CS 2750: Machine Learning

Introduction

Prof. Adriana Kovashka
University of Pittsburgh
January 6, 2016
Course Info

• **Website:** [http://people.cs.pitt.edu/~kovashka/cs2750](http://people.cs.pitt.edu/~kovashka/cs2750)

• **Instructor:** Adriana Kovashka (kovashka@cs.pitt.edu)
  → Use "CS2750" at the beginning of your Subject

• **Office:** Sennott Square 5325

• **Office hours:** Mon/Wed, 12:15pm-1:15pm

  → And other readings
TA

- Changsheng Liu
- **Office**: Sennott Square 6805
- **Office hours**: TBD
  - Wednesday 1PM to 3PM
  - Wednesday 4PM to 6PM
  - Any time Thursday
- **Do the Doodle by the end of Friday**: [http://doodle.com/poll/7qinfkgu2xwqrnxna](http://doodle.com/poll/7qinfkgu2xwqrnxna)
Schedule

http://people.cs.pitt.edu/~kovashka/cs2750
Grading

• Homework (40%)
• Project (20%)
  – Status presentation and report (5%)
  – Final presentation and report (15%)
• Midterm exam (15%)
• Final exam (20%)
• Participation (5%)
Homework

- Four homework assignments
- Will use Matlab (tutorial next class)
- Bias towards computer vision applications
- Submission through CourseWeb → CS2750 → Assignments → Homework 1, etc.
- Attach a zip file with your written responses and code, named YourFirstName_YourLastName.zip
- Homework is due at 11:59pm on the due
- Grades will appear on CourseWeb
Homework Late Policy

- You get 3 "free" late days, i.e., you can submit homework a total of 3 days late.
- For example, you can submit one problem set 12 hours late, and another 60 hours late.
- Once you've used up your free late days, you will incur a penalty of 25% from the total project credit possible for each late day.
- A late day is anything from 1 minute to 24 hours.
Project

• Encouraged to work in groups of two
• Proposal due Feb. 29
• Status report and in-class status presentations March 28
• Final report and presentations in last week of class
• Aim for workshop-level work
• See course website for resources and more info
Exams

• One mid-term and one final exam
• The final exam will be cumulative but will focus on the latter half of the course
Readings

- Posted on course website (tentatively)
- Subject to change until 6pm on the day of the previous class
Participation

• 5% of grade will be based on participation
• No attendance will be taken, but if you don’t attend, you can’t participate
• How to participate:
  – Answer questions asked by instructor and others
  – Ask meaningful questions
  – Bring in relevant articles about recent developments in machine learning
  – Contribute on Piazza
• Feedback is welcome!
Collaboration Policy

• You will work individually. The work you turn in must be your own work.

• You can discuss the problem sets with your classmates, but do not look at their code.

• You cannot use posted solutions, search for code on the internet or use or look at Matlab implementations of something you are asked to write.

• When in doubt, ask the instructor or TA!

• Plagiarism will cause you to fail the class and receive disciplinary penalty.
Disabilities

• If you have a disability for which you are or may be requesting an accommodation, you are encouraged to contact both your instructor and Disability Resources and Services (DRS), 140 William Pitt Union, (412) 648-7890, drsrecep@pitt.edu, (412) 228-5347 for P3 ASL users, as early as possible in the term. DRS will verify your disability and determine reasonable accommodations for this course.
Medical Conditions

• If you have a medical condition which will prevent you from doing a certain assignment or coming to class, you must inform the instructor of this before the deadline.

• You must then submit documentation of your condition within a week of the assignment deadline.
Prerequisites

• Linear algebra
• Probability
• Calculus
• Programming and complexity analysis
Should I take this class?

• It will be a lot of work!
  – But you will learn a lot

• Some parts will be hard and require that you pay close attention!
  – But I will have periodic ungraded pop quizzes to see how you’re doing
  – I will also pick on students randomly to answer questions
  – Use instructor’s and TA’s office hours!!!
Questions?
Plan for Today

• Course structure and policies

• Introductions

• What is machine learning?
  – Example problems
  – Framework
  – Challenges
Introductions

• What is your name?
• What is your department and year in the program?
• What are your research interests?
• What one thing outside of school are you passionate about?
• What do you hope to get out of this class?
What is machine learning?

- Finding patterns and relationships in data
- We can apply these patterns to make useful predictions
- E.g. we can predict how much a user will like a movie, even though that user never rated that movie
Example machine learning tasks

• Netflix challenge
  – Given lots of data about how users rated movies (training data)
  – But we don’t know how user $i$ will rate movie $j$ and want to predict that (test data)
Example machine learning tasks

- Spam or not?

Sebring, Tracy
To: Batra, Dhruv
ECE 4424 proposal

CUSP has approved ECE 4424 with the following copy of the proposal with these items addressed:

Thanks!!!
Tracy

nadia bamba
To: undisclosed recipients;
Reply-To: nadia bamba
From: Miss Nadia BamBa,

From Miss Nadia BamBa,
Greeting, Permit me to inform you of my desire of going into business relationship with you. I am Nadia BamBa the only Daughter of late Mr and Mrs James BamBa, My father was a director of cocoa merchant in Abidjan, the economic capital of Ivory Coast before he was poisoned to death by his business associates on one of their outing to discus on a business deal. When my mother died on the 21st October 2002, my father took me very special because i am motherless.

Before the death of my father in a private hospital here in Abidjan, He secretly called me on his bedside and told me that he had a sum of $6,800,000(SIX Million EIGHT HUNDRED THOUSAND), Dollars left in a suspense account in a Bank here in Abidjan, that he used my name as his first Daughter for the next of kin in deposit of the fund.

He also explained to me that it was because of this wealth and some huge amount of money That his business associates supposed to balance him from the deal they had that he was poisoned by his business associates, that I should seek for a God fearing foreign partner in a country of my choice where I will transfer this money and use it for investment purposes, (such as real estate Or Hotel management).please i am honourably seeking your assistance in the following ways.

1) To provide a Bank account where this money would be transferred to.
2) To serve as the guardian of this Money since I am a girl of 19 years old.
3)Your private phone number’s and your family background’s that we can know each order more.

Moreover i am willing to offer you 15% of the total sum as compensation for effort input after the successful transfer of this fund to your designated account overseas.

Anticipating to hear from you soon.
Thanks and God Bless.
Best regards.
Example machine learning tasks

- Weather prediction

Slide credit: Carlos Guestrin
Example machine learning tasks

• Who will win <contest of your choice>?
Example machine learning tasks

• Machine translation

\[ a_1=2 \quad a_2=0 \quad a_3=1 \quad a_4=3 \quad a_5=4 \quad a_6=2 \quad a_7=5 \]

\( x = \text{bringen sie bitte das auto zurück} \)

\( y = \text{please return the car} \)
Example machine learning tasks

- Speech recognition

Slide credit: Carlos Guestrin
Example machine learning tasks

- Pose estimation
Example machine learning tasks

- Face recognition
Example machine learning tasks

- Image categorization
Example machine learning tasks

Is it dangerous?

Is it alive?

How fast does it run?

Is it soft?

Does it have a tail?

Can I poke with it?
Example machine learning tasks

- Attribute-based image retrieval

Query: “black shoes”

Feedback: “more formal than these”

Feedback: “shinier than these”

Initial top search results

Refined top search results

Example machine learning tasks

- Dating car photographs

Lee et al., “Style-aware Mid-level Representation for Discovering Visual Connections in Space and Time”, ICCV 2013
Example machine learning tasks

- Inferring visual persuasion

Example machine learning tasks

- Answering questions about images

Example machine learning tasks

• What else?
Plan for Today

• Course structure and policies
• Introductions
• What is machine learning?
  – Example problems
  – Framework
  – Challenges
ML in a Nutshell

• Tens of thousands of machine learning algorithms

• Decades of ML research oversimplified:
  – Learn a mapping from input to output $f: X \rightarrow Y$
  – $X$: emails, $Y$: {spam, notspam}
ML in a Nutshell

- Input: x  (images, text, emails...)
- Output: y  (spam or non-spam...)
- (Unknown) Target Function
  - f: X \rightarrow Y  (the “true” mapping / reality)
- Data
  - (x_1, y_1), (x_2, y_2), ..., (x_N, y_N)
- Model
  - g: X \rightarrow Y
  - y = g(x) = \text{sign}(w^T x)
ML in a Nutshell

• Every machine learning algorithm has:
  – Representation
  – Evaluation / objective function
  – Optimization

• You also need:
  – A way to represent your data

Adapted from Pedro Domingos
Representation

- Decision trees
- Sets of rules / Logic programs
- Instances
- Graphical models (Bayes/Markov nets)
- Neural networks
- Support vector machines
- Model ensembles
- Etc.

Slide credit: Pedro Domingos
Evaluation / objective function

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence
- Etc.

Slide credit: Pedro Domingos
Optimization

- **Discrete / combinatorial optimization**
  - E.g. graph algorithms

- **Continuous optimization**
  - E.g. linear programming

\[
\begin{align*}
\text{maximize} & \quad c^T x \\
\text{subject to} & \quad Ax \leq b \\
\text{and} & \quad x \geq 0
\end{align*}
\]
Types of Learning

• Supervised learning
  – Training data includes desired outputs

• Unsupervised learning
  – Training data does not include desired outputs

• Weakly or Semi-supervised learning
  – Training data includes a few desired outputs

• Reinforcement learning
  – Rewards from sequence of actions
Tasks

**Supervised Learning**

- Classification
  - \( x \rightarrow \text{Classification} \rightarrow y \)
    - Discrete

- Regression
  - \( x \rightarrow \text{Regression} \rightarrow y \)
    - Continuous

**Unsupervised Learning**

- Clustering
  - \( x \rightarrow \text{Clustering} \rightarrow y \)
    - Discrete ID

- Dimensionality Reduction
  - \( x \rightarrow \text{Dimensionality Reduction} \rightarrow y \)
    - Continuous
Your Homework

• Fill out Doodle
• Read entire course website
• Read Chapter 1 and Section 3.2 of Bishop
• Homework 1 released, due 1/27
  – Read it and let me know if any questions
  – Will overview next time
Next Time

• Matlab tutorial
• Homework 1 overview
• Machine learning basics and challenges (in more detail)
• The bias-variance trade-off