18-461/661: Introduction to ML for Engineers

Course Overview

Spring 2020

ECE – Carnegie Mellon University

Registration

- Course is currently full and we can't increase class size
- Expect several students will drop the course
- Course will be offered again in Fall 2020

Direct all waitlist-related questions to Megan Oliver (Pittsburgh) or Brittany Walker (SV): mvoliver@andrew.cmu.edu and bmw2@andrew.cmu.edu

- If you're not registered, we encourage you to stay patient
- You are welcome to keep attending the lectures until the waitlists are sorted out

Instructors & TAs: One Class, Three Campuses!

- Gauri Joshi, Co-Instructor (Pittsburgh)
- Carlee Joe-Wong, Co-Instructor (SV)
- Moise Busogi, Co-Instructor (Kigali)

- Shreyas Chaudhari, TA (Pittsburgh)
- Ritwick Chaudhry, TA (Pittsburgh)
- Soham Deshmukh, TA (Pittsburgh)
- Jacob Hoffman, TA (Pittsburgh)
- Samarth Gupta, TA (Pittsburgh)
- Mike Weber, TA (SV)
- TJ Kim, TA (SV)
- Sweta Hari Kumar, TA (SV)

Credit & thanks to:

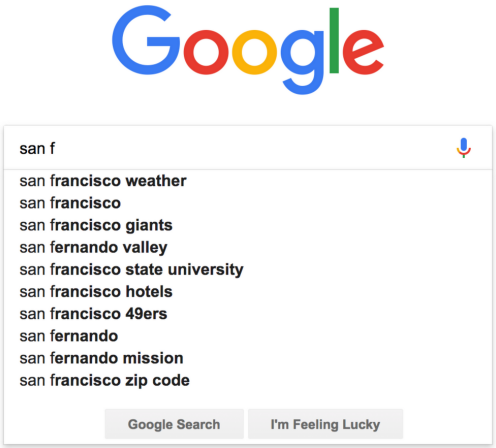
- Yuejie Chi and Virginia Smith, CMU
- Anit Sahu and Joao Saude, CMU
- Ameet Talwalkar, CMU
- Fei Sha, USC
- Emily Fox, UW

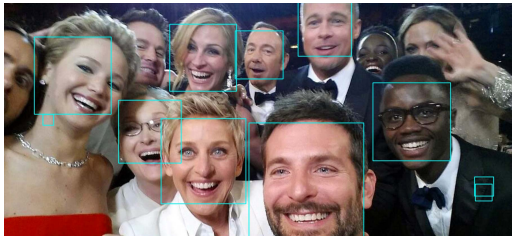
Outline

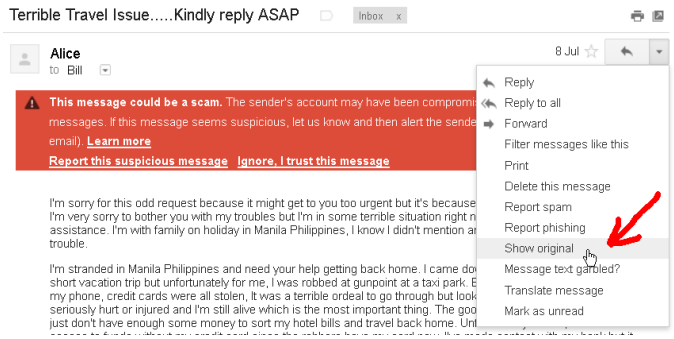
1. What is Machine Learning?
2. Course Goals
3. Course Logistics

What is Machine Learning?

What Is Machine Learning?







- **Machine learning is**: the study of methods that *improve their performance on some task with experience*

Machine Learning Pipeline

Examples

Task 1: Regression

How much should you sell your house for?

Task 1: Regression

How much should you sell your house for?

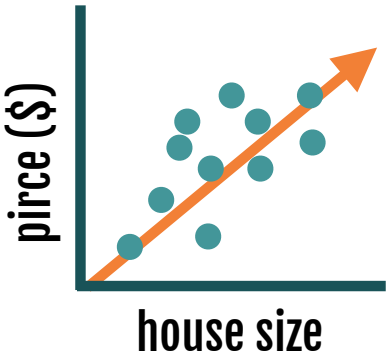
Task 1: Regression

How much should you sell your house for?

input: houses & features

Task 1: Regression

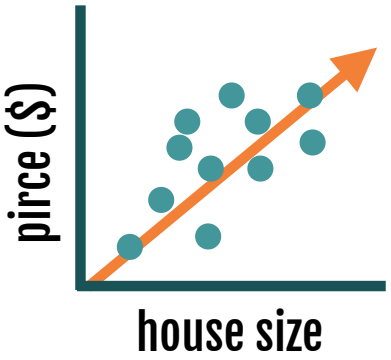
How much should you sell your house for?



input: houses & features **learn**: $x \to y$ relationship

Task 1: Regression

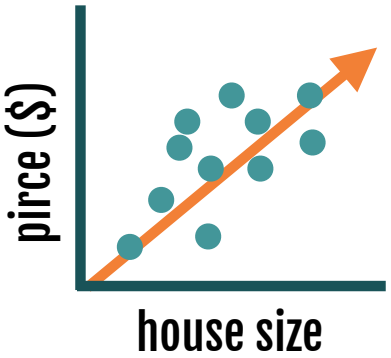
How much should you sell your house for?



input: houses & features **learn**: $x \to y$ relationship **predict**: y (*continuous*)

Task 1: Regression

How much should you sell your house for?



input: houses & features **learn**: $x \to y$ relationship **predict**: y (*continuous*)

Course Covers: Feature Scaling, Linear/Ridge Regression, Loss Function, SGD, Regularization, Cross Validation

Task 2: Classification

Cat or dog?

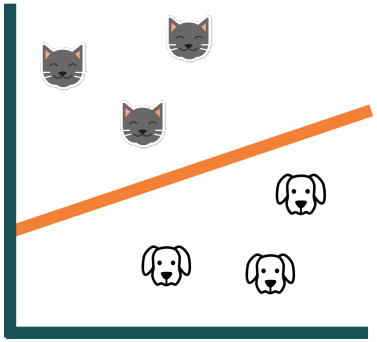
Task 2: Classification

Cat or dog?

input: cats and dogs

Task 2: Classification

Cat or dog?

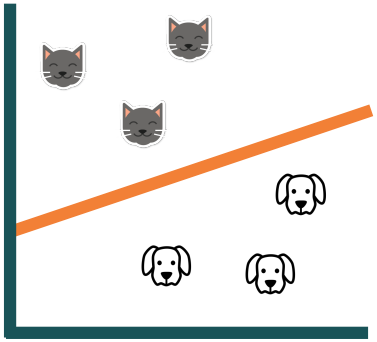


input: cats and dogs

learn: $x \to y$ relationship

Task 2: Classification

Cat or dog?



= ??

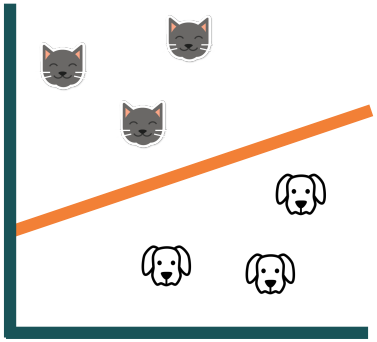
input: cats and dogs

learn: $x \rightarrow y$ relationship

predict: y (*categorical*)

Task 2: Classification

Cat or dog?



= ??

input: cats and dogs

learn: $x \rightarrow y$ relationship

predict: y (*categorical*)

Course Covers: Naive Bayes, Logistic Regression, SVMs, Neural Nets, Decision Trees, Boosting, Nearest Neighbors

Task 3: Clustering

How to segment an image?

Task 3: Clustering

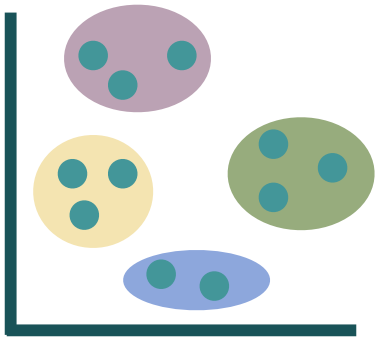
How to segment an image?

input: raw pixels $\{x\}$

Task 3: Clustering

How to segment an image?

input: raw pixels $\{x\}$

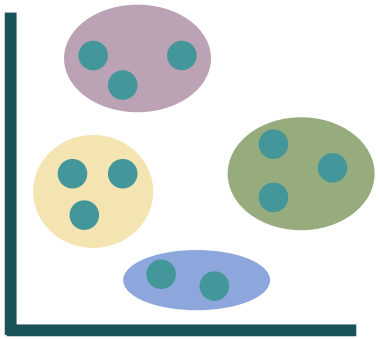


separate: $\{x\}$ into sets

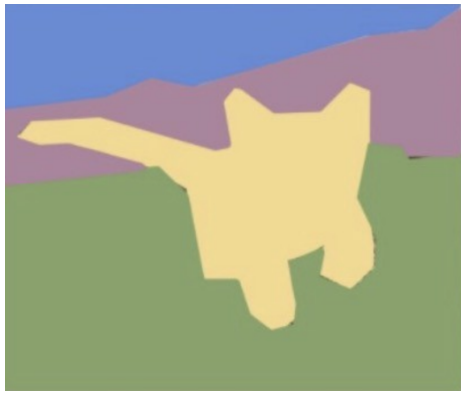
Task 3: Clustering

How to segment an image?

input: raw pixels $\{x\}$



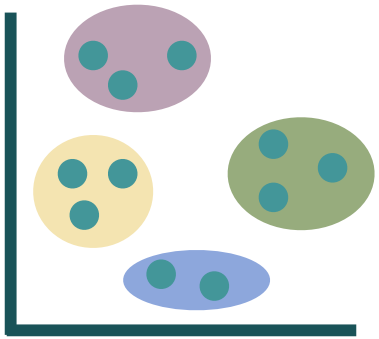
separate: $\{x\}$ into sets

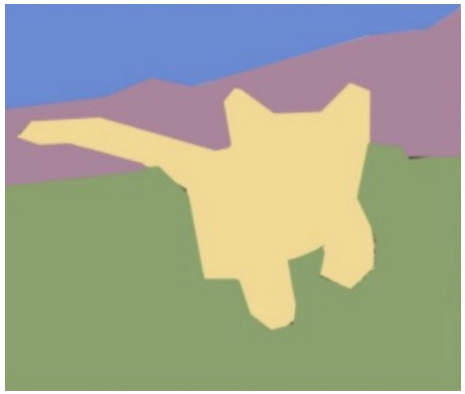


output: cluster labels $\{z\}$

Task 3: Clustering

How to segment an image?





input: raw pixels $\{x\}$ **separate**: $\{x\}$ into sets **output**: cluster labels $\{z\}$

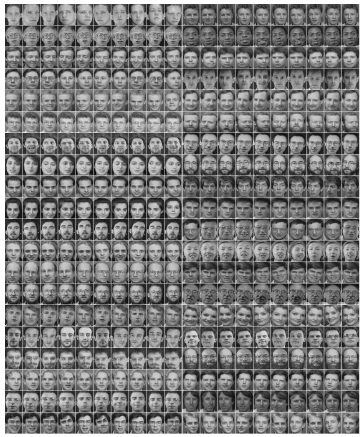
Course Covers: K-means, K-means++ clustering

Task 4: Embedding

How to reduce size of dataset?

Task 4: Embedding

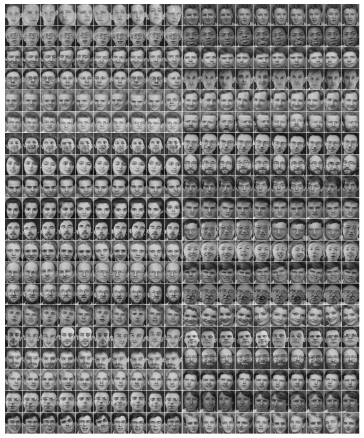
How to reduce size of dataset?

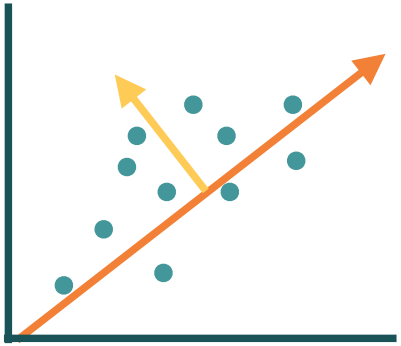


input: large dataset $\{x\}$

Task 4: Embedding

How to reduce size of dataset?

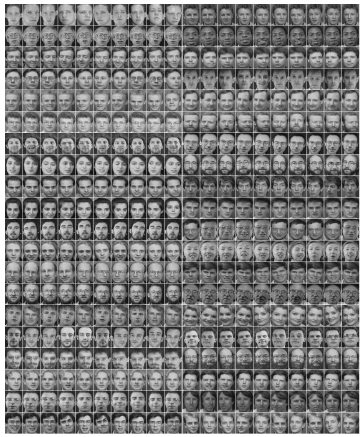


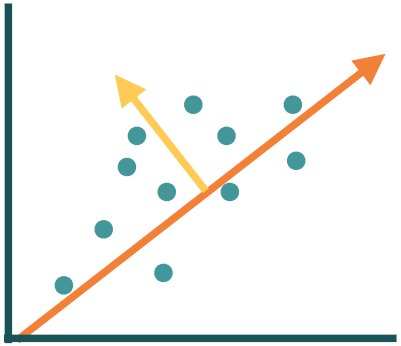


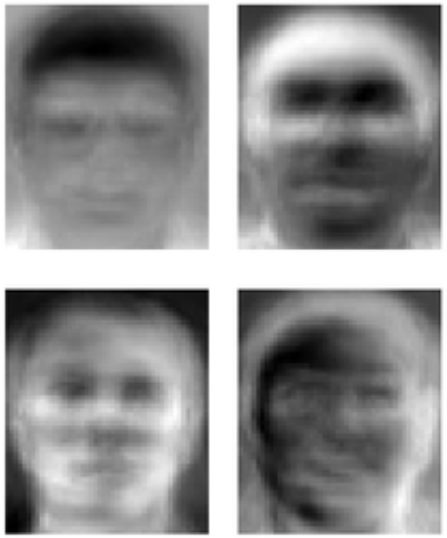
input: large dataset $\{x\}$ **find**: sources of variation

Task 4: Embedding

How to reduce size of dataset?



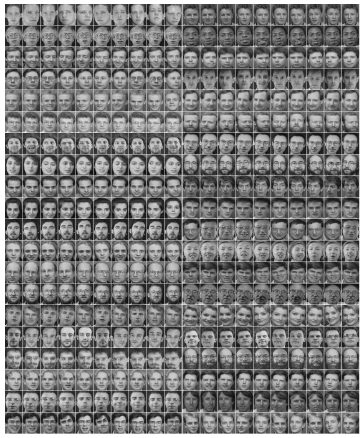


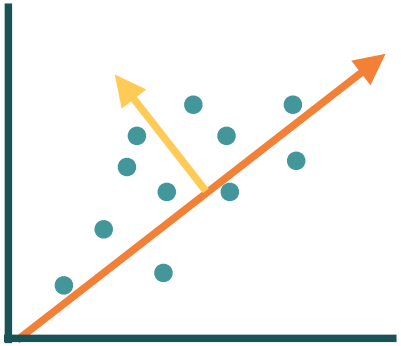


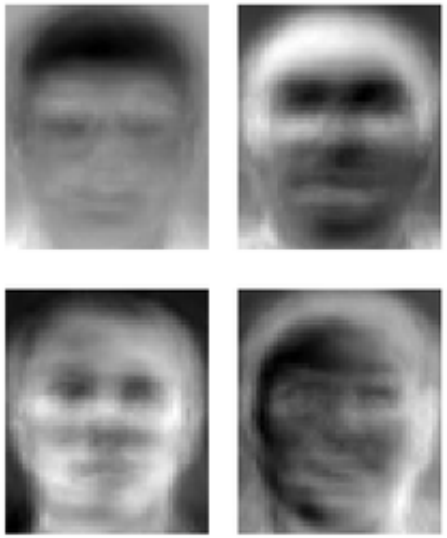
input: large dataset $\{x\}$ **find**: sources of variation **return**: representation $\{z\}$

Task 4: Embedding

How to reduce size of dataset?





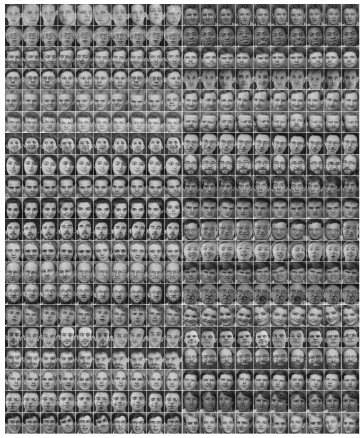


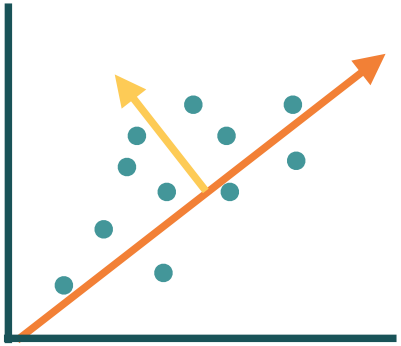
input: large dataset $\{x\}$ **find**: sources of variation **return**: representation $\{z\}$

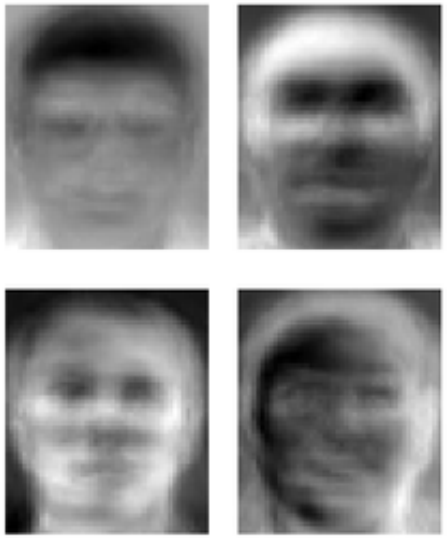
Course Covers: Dimensionality Reduction, PCA

Task 4: Embedding

How to reduce size of dataset?





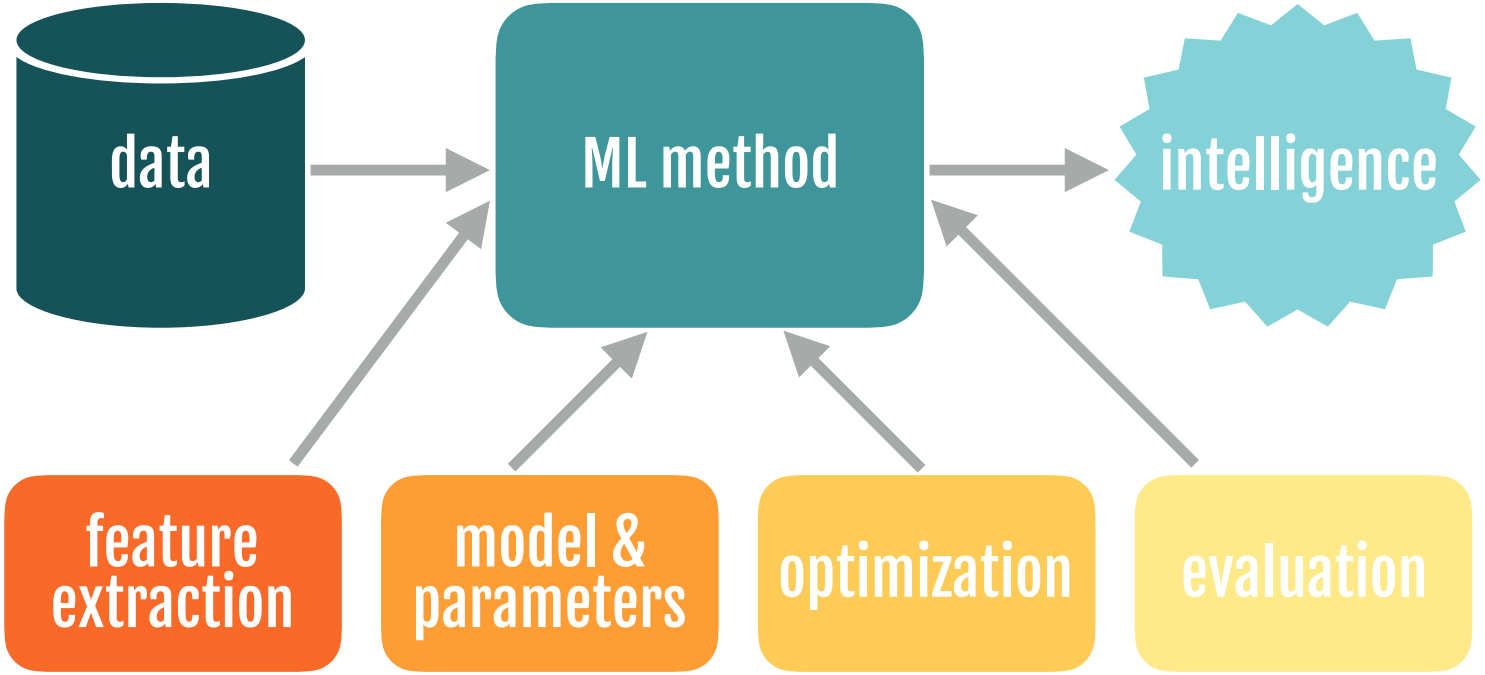


input: large dataset $\{x\}$ **find**: sources of variation **return**: representation $\{z\}$

Course Covers: Dimensionality Reduction, PCA

Course Goals

Goal of the Course



Equip you with the *tools* to *develop* and *deploy* machine learning for engineering applications

- Fundamental Understanding: Algorithms, Theoretical Analysis
- Applications: Implementation in Python, PyTorch

Key Topics

Models

- Linear regression
- Linear classification: logistic regression, SVM
- Nonlinear models: kernels, neural networks & deep learning, decision trees
- Nearest neighbors, clustering

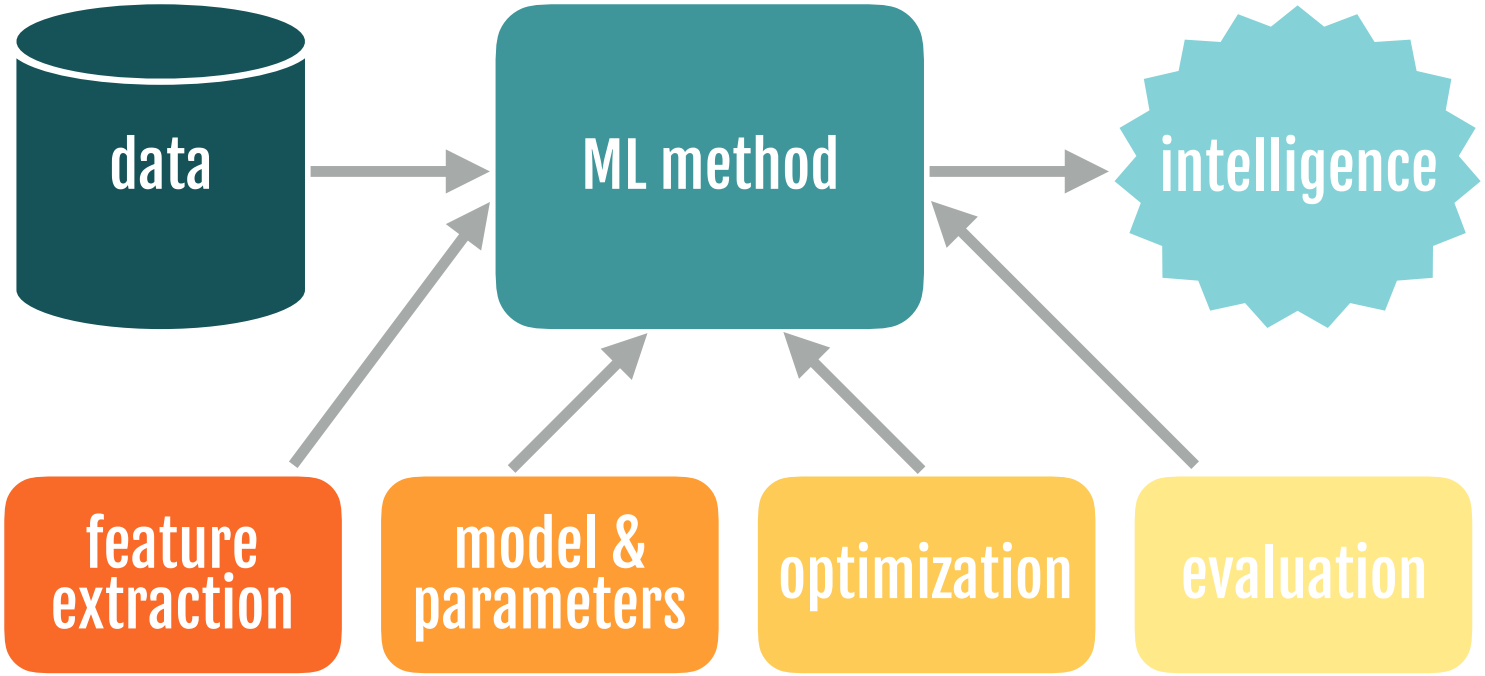
Methods

- Gradient descent
- Boosting
- k-means
- PCA

Concepts

- Point estimation, MLE, MAP
- Loss functions, bias-variance tradeoff, cross-validation
- Sparsity, overfitting, model selection
- Types of ML (supervised, unsupervised, reinforcement)

Similar Courses



- Most similar CMU Courses are 10-601 and 10-701
- This class is geared towards engineers and will include Python & PyTorch implementation of ML methods on real datasets

Course Logistics

Instructors & TAs

- Gauri Joshi, Co-Instructor (Pittsburgh)
- Carlee Joe-Wong, Co-Instructor (SV)
- Moise Busogi, Co-Instructor (Kigali)

- Shreyas Chaudhari, TA (Pittsburgh)
- Ritwick Chaudhry, TA (Pittsburgh)
- Soham Deshmukh, TA (Pittsburgh)
- Jacob Hoffman, TA (Pittsburgh)
- Samarth Gupta, TA (Pittsburgh)
- Mike Weber, TA (SV)
- TJ Kim, TA (SV)
- Sweta Hari Kumar, TA (SV)

Office Hours

- Instructors' office hours will be held immediately after class
- TA Office Hours (starting next week) TBD on course website

Lectures and Office Hours

Lectures

- Gauri and Carlee will co-teach, switching after every 3-4 lectures
- Lectures will be broadcast to the remote campuses
- Recorded lectures will be uploaded to Canvas

Office Hours

- Gauri's office hours – right after class (4:20-5 pm) at Pitt CIC 4105
- Carlee's office hours – right after class (1:20- 2:30 pm) at SV
- Moise's office hours – right before class (8-9 pm) in Kigali D110
- TA Office Hours (starting next week) TBD on course website

Recitations

- There will be ~ 1 recitation every week (see schedule on the website):
  - Friday 10:30am-11:50am ET, BH 136A (Pittsburgh)
  - Friday 1:30pm-2:50pm PT, B23 118 (SV)
- Attending discussion sections is not mandatory
  - But it is strongly encouraged!
  - Discussion will go into greater detail, cover supplementary material, answer questions, and review before exams

Homeworks and Exams

- Homeworks (50%): Both math and programming problems
- Midterm Exam (20%): Linear and Logistic Regression, SVMs
- Final Exam (30%): Nearest Neighbors, Neural Networks, Decision Trees, Boosting, Clustering, PCA, plus pre-midterm topics

We may curve up grades based on attendance and class participation

Homeworks and Exams

- Homeworks (50%): Both math and programming problems
- Midterm Exam (20%): Linear and Logistic Regression, SVMs
- Final Exam (30%): Nearest Neighbors, Neural Networks, Decision Trees, Boosting, Clustering, PCA, plus pre-midterm topics

We may curve up grades based on attendance and class participation

Difference between 18-461 and 18-661

- The major programming HW will be easier for 18-461
- Grading Curve is separate (and easier) for 18-461

Logistics

- Course Website: `https://www.andrew.cmu.edu/course/18-661/`: Slides and Other Reading Materials
- Gradescope: Homework submission and grading
- Piazza: Course Discussions

We are working on getting you access to Gradescope and Piazza ASAP. More instructions to follow

Student Wellness

Take care of yourself. Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

Student Wellness

Take care of yourself. Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

If you feel overwhelmed or want to talk to someone, please feel free to contact the following resources:

- Counseling and Psychological Services (CaPS) in Pittsburgh at 412-268-2922 or `http://www.cmu.edu/counseling/`.
- Director of Student Affairs in SV at 650-335-2846, Building 19, Room 1041 or student-services@sv.cmu.edu.

Quick Polls

- Undergrad / Masters / PhD?

Quick Polls

- Undergrad / Masters / PhD?
- ECE / Other Engineering / Computer Science / Other?

Quick Polls

- Undergrad / Masters / PhD?
- ECE / Other Engineering / Computer Science / Other?
- Registered / Waitlist / Hoping to Register / Other?

Registration

- Course is currently full and we can't increase class size
- Expect several students will drop the course
- Course will be offered again in Fall 2020

Direct all waitlist-related questions to Megan Oliver (Pittsburgh) or Brittany Walker (SV): mvoliver@andrew.cmu.edu and bmw2@andrew.cmu.edu

- If you're not registered, we encourage you to stay patient
- You are welcome to keep attending the lectures until the waitlists are sorted out

Registration

- Today's math quiz will hopefully mitigate attrition later
  - Representative of mathematical concepts you are excepted to know
  - Graded to assess your background (but not part of final grade)
  - We may contact students who perform poorly

Registration

- Today's math quiz will hopefully mitigate attrition later
  - Representative of mathematical concepts you are excepted to know
  - Graded to assess your background (but not part of final grade)
  - We may contact students who perform poorly
- Be honest / realistic with yourself about your background
  - It's better for you, me, and your classmates to drop the course now rather than a month from now, so that others can be admitted off the waitlist

Registration

- Today's math quiz will hopefully mitigate attrition later
  - Representative of mathematical concepts you are excepted to know
  - Graded to assess your background (but not part of final grade)
  - We may contact students who perform poorly
- Be honest / realistic with yourself about your background
  - It's better for you, me, and your classmates to drop the course now rather than a month from now, so that others can be admitted off the waitlist

TAs will discuss the math quiz during this Friday's recitations

Questions?

Math Quiz

Score won't affect grade

But is an indication of your preparedness for the course