10701 Introduction to Machine Learning (PhD)

Spring 2019

Class Information

http://cs.cmu.edu/~lwehbe/10701_S19

- You can also find the link on my webpage

Make sure to sign up for Piazza

- https://piazza.com/cmu/spring2019/10701
- (also on the course website)
- Will be used for course communication, in class surveys, questions about homework.

Leila Wehbe

- Assistant Professor in MLD
- Interests:
  - Computational Neuroscience
  - Natural Language Processing
  - Scientific Reproducibility

- Contact Info:
  - lwehbe@cmu.edu
  - Office: GHC 8217
  - Office hours: Tuesday 2-3pm

Brynn Edmunds

Education Associate for MLD

Contact Info:

- bedmunds@andrew.cmu.edu
- Office: GHC 8110
- Office hours: TBA

Byungsoo Jeon (byungsoj@andrew)

Fun Fact

✓ I believe human learning is more important than machine learning

Research Interest

✓ **Educational Data Science**

✓ Deep Learning

✓ Reinforcement Learning

Office Hours

✓ TBD

Gi Bum Kim (gibumk@andrew)

- Office: GHC 7409
- Office hrs: TBD
- Interests: computational genomics, graphical models

Mauro Moretto (mmoretto@andrew.cmu.edu)

4th Year PhD Student in Economics (@Tepper)

Research Interest

✓Data Science

✓ML Applications to Economics and Finance

✓Hierarchical Models

Office Hours

✓Tuesday 9-10 am

✓Location: TBD

Sarah Mallepalle (smallepa@andrew)

Fun Fact

✓ I start working for the Philadelphia Phillies' Baseball Research & Development Office in June!

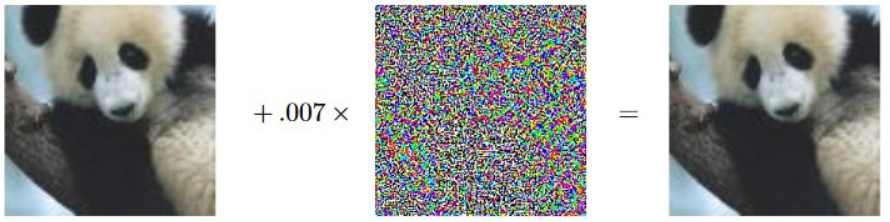
Research Interests

✓ Sports Analytics

✓ Generative Adversarial Networks

✓ Adversarial Attacks

Office Hours: TBD



Yimeng Zhang
(yimengzh@cs.cmu.edu)

Academic Interests

✓ Computational Neuroscience

✓ CV/ML/DL

✓ Writing libraries & toolboxes

Hobbies

✓ Music & Video games

Office Hours

✓ TBD

zym1010.github.io

Ziheng (George) Cai (zcai@andrew)

Fun Fact:

✓ Studying Machine Learning gives me better understanding of my own learning process

Research Interests:

✓ Deep learning

✓ Generative Models

Office Hours:

✓ TBD

Class times

MW lecture 10:30am to 11:50 am, Rashid Auditorium

F recitation 10:30am to 11:50 am, Rashid Auditorium

Office hours scattered through the week.

Just for this week the TA will have office hours during recitation (Friday 1/18 at 10:30am in Rashid Auditorium)

Office hours will start normally on 1/21.

Class structure

- 4 Homework assignments (each 10%)
- Midterm (after spring break, 15%)
- Project in groups of 3-5 (25%)
- Final (will be scheduled by the university,15%)
- Participation (5%)

Schedule

Date	*Note*	*Topic*	*Resources*
		Basics	
01/14		Lecture 1: Introduction - What is Machine Learning	
01/16	HW1 Out	Lecture 2: Building blocks - MLE, Bayesian Estimation, MAP, Decision Theory, Model-free, Risk Minimization	
01/18	Special Office Hours		
01/21	MLK day, no class		
01/23		Lecture 3: Classification, kNN	
01/25		Recitation	
		Parametric Estimation and Prediction	
01/28		Lecture 4: Linear Regression, Regularization	
01/30	HW1 due, HW2 Out	Lecture 5: Logistic Regression	
02/01		Recitation	
02/04		Lecture 6: Naive Bayes	
02/06		Lecture 7: Disciminative vs generative	
02/08		Recitation	
02/11		Lecture 8: Decision Trees	
02/13	HW2 due, HW3 Out	Lecture 9: Neural Networks (perceptron, neural nets)	
02/15		Recitation	
02/18	Project Topic Selection	Lecture 10: Neural Networks (deep nets, backprop)	
02/20		Lecture 11: SVMs	
02/22		Recitation	
02/25	Course Drop Deadline	Lecture 12: SVMs	
02/27	HW3 due	Lecture 13: Boosting	
03/01		Recitation	

Schedule

Learning Theory

03/04		Lecture 14: Generalization, Model selection
03/06	Midway Report Due	Lecture 15: Learning Theory
03/08	Mid-Semester Break, no class	
03/11	Spring break, no class	
03/13	Spring break, no class	
03/15	Spring break, no class	

Unsupervised Learning

03/18		Lecture 16: Clustering
03/20		Lecture 17: Clustering
03/21		Midterm Exam (Thursday 3/21 6:30pm)
03/22		Recitation
03/25	HW4 Out	Lecture 18: Representation Learning
03/27		Lecture 19: Representation Learning
03/29		Recitation

Graphical Models and Structured Prediction

04/01		Lecture 20: Graphical Models
04/03		Lecture 21: Graphical Models
04/05		Recitation
04/08	HW4 Due	Lecture 22: Graphical Models
04/10		Lecture 23: Sequence Models
04/12	Spring Carnival, No Class	
04/15		Lecture 24: Sequence Models

Schedule

Special Topics

04/17		Lecture 25: Semi-supervised learning
04/19		Recitation
04/22		Lecture 26: Online learning
04/24		Lecture 27: Reinforcement learning
04/26		Recitation
04/29		Lecture 26: Deep Reinforcement learning
05/01	Final Reports Due	Lecture 27: Review, discussion
05/03	Last day of class	

Homework

4 homework assignments submitted through gradescope and autolab:

- allows TAs to grade and also checks for copied code

HW1 will be out on Wednesday, and due two weeks from Wednesday.

Extensions

8 late days total

- Maximum 4 for a single homework

Special cases / emergencies:

- late days are supposed to be used for conferences etc.
- if you have special case, let us know 5 days in advance
- (except for emergencies)

Project

- groups of 3-5
- pick one project from 7 fixed topics
- Feb 18: select group and topic
- March 6: Midway report Due
- Sometime in April: Poster presentation
- May 1st: Final reports Due

No late days for Projects. For special circumstances you should talk to us in advance

Academic Integrity

Collaboration is allowed:

- You must write your answers separately
- **You must always disclose it**

It is explicitly disallowed to look for answers to problems online (for analytical or coding problems)

- If you do look up sources (not the answer) you should disclose it

Disclosing is always the right strategy.

Accommodations for Students with Disabilities

Office of Disability Resources

access@andrew.cmu.edu

Taking care of yourself

It's important to make sure to stay healthy:

- Eat well, sleep enough, exercise, hang out with friends
- Being unhealthy doesn't lead to more productivity on the long run

Take care of your mental health and of others close to you

- Talk to CaPS: 412-268-2922
- Reach out to your friends
- Reach out to me or to one of us

Prerequisites

In the next two week, I advise you to catch up on your prerequisites using the following videos:

- Prof. Zico Kolter's videos on basics of linear algebra:

http://www.cs.cmu.edu/~zkolter/course/linalg/index.html

- Prof. Aaditya Ramdas's videos on some basic and some advanced topics (from when he was a PhD student) :
Multivariate prob/stats: https://www.youtube.com/playlist?list=PLRCdqbn4-qwoRTW3OpaB8-GnQwr6ta756
Multivariate calculus: https://www.youtube.com/playlist?list=PLRCdqbn4-qwoT1uhC7kfIjpPIg28TC0x1
SVD review: https://www.youtube.com/playlist?list=PLRCdqbn4-qwqqC-ksijw-PCQbt92pugvV
Real and functional analysis: https://www.youtube.com/playlist?list=PLRCdqbn4-qwrecUfMtAEhGOfLsnJh_IXY

Ressources

Textbooks:

- [CB] ***Machine Learning: A probabilistic perspective***, Kevin Murphy.
- [KM] ***Pattern Recognition and Machine Learning***, Christopher Bishop.
- [HTF] ***The Elements of Statistical Learning: Data Mining, Inference and Prediction***, Trevor Hastie, Robert Tibshirani, Jerome Friedman.
- [TM] ***Machine Learning***, Tom Mitchell.

Introduction to Machine Learning

It is **not** an introduction to Deep Learning!

- We will cover the basics of Deep Learning,
- And many other areas of Machine Learning!

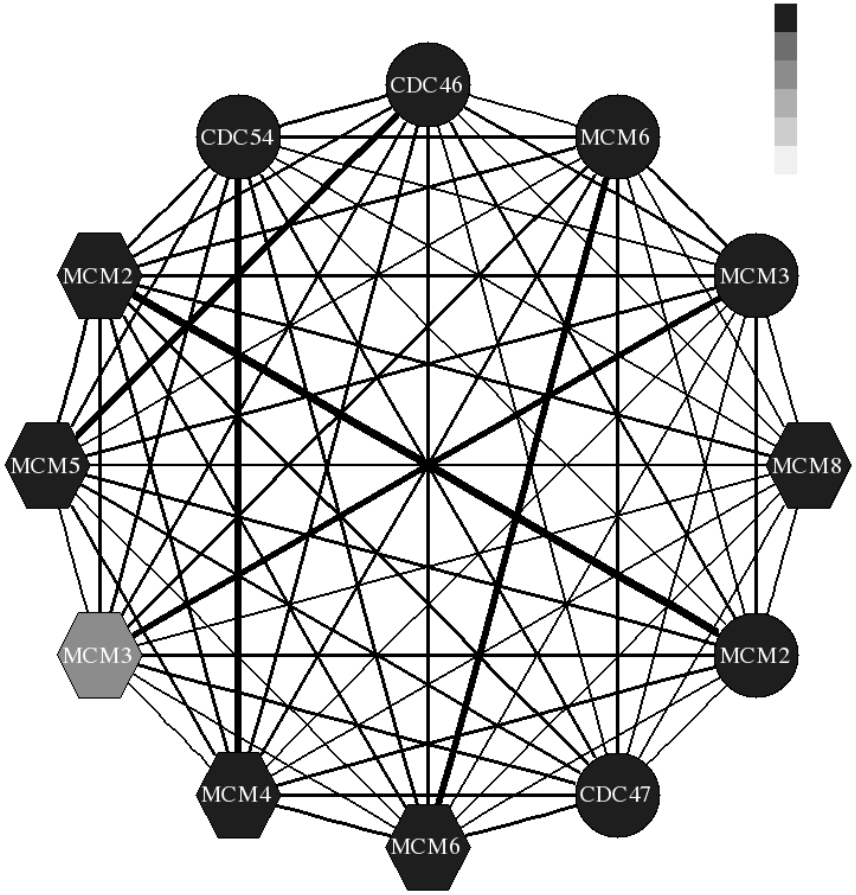
So what is Machine Learning?

Machine Learning

How to learn from experience / data (based on some assumptions) to be able to:

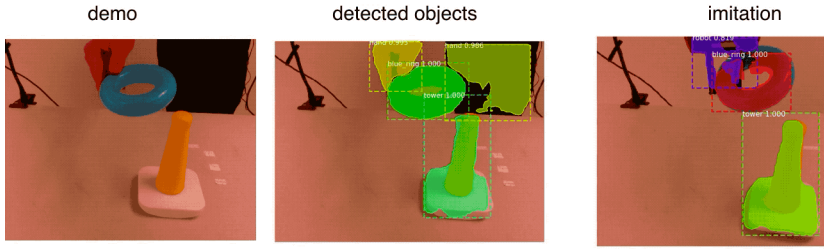
- Perform predictions
- Make decisions and perform actions
- Plan
- Assess risk
- Produce/understand (e.g. understand a scene, produce linguistic descriptions)
- …

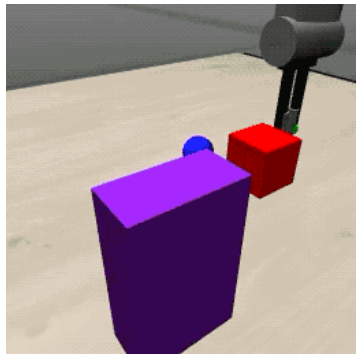
Just a few of the research projects going on in MLD



Conservation of biological systems

(Systems Biology group)





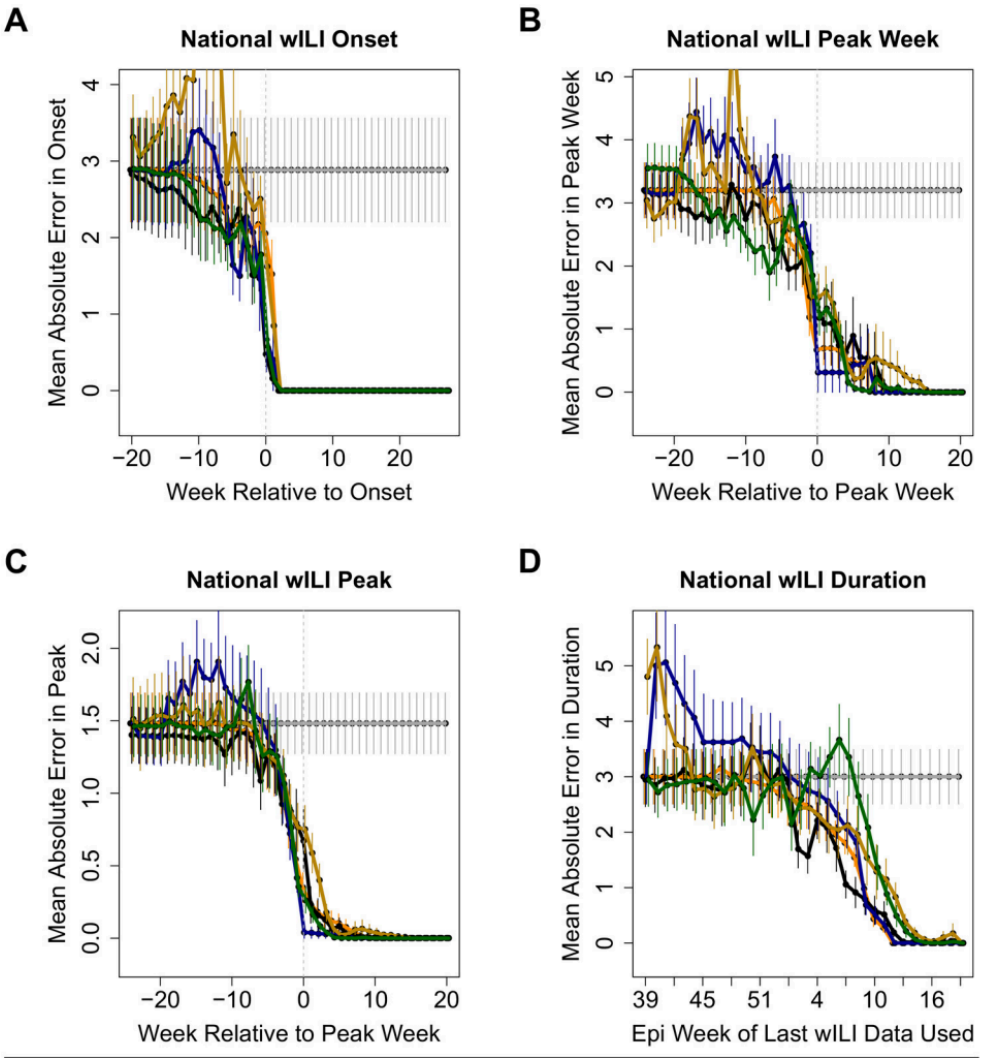
Objects detection and scene understanding for reinforcement learning (Katerina Fragkiadaki)

Just a few of the research projects going on in MLD

Recently-Learned Facts twitter

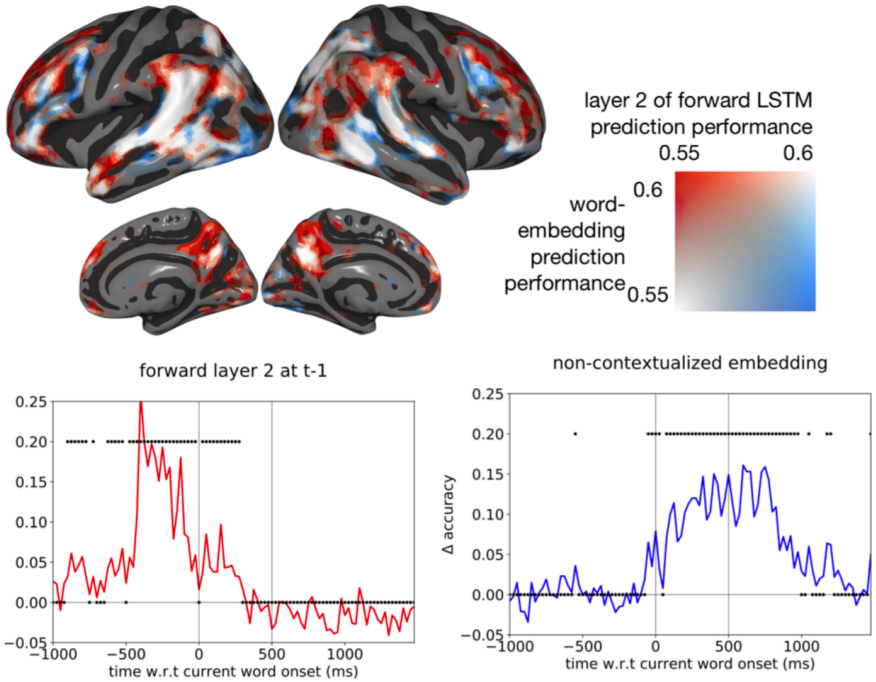
instance
diversion_magazine is a website
barton_road_swing_bridge is a bridge
jennifer_finnegan is a chef
the_summer_smothers_brothers_show is a TV show
anton_refregier is a visual artist
bed is often found in lodge_rooms
us_dept is an organization also known as administration
eu_member_states is a generalization of eastern_european_countries
edinburgh_airport is a building located in the city central_london
miller001 is a person who graduated from the university college

NELL (Never Ending Language Learner)

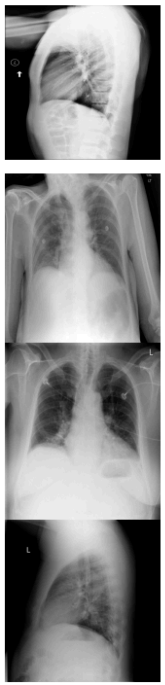


Forecasting epidemics (Delphi Group)

Just a few of the research projects going on in MLD



Aligning NLP Neural Networks and Brain Activity (BrainML group)



Ground Truth

The cardiomediastinal silhouette is within normal limits. **Calcified right lower lobe granuloma**. No focal airspace consolidation. No visualized pneumothorax or large pleural effusion. No acute bony abnormalities.

Exam limited by patient rotation. Mild rightward **deviation of the trachea**. Stable **cardiomegaly**. Unfolding of the thoracic aorta. Persistent right **pleural effusion** with adjacent **atelectasis**. **Low lung volumes**. No focal airspace consolidation. There is severe **degenerative changes of the right shoulder**.

Frontal and lateral views of the chest with overlying external cardiac monitor leads show **reduced lung volumes** with bronchovascular crowding of basilar **atelectasis**. No definite focal airspace consolidation or pleural effusion. **The cardiac silhouette appears mildly enlarged**.

Apparent cardiomegaly partially accentuated by **low lung volumes**. No focal consolidation, pneumothorax or large pleural effusion. Right base **calcified granuloma**. Stable right infrahilar nodular density (lateral view). Negative for acute bone abnormality.

CoAtt [16]

The heart is normal in size. The mediastinum is unremarkable. The lungs are clear.

The heart size and pulmonary vascularity appear within normal limits. The lungs are free of focal airspace disease. No pleural effusion or pneumothorax. No acute bony abnormality.

The heart size and pulmonary vascularity appear within normal limits. The lungs are free of focal airspace disease. No pleural effusion or pneumothorax. no acute bony abnormality.

The heart is normal in size. The mediastinum is unremarkable. The lungs are clear.

HRGR-Agent

The cardiomediastinal silhouette is
mal size and configuration. Pulmon
vasculature within normal limits. Th
is **right middle lobe airspace dise**
may reflect **granuloma** or pneumo
No pleural effusion. No pneumotho
No acute bony abnormalities.

The heart is enlarged. Possible
diomegaly. There is pulmonary vasc
congestion with diffusely increased i
stitial and mild patchy airspace opaci
Suspicious **pleural effusion**. *There i*
pneumothorax. There are no acute b
findings.

The heart is mildly enlarged.
aorta is atherosclerotic and ecta
Chronic parenchymal changes are n
with mild scarring and/or subsegme
atelectasis in the right lung base.
focal consolidation or significant ple
effusion identified. **Costophrenic U**
are blunted.

The heart size and pulmonary vascu
ity appear within normal limits. **I**
lung volumes. Suspicious **calci**
granuloma. *No pleural effusion or p*
mothorax. No acute bony abnormali

Machine Learning for Healthcare

Many other research projects

Active Learning

Optimization

Deep Learning

Adversarial Learning

Non-parametric methods

…

Machine Learning areas can be categorized in multiple ways (to be explored throughout the course)

- Supervised learning vs. Unsupervised learning
- Parametric vs. non-parametric
- Generative vs. Discriminative
- Prediction: regression vs. classification
- Graphical Models / Sequence Models
- Clustering / Density Estimation / Dimensionality reduction / Representation Learning
- Reinforcement Learning
- Active Learning
- …

Now let's jump right into an exercise to revise basics and learn foundational tools