Machine Learning
CS229/STATS229

Instructors: Tengyu Ma and Chris Re
1. Administrivia

[cs229.stanford.edu]

2. Topics Covered in This Course
Pre-requisite

- Probability (CS109 or STAT 116)
  - distribution, random variable, expectation, conditional probability, variance, density
- Linear algebra (Math 104, Math 113, or CS205)
  - matrix multiplication
  - eigenvector
- Basic programming (in Python)
- Will be reviewed in Friday sections (recorded)

This is a mathematically intense course. 😬
But that’s why it’s exciting and rewarding!
Honor Code

Do’s

- write down the solutions independently
- write down the names of people with whom you’ve discussed the homework
- read the longer description on the course website

Don’ts

- copy, refer to, or look at any official or unofficial previous years’ solutions in preparing the answers
Course Project

- We encourage you to form a group of 1-3 people
- same criterion for 1-3 people
- More information and previous course projects can be found on course website
- List of potential topics

- Athletics & Sensing Devices
- Audio & Music
- Computer Vision
- Finance & Commerce
- General Machine Learning
- Life Sciences
- Natural Language
- Physical Sciences
- Theory
- Reinforcement Learning
Other Information on Course Website

cs229.stanford.edu

- Piazza:
  - technical and logistical question (anonymous or non-anonymous, private or public)
  - to find study groups friends
  - all announcement

- Videos on canvas

- Course calendar: office hours and deadlines

- Section (not Fri section) vs office hour

- Gradescope
  - you will receive invite after Axess enrollment within 24hrs

- Late days policy

- FAQ
1. Administrivia

2. Topics Covered in This Course
Definition of Machine Learning

Arthur Samuel (1959): Machine Learning is the field of study that gives the computer the ability to learn without being explicitly programmed.

A. L. Samuel*

Some Studies in Machine Learning
Using the Game of Checkers. II—Recent Progress

Photos from Wikipedia
Definition of Machine Learning

Tom Mitchell (1998): a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

Experience (data): games played by the program (with itself)
Performance measure: winning rate
Taxonomy of Machine Learning
(A Simplistic View Based on Tasks)

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
Taxonomy of Machine Learning
(A Simplistic View Based on Tasks)

Supervised Learning

Unsupervised Learning

Reinforcement Learning

can also be viewed as tools/methods
Supervised Learning
Housing Price Prediction

Given: a dataset that contains \( n \) samples

\[
(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})
\]

Task: if a residence has \( x \) square feet, predict its price?

\[
x = 800 \\
y = \, ?
\]

15th sample
\[
(x^{(15)}, y^{(15)})
\]
Housing Price Prediction

- Given: a dataset that contains $n$ samples
  
  \[ (x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)}) \]

- Task: if a residence has $x$ square feet, predict its price?

Lecture 2&3: fitting linear/quadratic functions to the dataset

\[ x = 800 \]
\[ y = ? \]
More Features

- Suppose we also know the lot size
- Task: find a function that maps 
  \[(\text{size, lot size}) \rightarrow \text{price}\]
  features/input \[x \in \mathbb{R}^2\]
  label/output \[y \in \mathbb{R}\]
- Dataset: \[(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})\]
  where \[x^{(i)} = (x_1^{(i)}, x_2^{(i)})\]
- “Supervision” refers to \(y^{(1)}, \ldots, y^{(n)}\)
High-dimensional Features

- $x \in \mathbb{R}^d$ for large $d$
- E.g.,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ \vdots \\ x_d \end{bmatrix}$$

--- living size
--- lot size
--- # floors
--- condition
--- zip code

$y$ --- price

- Lecture 6-7: infinite dimensional features
- Lecture 10-11: select features based on the data
Regression vs Classification

- **regression**: if $y \in \mathbb{R}$ is a continuous variable
  - e.g., price prediction

- **classification**: the label is a discrete variable
  - e.g., the task of predicting the types of residence
    
    (size, lot size) $\rightarrow$ house or townhouse?

---

Lecture 3&4: classification

![Graph showing the prediction of house or townhouse based on lot size and size. The graph has points indicating different lot sizes, with a trend line showing how lot size affects the classification.]
Supervised Learning in Computer Vision

- Image Classification

$x = \text{raw pixels of the image}, \ y = \text{the main object}$

Supervised Learning in Computer Vision

- Object localization and detection
- $x = \text{raw pixels of the image}, \ y = \text{the bounding boxes}$
Supervised Learning in Natural Language Processing

- Machine translation

**Note:** this course only covers the basic and fundamental techniques of supervised learning (which are not enough for solving hard vision or NLP problems.)

- CS224N and CS231N would be more suitable if you are interested in the particular applications.
Unsupervised Learning
Unsupervised Learning

- Dataset contains no labels: $x^{(1)}, \ldots, x^{(n)}$
- Goal (vaguely-posed): to find interesting structures in the data

**supervised**

**unsupervised**
Clustering
Clustering

- Lecture 12&13: k-mean clustering, mixture of Gaussians

![Diagram showing clustering](image-url)
Clustering Genes

Identifying Regulatory Mechanisms using Individual Variation Reveals Key Role for Chromatin Modification. [Su-In Lee, Dana Pe'er, Aimee M. Dudley, George M. Church and Daphne Koller. ’06]
Lecture 14: principal component analysis (tools used in LSA)

Image credit: https://commons.wikimedia.org/wiki/File:Topic_detection_in_a_document-word_matrix.gif
Word Embeddings

Represent words by vectors

- word $\quad$ encode $\quad$ vector
- relation $\quad$ encode $\quad$ direction

Unlabeled dataset

Word2vec [Mikolov et al’13]
GloVe [Pennington et al’14]
Clustering Words with Similar Meanings (Hierarchically)

[Arora-Ge-Liang-M.-Risteski, TACL’17,18]
Reinforcement Learning
learning to walk to the right

Iteration 10
learning to walk to the right

Iteration 20
learning to walk to the right

Iteration 80

[Luo-Xu-Li-Tian-Darrell-M.’18]
learning to walk to the right

Iteration 210
Reinforcement Learning

- The algorithm can collect data interactively

Try the strategy and collect feedbacks

Data collection

Training

Improve the strategy based on the feedbacks
Supervised Learning
Unsupervised Learning
Reinforcement Learning

can also be viewed as tools/methods
Other Tools/Topics In This Course

- Deep learning basics
- Introduction to learning theory
  - Bias variance tradeoff
  - Feature selection
  - ML advice
Questions?

Thank you!