18-461/661: Introduction to ML for Engineers

Course Overview

Spring 2019

Prof. Gauri Joshi
Outline

1. What is Machine Learning?
2. Course Goals and Overview
3. Course Logistics
Instructors & TAs

- Gauri Joshi, Instructor
- Umang Bhatt, TA
- Haewon Jeong, TA
- Ankur Mallick, TA
- Abhishek Sawarkar, Grader & Web admin
• Course is currently full and we can’t increase class size
• Some on the waitlist may be cleared, but not all
• Course will be offered again in Fall 2019
• Course is currently full and we can’t increase class size
• Some on the waitlist may be cleared, but not all
• Course will be offered again in Fall 2019

• If you’re not registered, I’d encourage you to stay patient
• You are welcome to keep attending the lectures until the waitlists are sorted out

Direct all waitlist-related questions to Megan Oliver (Pittsburgh): mvoliver@andrew.cmu.edu
What is Machine Learning?
Machine learning is: the study of methods that improve their performance on some task with experience
Machine learning is: the study of methods that improve their performance on some task with experience.
Course Goals and Overview
Goal: Choose the Right ML Method for a Given Task

- data
- ML method
- intelligence
- feature extraction
- model & parameters
- optimization
- evaluation
How much should you sell your house for?
Task 1: Regression

How much should you sell your house for?

input: houses & features

predict: \( y \) (continuous)

Course Covers:
Feature Scaling, Linear/Ridge Regression, Loss Function, SGD, Regularization, Cross Validation
Task 1: Regression

How much should you sell your house for?

input: houses & features

Course Covers:
Feature Scaling, Linear/Ridge Regression, Loss Function, SGD, Regularization, Cross Validation
Task 1: Regression

How much should you sell your house for?

input: houses & features  learn: \( x \rightarrow y \) relationship

Course Covers:
Feature Scaling, Linear/Ridge Regression, Loss Function, SGD, Regularization, Cross Validation
Task 1: Regression

How much should you sell your house for?

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

Course Covers: Feature Scaling, Linear/Ridge Regression, Loss Function, SGD, Regularization, Cross Validation
Task 1: Regression

How much should you sell your house for?

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

Course Covers: Feature Scaling, Linear/Ridge Regression, Loss Function, SGD, Regularization, Cross Validation
Task 2: Classification

Cat or dog?

Data → ML method → Intelligence

Course Covers: Naive Bayes, Logistic Regression, SVMs, Neural Nets, Decision Trees, Boosting
Task 2: Classification

Cat or dog?

input: cats and dogs
Task 2: Classification

Cat or dog?

**Input:** cats and dogs  
**Learn:** $x \rightarrow y$ relationship
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
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
Task 3: Clustering

How to segment an image?

- **Input**: raw pixels \( \{x\} \)

- **Clustering**: separate \( \{x\} \) into sets

- **Output**: cluster labels \( \{z\} \)

Course Covers: Nearest Neighbors, K-means clustering
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:** Nearest Neighbors, K-means clustering
Task 4: Embedding

How to reduce size of dataset?

1. Data
2. ML method
3. Intelligence

Course Covers: Dimensionality Reduction, PCA
Task 4: Embedding

How to reduce size of dataset?

input: large dataset \{x\}
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
Topics and Our Approach

Topics:

- **Models:** Linear Regression, Naive Bayes, Logistic Regression, SVMs, Neural Networks, Decision Trees, Clustering
- **Algorithms:** Gradient Descent, k-means, Boosting, PCA
- **Concepts:** Loss Functions, Bias-Variance Trade-off, Empirical Risk Minimization, Cross-Validation
- **Extra Topics:** Large-scale Learning, Reinforcement Learning

Goal:

Equip you with tools necessary to develop and deploy machine learning for engineering applications
- Fundamental Understanding: Algorithms, Theoretical Analysis
- Application to Engineering Problems: Implementation in Python, TensorFlow, PyTorch
Topics and Our Approach

**Topics:**

- **Models:** Linear Regression, Naive Bayes, Logistic Regression, SVMs, Neural Networks, Decision Trees, Clustering
- **Algorithms:** Gradient Descent, k-means, Boosting, PCA
- **Concepts:** Loss Functions, Bias-Variance Trade-off, Empirical Risk Minimization, Cross-Validation
- **Extra Topics:** Large-scale Learning, Reinforcement Learning

**Goal:** Equip you with *tools* necessary to *develop* and *deploy* machine learning for engineering applications

- Fundamental Understanding: Algorithms, Theoretical Analysis
- Application to Engineering Problems: Implementation in Python, TensorFlow, PyTorch
Credit & thanks to:

- Virginia Smith, Carlee Joe-Wong, CMU
- Anit Sahu, Joao Saude, CMU
- Ameet Talwalkar, CMU
- Fei Sha, USC
- Emily Fox, UW
Course Logistics
Instructors & TAs

- Gauri Joshi, Instructor
- Umang Bhatt, TA
- Haewon Jeong, TA
- Ankur Mallick, TA
- Abhishek Sawarkar, Grader & Web admin

Office Hours

- My office hours will be held immediately after class
- TA Office Hours (starting next week) will be posted on the course website
There will be one discussion section:
  - Friday 10:30 am – 11:50 am, POS 152

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%): Neural Networks, Decision Trees, Boosting, Clustering, PCA, plus pre-midterm topics

We may curve up grades based on attendance and class participation
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
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?
• **Today’s math quiz will hopefully mitigate attrition later**
  
  • Representative of probability and linear algebra concepts you are expected to know
  
  • If you have learnt but forgotten these concepts, you should be able to solve the quiz after Thursday’s lecture and Friday’s recitation
  
  • Graded to assess your background (but not part of final grade)
Math Quiz

• **Today’s math quiz will hopefully mitigate attrition later**
  ● Representative of probability and linear algebra concepts you are excepted to know
  ● If you have learnt but forgotten these concepts, you should be able to solve the quiz after Thursday’s lecture and Friday’s recitation
  ● Graded to assess your background (but not part of final grade)

• 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
Questions?
Math Quiz

*Score won’t affect grade*

*But is an indication of your preparedness for the course*