Pop Quiz 1 - Why Are You Here?

Because Machine Learning is everywhere!

Welcome to CS260!

Teaching Staff

- Instructor: Dr. Yutao He (yutao@cs.ucla.edu)
- Office Hours: 8-10pm Thursday on Zoom
- TA: Aashna Agarwal (aashna@cs.ucla.edu)
- Office Hours: 2-4pm Saturday on Zoom
Prerequisite

- You are expected to have the following background to do well in the course:
  - Linear algebra
  - Multivariate calculus
  - Probability theory and statistics
  - Algorithms and complexity
  - Hands-on programming experience in Python

Textbooks

- Required
  - Learning from Data
    - By Yaser S. Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin
    - http://www.amlbook.com

- Optional
  - Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow
    - By Sebastian Raschka and Vahid Mirjalili
    - 2nd Edition
  - Pattern Recognition and Machine Learning
    - By Christopher M. Bishop
Course website

- URL
  - http://ccle.ucla.edu/course/view/18W-COMSCI260-1
- Contains:
  - Course Information
  - Up-to-date news
  - Homework Assignments
  - Projects
  - Lecture Videos and Notes
  - Resources
  - Link to Discussion Forum

Check it frequently!

Piazza As A Discussion Forum

Piazza As A Discussion Forum

- It is highly catered to getting you help fast and efficiently from your classmates, TA, and myself
- Rather than emailing questions to the teaching staff, please post your questions on Piazza. You can even post anonymously
- Should be used for any general or specific questions on the class
- No flame, no spam, no commercials, no for-sales …
- You’re welcome to email to the teaching staff at the ordinary email address any private questions

Lecture Style
Lecture Style

- Slides will be used
- Each lecture is structured as
  - 2-Minute administrative matters
  - 3-Minute review
  - 110-Minute new contents
  - 5-Minute summary

Course Structure

Course Grade Breakdown
• Homework: 20%
• Project: 30%
• Midterm: 20%
• Final: 30%
• The final course grade will be curved

Homework Assignment

• There are totally about 10 weekly assignments
• Each assignment will be released on-line by 8am every Monday
• Due at midnight on the following Monday
• Work is submitted electronically and time-stamped via CCLE
• It will include both theoretical Q&A questions and small hands-on programming assignments in Python.

Lateness Policy
• 2 LAW (Late Work) token for the entire quarter
  – Each student can submit his/her work one day later (12pm Tuesday following the deadline) without any penalty
  – Must inform the teaching staff before the original submission of using a LAW token
  – When it is used up, no late work will be accepted

Projects

• There will be totally two projects
• Each project is designed to explore how best to solve some real-world problems by applying what is taught in class and going through the complete application process by using the state-of-the-practice tools

Exams
• One Midterm
  – Date: Saturday, February 10
  – Location: Classroom on UCLA campus

• Final
  – Date: Saturday, March 17
  – Location: Classroom on UCLA campus

• No make-up is granted for the midterm or final unless you run into unexpected medical or family emergency

---

Computing Resources and Software Tools

Both in class and the following Python libraries will be used extensively for homework assignments and projects in the course

- scikit-learn, scikit.learn, scikit.image, SciPy, NumPy, Matplotlib, seaborn, Pandas, keras, and Tensorflow

• They will be available on the Seasnet workstations or better be installed on your own personal computer.

• Work at school:
  – SEASNet Lab

• Work at home:
  – Remote login or install all tools on your own laptop

• Other setup
  • Available on MacOS, Windows, and Linux

---

Academic Integrity

• Studying in groups is encouraged
• Studying in groups is encouraged
• Actual work must be your own
• Common cheating scenarios:
  – Running out of time on an assignment and then cut-and-paste from your buddy’s work
  – Cross-talking and cross-eyeing during exams
• Zero-tolerance
  – Will be reported to Dean’s Office if any violation is caught

Class Survey aka Homework #0

• Help us better understand your background
• Help us better understand your background
• Develop more effective teaching portfolios
• Maximize your learning benefit
• Not graded but mandatory

Outline
Objectives

- Grasp the state-of-the-art ML theories
- Understand the key ML algorithms as building blocks of solutions
- Know how to get them to work in the real world problems
- Master the state-of-the-practice framework/tools

What You Will Learn From CS260

- Key ML algorithms
  - Linear regression, Logistics regression, KNN
- Linear regression, Logistic regression, KNN
- Decision Tree, Random Forests, ...

- Underlying statistics/computational theories
  - text, image, unstructured
  - self-generated

- Viable methodologies for tackling the real world data

- Hands-on experiences with frameworks/toolkits
  - Python as a programming language
  - NumPy, SciPy, and pandas for numerical computation
  - Matplotlib and seaborn for data visualization
  - Scikit-learn for machine learning libraries
  - Keras and Tensorflow for deep learning

---

**Course Contents At-A-Glance**

<table>
<thead>
<tr>
<th>Week</th>
<th>Lecture Topics</th>
<th>Reading</th>
</tr>
</thead>
</table>
| 1    | - Introduction and Course Overview  
      - Probability and Linear Algebra Review |         |
| 2    | - Learning Problem and Perceptron  
      - Linear Classification | 1.1-1.3, 3.1 |
| 3    | - Linear Regression  
      - Logistic Regression and Gradient Descent | 3.2-3.4 |
| 4    | - Theory of Generalization, VC Dimension, and Bias-Variance Tradeoff  
      - Overfitting, Regularization, and Validation | 2 and 4 |
| 5    | - Decision Tree and Random Forrest  
      - Gaussian Discriminant Analysis and Naive Bayes Algorithm |         |
| 6    | - Nearest Neighbors (NN) and k-NN  
      - Support Vector Machine I | e-6, e-8 |
| 7    | - Support Vector Machine II  
      - Boosting | e-8 |
| 8    | - K-means Clustering  
      - Mixture Models, Gaussian Mixture Model, EM Algorithm | e-6 |
| 9    | - Dimensionality Reduction and PCA  
      - Reinforcement Learning | e-9 |
| 10   | - Deep Learning  
      - Machine Learning Trends and Course Summary | e-9 |

---

**What is Learning?**
"Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the task or tasks drawn from the same population more efficiently and more effectively the next time."

—Herbert Simon

What is Machine Learning?

"Field of study that gives computers the ability to learn without being explicitly programmed."

—Arthur Samuel

Why Now?

- More Mature Theory and Algorithms
- More Data
- More power computer hardware

Machine and Traditional Programming

Traditional Programming

Data → Computer → Output
Program

Machine Learning

Data → Computer → Program
Output

ML In Action

Learning to Predict Emergency C-Sections
One of 18 learned rules:

If No previous vaginal delivery, and Abnormal 2nd Trimester Ultrasound, and Malpresentation at admission
Then Probability of Emergency C-Section is 0.6

Over training data: 26/41 = .63,
Over test data: 12/20 = .60

ML In Action

Learning to classify text documents

spam vs not spam
ML In Action

Learn to classify the word a person is thinking about based on fMRI brain activity.