Lecture Notes for

# Machine Learning in Python

Professor Eric Larson

## Final Lecture: Ethics and Retrospective

# Lecture Agenda

- Logistics
  - CNN Grades update
  - RNNs due **Last Day of Finals**
- Agenda
  - RNN town hall
  - Ethical Principles
  - Retrospective and Evaluations

# Class Overview, by topic

Table Data Visualization

Numpy, Pandas, Seaborn

Overviews with some in-depth discussion

Dimension Reduction and Image Processing

Scikit-learn, Scikit Image,

Intuition only, Some mathematics

Linear and Logistic Regression

Numpy, Recreate API for Scikit-learn

Detailed mathematics for simple optimization

intuition for advanced optimization

Neural Networks and Back Prop.

Numpy

Detailed mathematics for NN operations

Wide and Deep Networks

Convolutional Networks

Recurrent Networks

Keras, Tensorflow

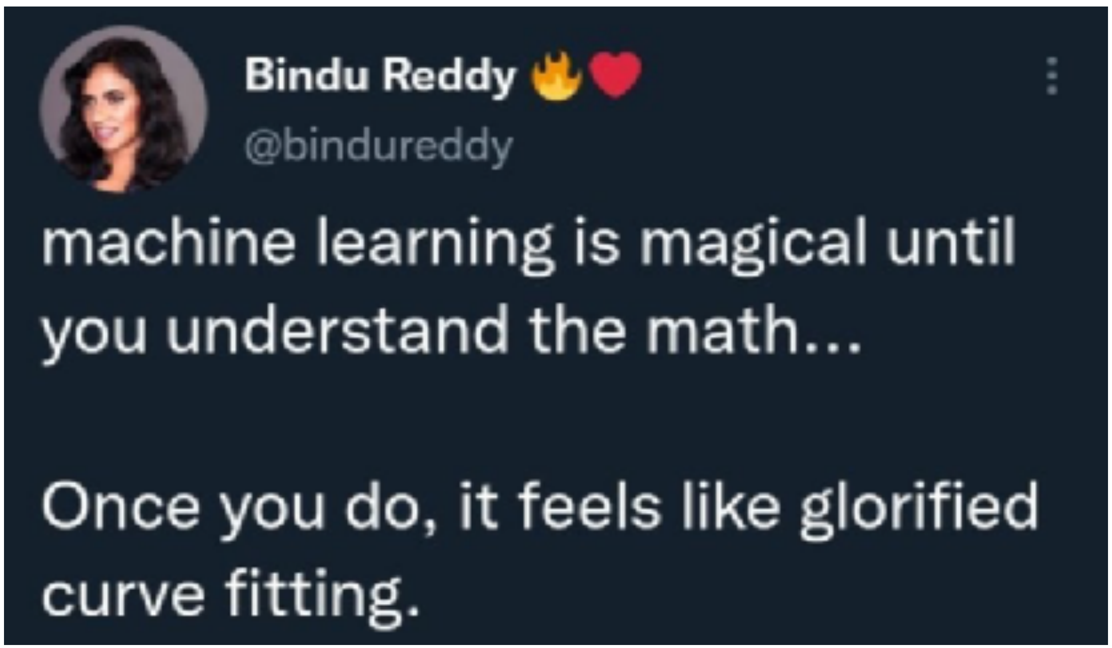
Intuition, Detailed implement.

Ethics in Language Models

ConceptNet

Case studies

# Sequential Networks Town Hall



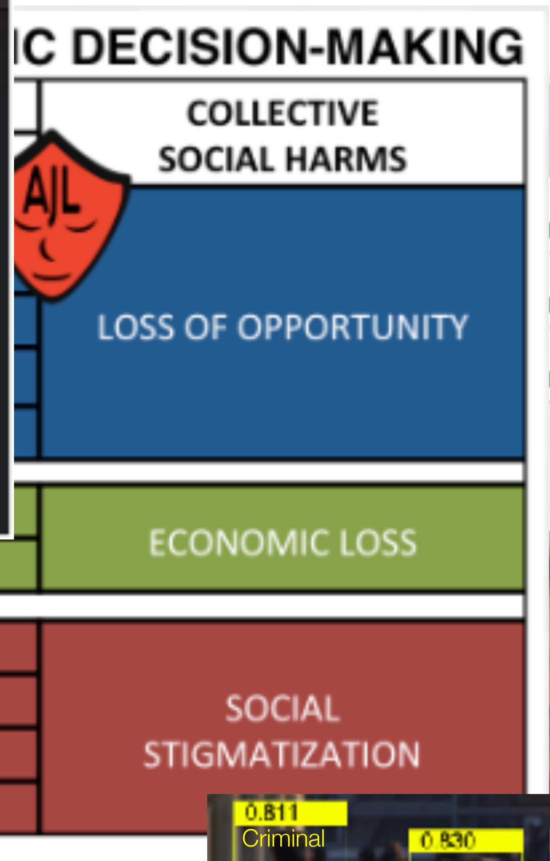

# AI Ethics Principles

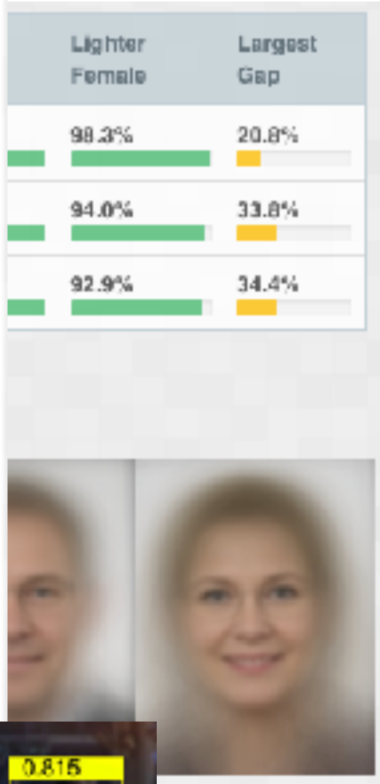
**Janelle Shane** @JanelleCShane · 1d
Predictive policing algorithms don't predict who commits crime. They predict who the police will arrest.

**Emily M. Bender, professionally...** · 11h
"AI" can NOT:
* Predict who will commit a crime

"AI" can:
* Make biased policing look "objective"

# Timnit Gebru: Gender Shades



**Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification***


**Joy Buolamwini** JOYAB@MIT.EDU
*MIT Media Lab 75 Amherst St. Cambridge, MA 02139*

**Timnit Gebru** TIMNIT.GEBRU@MICROSOFT.COM
*Microsoft Research 641 Avenue of the Americas, New York, NY 10011*


http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

# Ethical Principles in ML

*From Australian Goverment, Department of Science*

- **Reliability**: reliably operate in accordance with their intended purpose
- **Beneficence**: individuals, society and the environment.
- **Respect**: respect human rights, diversity, and autonomy of individuals.
- **Fairness**: be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups
- **Privacy**: respect and uphold privacy rights and data protection, and ensure the security of data
- **Transparency**: ensure people know when they are being significantly impacted by an AI system, and can find out when engaging with them
- **Contestability**: should be a timely process to allow people to challenge the use or output of the AI system
- **Accountability**: Those responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and *human oversight* of AI systems should be enabled.

**To enforce these principles a board with autonomy must exist**

https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework/ai-ethics-principles

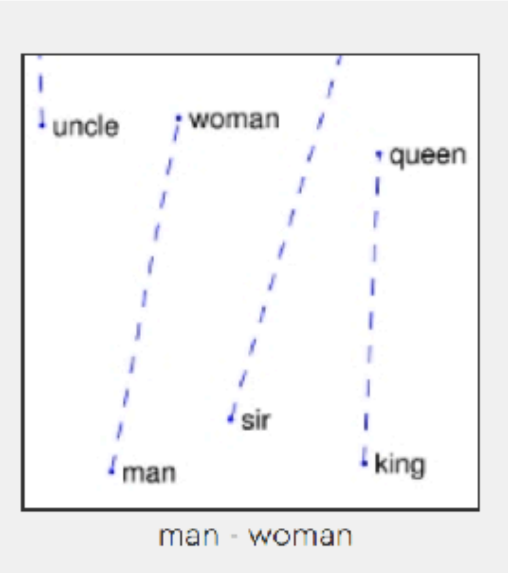
# Bias Case Study in NLP

Timnit Gebru @timnitGebru

I'm sick of this framing. Tired of it. Many people have tried to explain, many scholars. Listen to us. You can't just reduce harms caused by ML to dataset bias.

> Yann LeCun @ylecun · 19h
>
> ML systems are biased when data is biased. This face upsampling system makes everyone look white because the network was pretrained on FlickFaceHQ, which mainly contains white people pics....

# Ethics in Language: Word Embedding Analogy




**Trained on New York Times**

😳

$$W(\text{``woman''}) - W(\text{``man''}) \simeq W(\text{``aunt''}) - W(\text{``uncle''})$$

$$W(\text{``woman''}) - W(\text{``man''}) \simeq W(\text{``queen''}) - W(\text{``king''})$$

$$\overrightarrow{\text{man}} - \overrightarrow{\text{woman}} \approx \overrightarrow{\text{computer programmer}} - \overrightarrow{\text{homemaker}}.$$

**Extreme *she* occupations**

| | | |
|---|---|---|
| 1. homemaker | 2. nurse | 3. receptionist |
| 4. librarian | 5. socialite | 6. hairdresser |
| 7. nanny | 8. bookkeeper | 9. stylist |
| 10. housekeeper | 11. interior designer | 12. guidance counselor |

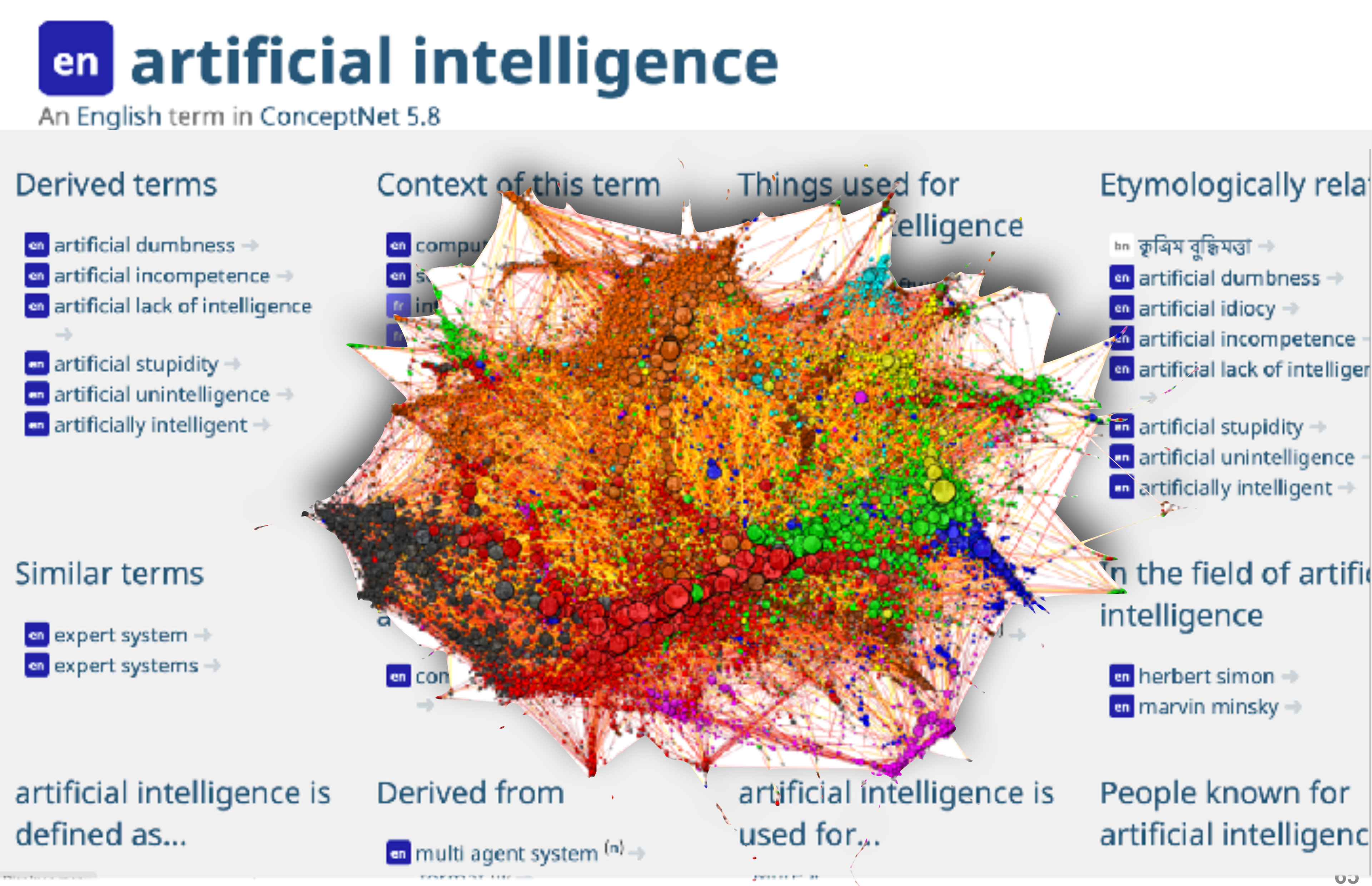
**Extreme *he* occupations**

| | | |
|---|---|---|
| 1. maestro | 2. skipper | 3. protege |
| 4. philosopher | 5. captain | 6. architect |
| 7. financier | 8. warrior | 9. broadcaster |
| 10. magician | 11. figher pilot | 12. boss |

https://nlp.stanford.edu/projects/glove/

Bolukbasi et al., NeurIPs 2016
https://arxiv.org/pdf/1607.06520.pdf

# ConceptNet, a Multi-lingual Knowledge Graph



# ConceptNet Numberbatch


- **Step One**: Create a Knowledge Graph (from multiple sources with relations like *UsedFor*, *PartOf*, *etc*.)
- **Step Two**: Based on this KG, perturb existing embeddings (like GloVe) to minimize:

$$\Psi(Q) = \sum_{i=1}^{n} \left[ \alpha_i \| q_i - \hat{q}_i \|^2 + \sum_{(i,j) \in E} \beta_{ij} \| q_i - q_j \|^2 \right]$$

new embed old embed neighbors from KG

(keep similar to original) (make similar according to other knowledge)

- Straight forward  to optimize the objective by averaging neighbors in the ConceptNet Knowledge Graph
- Multiple embeddings achieved by merging through “retrofitting” which projects onto a shared matrix space (with SVD)

ConceptNet 5.5: An Open Multilingual Graph of General Knowledge, Speer et al., 2017

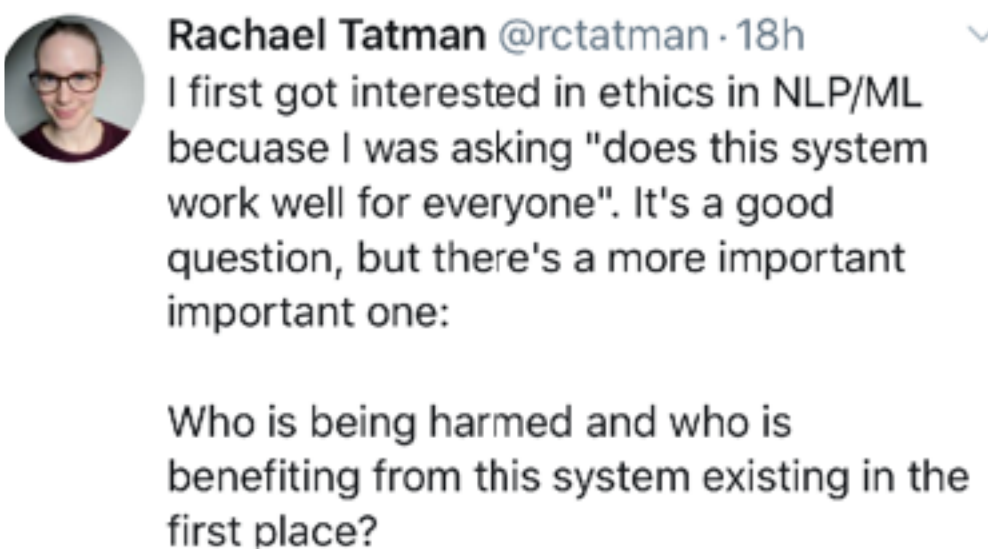
# Lightning Demo (or self guided demo)

## How to Make a Racist AI without Really Trying

Robyn Speer, 2017
http://blog.conceptnet.io/posts/2017/how-to-make-a-racist-ai-without-really-trying/

**Debiasing: Man is to Computer Programmer as Woman is to Homemaker? De-biasing Word Embeddings**
Bolukbasi et al., NeurIPs 2016
https://arxiv.org/pdf/1607.06520.pdf

**ConceptNet 5.5: An Open Multilingual Graph of General Knowledge**
Speer et al., AAAI 2017
https://arxiv.org/pdf/1612.03975.pdf



François Chollet @fchollet · 11h

When faced with tech ethics problems, you can either ask hard questions, seek solutions, and take responsibility, or you

Devin Guillory @databoydg · 13h

Watching one of the most influential

## Timnit Gebru

A lot of times, people are talking about bias in the sense of equalizing performance across groups. They're not thinking about the underlying foundation, whether a task should exist in the first place, who creates it, who will deploy it on which population, who owns the data, and how is it used?

The root of these problems is not only technological. It's social. Using technology with this underlying social foundation often advances the worst possible things that are happening. In order for technology not to do that, you have to work on the underlying foundation as well. You can't just close your eyes and say: "Oh, whatever, the foundation, I'm a scientist. All I'm going to do is math."

# Course Retrospective

- AI winters exi
machine lear
  - **and histo**
- Formal meth
- At the end of
- **Open sourc**
advancemen
  - http://ww

**Leading ML researchers issue statement of support for JMLR**

From: Michael Jordan [mailto:jordan@CS.Berkeley.EDU]
Sent: Monday, October 08, 2001 5:33 PM
Subject: letter of resignation from Machine Learning journal

Dear colleagues in machine learning,

The forty people whose names appear below have resigned from the Editorial Board of the Machine Learning Journal (MLJ). We would like to make our resignations public, to explain the rationale for our action, and to indicate some of the implications that we see for members of the machine learning community worldwide.

The machine learning community has come of age during a period of enormous change in the way that research publications are circulated. Fifteen years ago research papers did not circulate easily, and as with other research communities we were fortunate that a viable commercial publishing model was in place so that the fledgling MLJ could begin to circulate. The needs of the community, principally those of seeing our published papers circulate as widely and rapidly as possible, and the business model of commercial publishers were in harmony.

Times have changed. Articles now circulate easily via the Internet, but unfortunately MLJ publications are under restricted access. Universities and research centers can pay a yearly fee of $1050 US to obtain unrestricted access to MLJ articles (and individuals can pay $120 US). While these fees provide access for institutions and individuals who can afford them, we feel that they also have the effect of limiting contact between the current machine learning community and the potentially much larger community of researchers worldwide whose participation in our field should be the fruit of the modern Internet.

None of the revenue stream from the journal makes its way back to authors, and in this context authors should expect a particularly favorable return on their intellectual contribution---they should expect a service that maximizes the distribution of their work. We see little benefit accruing to our community from a mechanism that ensures revenue for a third party by restricting the communication channel between authors and readers.

Sincerely yours,

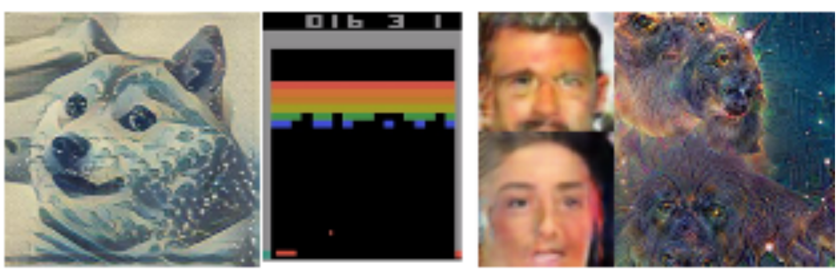
Chris Atkeson
Peter Bartlett
Andrew Barto
Jonathan Baxter
Yoshua Bengio
Kristin Bennett
Chris Bishop
Justin Boyan
Carla Brodley
Claire Cardie
William Cohen
Peter Dayan
Tom Dietterich
Jerome Friedman
Nir Friedman
Zoubin Ghahramani
David Heckerman
Geoffrey Hinton
Haym Hirsh
Tommi Jaakkola
Michael Jordan
Leslie Kaelbling
Daphne Koller
John Lafferty
Sridhar Mahadevan
Marina Meila
Andrew McCallum
Tom Mitchell
Stuart Russell
Lawrence Saul
Bernhard Schoelkopf
John Shawe-Taylor
Yoram Singer
Satinder Singh
Padhraic Smyth
Richard Sutton
Sebastian Thrun
Manfred Warmuth
Chris Williams
Robert Williamson

# Topics review

- Data **munging** in pandas and numpy and **visualization** with matplotlib, pandas, seaborn
- Data preprocessing: **dim reduction**, images, text, categorical features, **embeddings**
- **Linear models**: linear regression, logistic regression, simple neural networks
- **Optimization** strategies: Gradient ascent, Quasi-Newton, Extensions of SGD (RMSProp, AdaM)
- **Back propagation** in MLP (from scratch)
- Tensorflow/Keras for **wide and deep networks**
- **Convolutional** neural networks (up to modern day)
- **Recurrent** neural networks (scratched surface only)

# Topics Not Covered

- Transfer/Multi-Task Learning
- Visualizing Deep Convolutional Networks
- Fully Convolutional Networks
- Style Transfer
- Generative Adversarial Networks
- (*partial*) Reinforcement Learning



**Syllabus for CSE8321: Machine Learning and Neural Networks**

**Course Schedule**

| Week | Lecture A | Lecture B | Lab Due |
|---|---|---|---|
| 1 | Lecture: Course Introduction and Syllabus | Lecture: [illegible] Neural Networks | |
| 2 | [illegible] Lecture: [illegible] Overview | Lecture: CNN Visualization Methods Reading: Chapter 5, Section 4 | |
| 3 | Lecture/Demo: CNN Visuals | Lecture: Image Style Transfer Overview Reading: Chapter [illegible] | Lab One: CNN Visuals |
| 4 | Student Presentations: Reading: [illegible] Lecture: [illegible] and Content Loss | Student Presentations: Reading: [illegible] Lecture: [illegible] | |
| 5 | Lecture/Demo: Image Style Transfer in Keras | Lecture: Transfer Learning in CNNs Reading: Chapter 5, Section 2 and 3 | Lab Two: Style Transfer |
| 6 | Lecture/Demo: [illegible] | Lecture: Multi-modal Learning Overview | |
| 7 | Student Presentations: Reading: [illegible] Lecture/Demo: [illegible] | Student Presentations: Reading: An Overview of Multi-Task Learning [illegible] Lecture/Demo: [illegible] | |
| 8 | Lecture: Generative Adversarial Networks Overview Reading: Chapter [illegible] | Student Presentations: Reading: [illegible] Lecture: Problems with GANs | Lab Three: Multimodal and Multitask |
| 9 | Student Presentations: Reading: [illegible] Lecture: [illegible] | Lecture/Demo: GAN training on [illegible] | |
| 10 | Lecture: Deep Reinforcement Learning Overview | Lecture: Q-Learning | Lab Four: GANs |
| 11 | Lecture: Trust Region Policy Optimization | Lecture: Trust Region Policy Optimization, cont. | |
| 12 | Student Presentations: Reading: [illegible] Lecture/Demo: [illegible] | Lecture/Demo: Training Deep Reinforcement Learning [illegible] | |
| 13 | Lecture: The Future of Deep Learning Reading: Chapter 9 | Lecture: Overflow | Lab Five: Reinforcement Learning |
| 14 | Work on Final Project/Paper | Work on Final Project/Paper | |
| 15 | Work on Final Project/Paper | Work on Final Project/Paper | Research Paper Due |

**Syllabus for CSE8321: Machine Learning and Neural Networks**

**Overview**

This course extends basic knowledge of the use of Neural Networks in machine learning beyond simple prediction, especially targeted outputs that are generation or alteration of images, text, and audio. This course emphasizes topics of neural networks in the "deep learning" subdomain. This course will survey of important topics and current areas of research, including transfer learning, multi-task and multi modal learning, image style transfer, neural network visualization, deep convolutional generative adversarial networks, and deep reinforcement learning. For grading, students are expected to complete smaller team-based projects throughout the semester, present one research paper in a 15-20 minute group presentation (covering topics in the course), and complete a comprehensive final project that involves a number of different deep learning architectures.

# Thank you for a great semester!

- but it could **have been better** somehow, right?
  - how could you learn better, more reliably for an interview?
  - what should **not be cut** or **not changed**?
  - **Already cut**: SVMs, Ensembles, Transformers, many-to-many RNNs,
  - More RNNs? Less RNNs? No RNNs?
  - More convolutional approaches/depth?
  - More APIs? Turi / PyTorch?
  - More flipped Assignments?
  - Self-guided Jupyter notebooks?

# Thank You for an Excellent Semester!


Courtesy of Omar Roa

**Please fill out the course evaluations!!!!**