18-661 Introduction to Machine Learning

Course Review and Summary

Spring 2020

ECE – Carnegie Mellon University

Final Exam

- Part I: In class on Wednesday, April 29.
- Part II: Take-home posted Friday, May 1 and due Sunday, May 3.

Logistics for Part I will be discussed at the end of this lecture. Some details on Part II:

- Open everything (notes, readings, online resources, calculators, etc.) However, you may not receive any assistance from other people.
- Any updates and clarifications to the exam will be posted on Piazza. You may ask clarification questions as private questions on Piazza, but we will not help you solve any of the problems.
- You will submit your answers on Gradescope (similar to your homework assignments). Exceptions or extensions will be granted only in the most extreme of unforeseen circumstances.
- Expect 4 to 7 descriptive problems with multiple parts. There is no time limit, but we expect you will need about 2 or 3 hours.

Studying for the Final Exam

Please see the April 20 lecture for a list of exam topics. We will not ask any questions (on either part of the exam) that require coding or programming.

- Re-solve homework and midterm problems, and/or go over their solutions.
- Read over lectures to make sure you understand the basic concepts.
- Take the practice final and make sure that you understand the answers.
- Expect questions at the same difficulty level as the midterm. Take-home questions may be slightly more difficult (more like your homework questions).
- Prepare a cheat sheet for Part I of the exam.

What is Machine Learning?

Definition

Machine learning is: the study of methods that

improve their performance

on some task

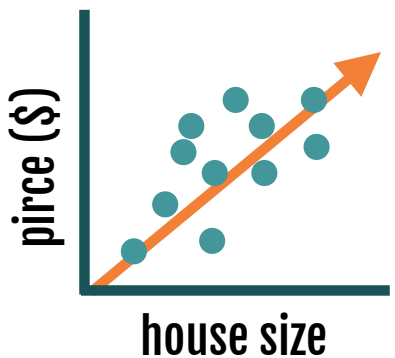
with experience

Machine Learning Pipeline

Examples

Example 1: Regression

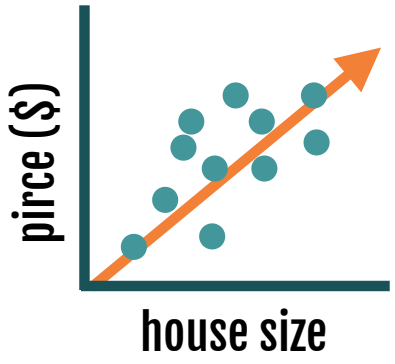
How much should you sell your house for?



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

Example 1: Regression

How much should you sell your house for?



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

Models

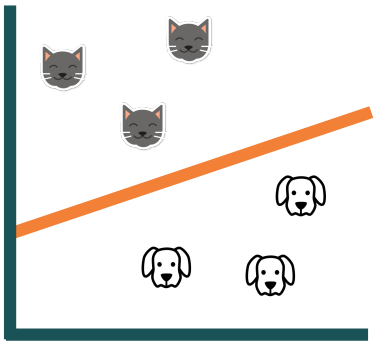
- Linear regression
- Nonlinear models: neural networks/deep learning, decision trees

Methods

- Maximum likelihood
- Gradient descent

Example 2: Classification

Cat or dog?



= ??

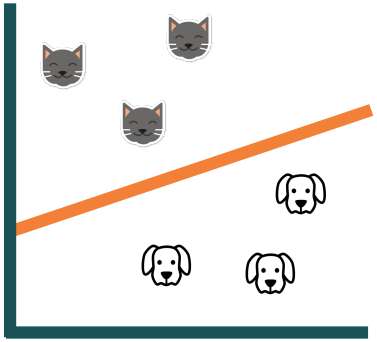
input: cats and dogs

learn: $x \rightarrow y$ relationship

predict: y (*categorical*)

Example 2: Classification

Cat or dog?



= ??

input: cats and dogs

learn: $x \to y$ relationship

predict: y (*categorical*)

Models

- Linear classification: logistic regression, SVM
- Nonlinear models: kernels, neural networks, decision trees

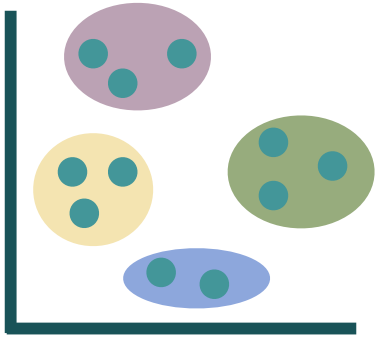
Methods

- Gradient descent
- Boosting

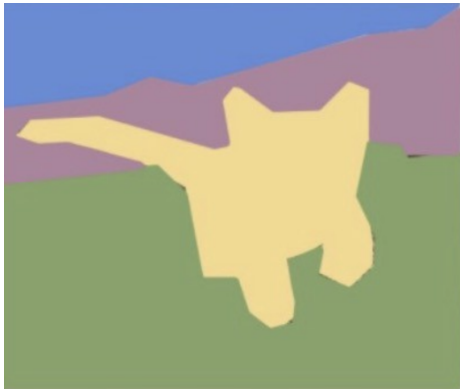
Example 3: Clustering

How to segment an image?

input: raw pixels $\{x\}$



separate: $\{x\}$ into sets



output: cluster labels $\{z\}$

Example 3: Clustering

How to segment an image?

Models

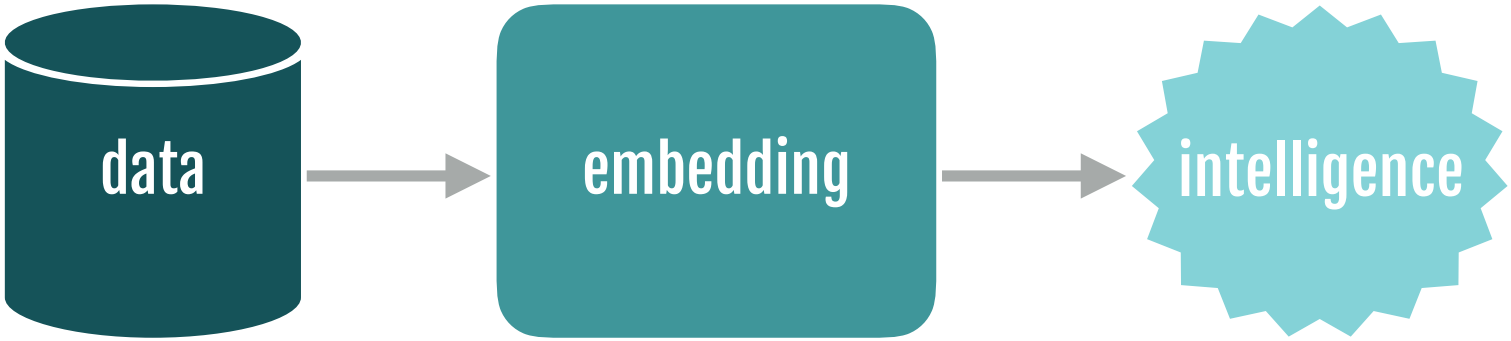
- Nearest neighbors, clustering

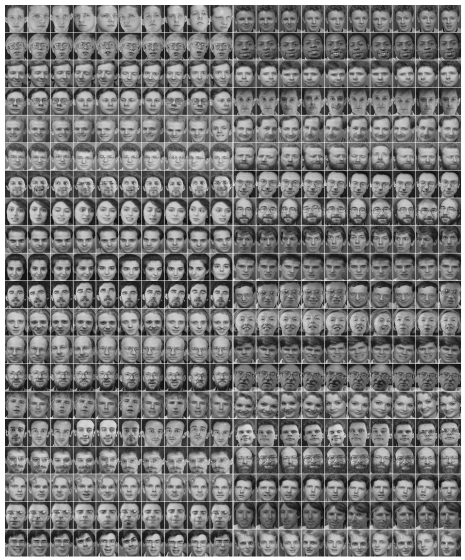
Methods

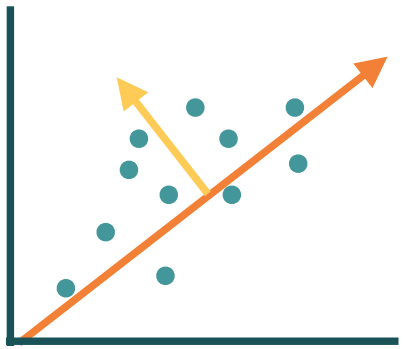
- k-means
- GMM
- EM

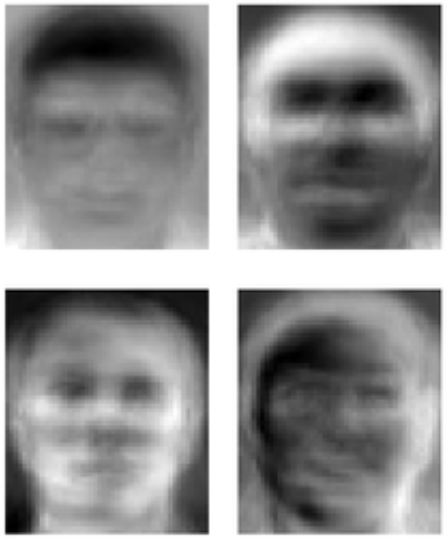
Example 4: Embedding

How to reduce size of dataset?





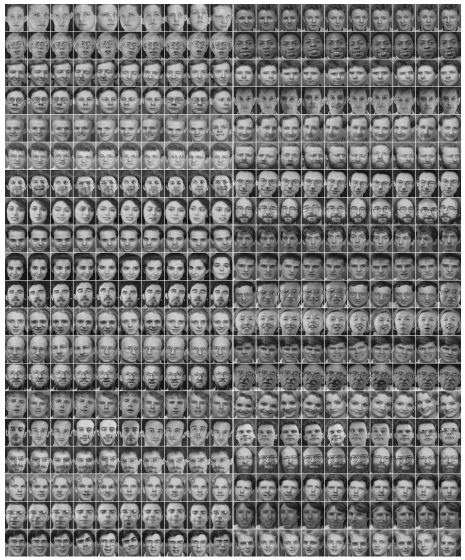


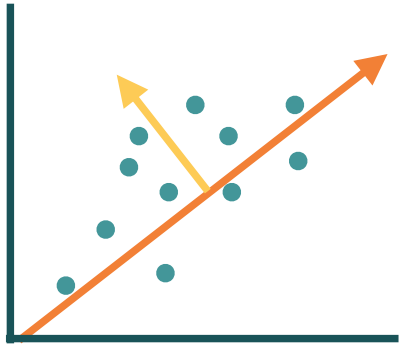


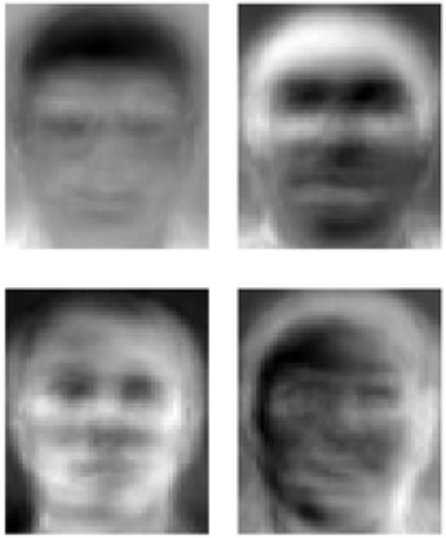
input: large dataset $\{x\}$ **find**: sources of variation **return**: representation $\{z\}$

Example 4: Embedding

How to reduce size of dataset?







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

Methods and Concepts

- PCA
- Sparsity

Key Topics

Models

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

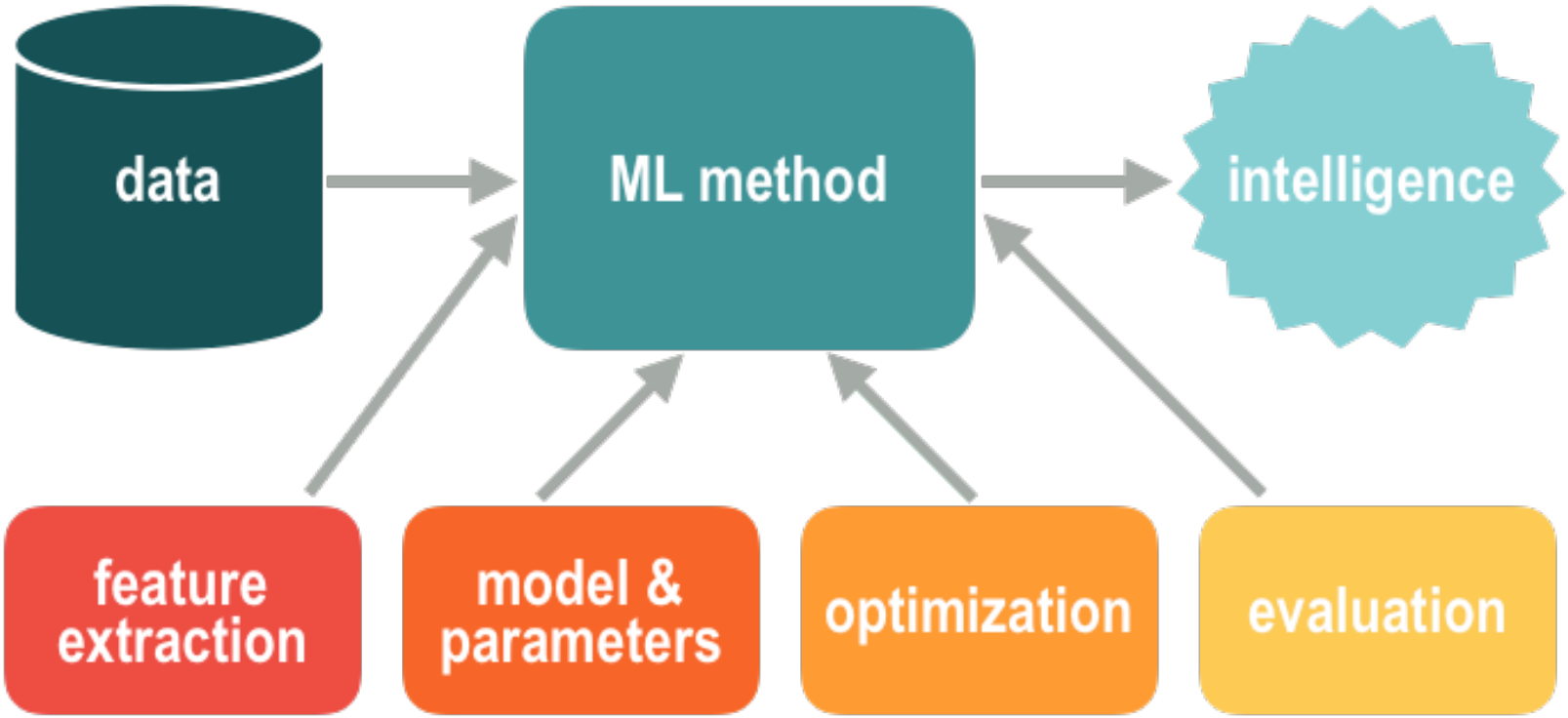
Methods

- Gradient descent
- Boosting
- k-means
- EM
- PCA

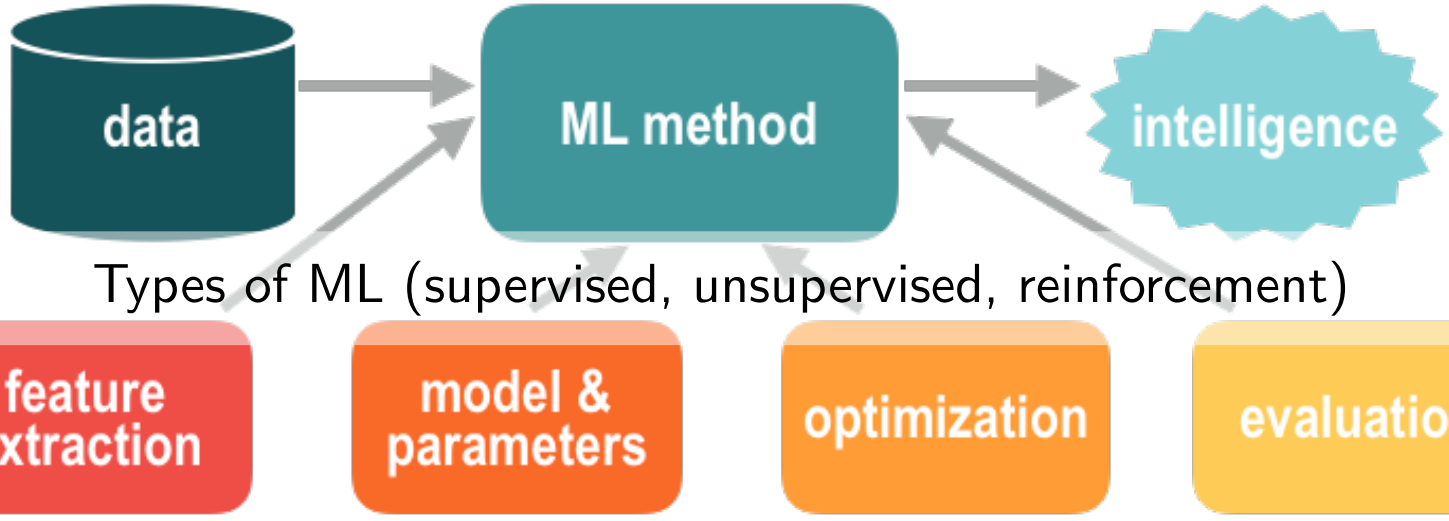
Concepts

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

Goal: Learn about ML Pipeline



Fitting the Course into the Pipeline



- PCA

- Point estimation, MLE, MAP
- Linear regression
- Linear classification
- Nonlinear models
- Nearest neighbors, clustering

- Gradient descent
- Boosting
- EM

- Loss functions, bias-variance tradeoff, cross-validation
- Sparsity, overfitting

Please Complete Course Evaluations!

We will appreciate your constructive feedback on:

- Course Content/Structure
- Level of Math
- Homework and Pytorch assignment
- Our teaching
- and other suggestions/feedbacks

Completing course evaluations will earn you 1 bonus point towards your final grade. Please upload the completion screenshot on Gradescope. Your answers will remain anonymous.

Thank you for taking the class!!

We hope you've enjoyed it as much as we have.

Final Exam, Part I: Logistics

- Expect 14 to 18 multiple choice problems. A few problems will have sub-problems. There are no descriptive questions.
- Unless otherwise stated, only one option is correct in each question. No partial credit will be given.
- If a problem asks you which of its choices is TRUE, you should treat choices that may be either true or false as FALSE.
- You are allowed one **handwritten** US-letter sized cheat sheet (two-sided). No other notes or material may be used. Calculators are not necessary and are not permitted.
- You may only use a pen/pencil, eraser, and scratch paper.
- Proctoring will be done over Zoom. If you have a question, please raise your hand and we will invite you into a breakout room to talk with a TA or instructor. Any class-wide announcements will be made in the Zoom chat box.
- You do not need to turn on your video in Zoom.

Final Exam, Part I: Gradescope

The exam will be released on Gradescope at 1:30pm ET on April 29. **Do not open the exam on Gradescope before the class time (2:30pm ET) unless we have made prior arrangements with you.** *Gradescope gives you 110 minutes to complete the exam once you have started, and we will not be able to extend this if you open the exam early.*

- We will post a PDF version of the questions on Piazza at 2:30pm ET. You may refer to this PDF as needed.
- Gradescope will give you 110 minutes to complete the exam, beginning from the time you open it.
- You may change your answers as many times as you wish. Your latest answer when time is called will be the one that is graded.
- If you experience an Internet or Gradescope outage at the end of the exam period, you have 5 minutes to send an email to the instructors with your final choices for each answer and a screenshot of Gradescope failing to load. ONLY DO THIS AS A LAST RESORT.

Practice Final Exam Instructions

- Fill in your answers for the "Practice Final Exam" on Gradescope.
- There are 9 questions and we have given you 60 minutes.
- The practice final allows late submissions, so don't worry about submitting the practice exam answers on time.
- You will be able to see the solutions once you complete the exam. We suggest you review them as you finish studying for Wednesday.
- This practice exam does not count towards your grade in the course. It is solely intended to help you prepare for the real final.

Practice for Part I of the Final Exam

Score won't affect grade

But should help you prepare