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
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
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

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

## Basics

## Parametric Estimation and Prediction
Schedule

Learning Theory

03/04  Lecture 14: Generalization, Model selection
03/06  Midway Report Due
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
03/27  Lecture 18: Representation Learning
03/29  Lecture 19: Representation Learning

Graphical Models and Structured Prediction

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

<table>
<thead>
<tr>
<th>Date</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>04/17</td>
<td>Lecture 25: Semi-supervised learning</td>
</tr>
<tr>
<td>04/19</td>
<td>Recitation</td>
</tr>
<tr>
<td>04/22</td>
<td>Lecture 26: Online learning</td>
</tr>
<tr>
<td>04/24</td>
<td>Lecture 27: Reinforcement learning</td>
</tr>
<tr>
<td>04/26</td>
<td>Recitation</td>
</tr>
<tr>
<td>04/29</td>
<td>Lecture 26: Deep Reinforcement learning</td>
</tr>
<tr>
<td>05/01</td>
<td>Final Reports Due</td>
</tr>
<tr>
<td>05/03</td>
<td>Last day of class</td>
</tr>
<tr>
<td></td>
<td>Lecture 27: Review, discussion</td>
</tr>
</tbody>
</table>
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 weeks, I advise you to catch up on your prerequisites using the following videos:


- 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-qwoT1uhC7kfljpPlg28TC0x1 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
Textbooks:
• [HTF] *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Trevor Hastie, Robert Tibshirani, Jerome Friedman.
Introduction to Machine Learning

It is \textbf{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

- diversion_magazine is a website
- bordon_road_swing_bridge is a bridge
- jennifer_finnegan is a chef
- the_summer_smother_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
- miler001 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)

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