
  - Return forecasting
  - …

- Companies
  - Goldman Sachs, Morgan Stanley, American Express, Citadel LLC, Barclays Capital, Rotella Capital Management, Citi Bank, Pequot Capital, Zestfinance, Federal Reserve Board, WorldQuant LLC, …

Who Wants Machine Learning People?

- Medical outcomes analysis
- Robot control
- Computational biology
- Sensor networks
- …

Required Reading

- PRML – Sec 1.2
- Deadline: 1/20