CS244 Machine Learning

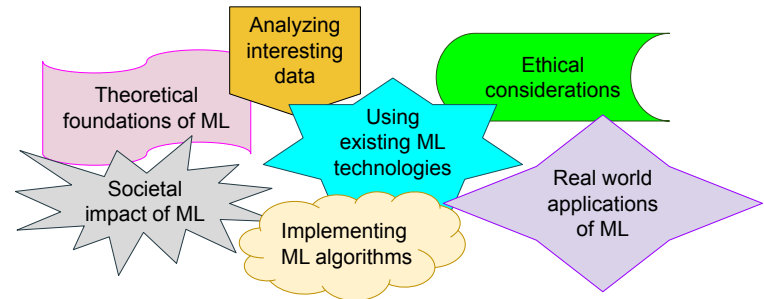## Course Information

---

Programming Environment

Check out Project 0!

- We will both implement our own machine learning algorithms and use existing machine learning algorithm implementations

- There are *many* software libraries for studying and programming machine learning applications

- In this course, we will use Python and its libraries `numpy`, `matplotlib`, and `sklearn`

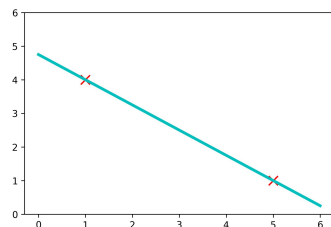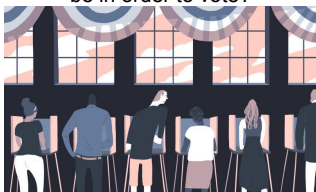- For development, we will use Anaconda together with Jupyter notebooks

---

Aspects of the Course

Different machine learning (ML) courses emphasize different things

- Analyzing interesting data
- Ethical considerations
- Theoretical foundations of ML
- Using existing ML technologies
- Societal impact of ML
- Implementing ML algorithms
- Real world applications of ML

Slide 1: What is Learning?

How old do you need to be in order to vote?

There is a line that passes through (1,4) and (5,1). Does (3,3) also lie along this line?

In machine learning, computers learn with experience. They solve a problem based on data or they get better at a task as they have more examples.

Slide 2: Is this how humans perform tasks?

- Do we pattern match from examples?

- Or do we have a deeper understanding / intuition for how to recognize faces, drive safely, and play chess?

Slide 3: ML Application
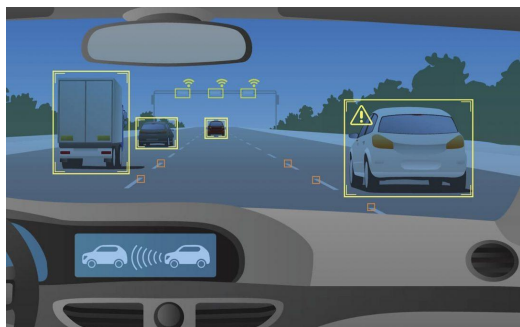
Slide 4: Observe these photos of two people

Z

TS

Quick!

Is this **Z** or **TS**?

How would you write a program to recognize a particular person in a picture?

Could this program be adapted easily to recognize some other person?

Automatically learn from examples

Program should *observe patterns* in pictures of faces to recognize new ones

---

Flavors of Machine Learning

1. Supervised learning:
   given **training examples with labels**, learn to automatically label (aka classify) a new example

   Z

   TS

Flavors of Machine Learning

2. Unsupervised learning:
   given **data <u>without</u> labels**, extract hidden structure
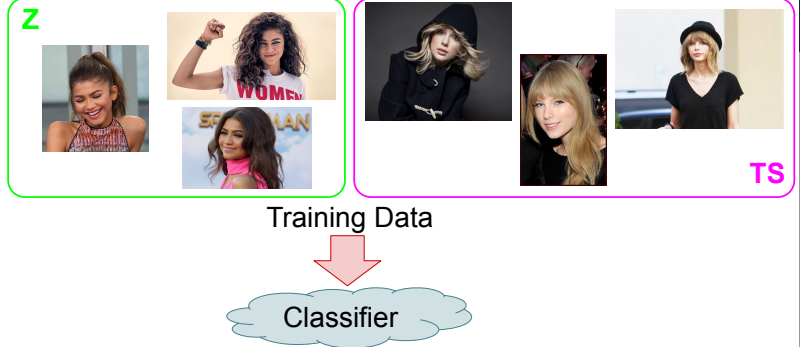
   ჩ ჳ რ დ ი ш η ლ ף

Flavors of Machine Learning

3. Reinforcement learning:
   learn how to take actions to maximize total reward

Supervised Learning Phases: (1) Training

Z

TS

Training Data

Classifier

Supervised Learning Phases: (2) Prediction + Evaluation

Testing Data

Z

Classifier

Are these equal?

Predicted Label

Training and Testing ML Algorithms

Testing Data

Training Data

Amy Tan   Judy Bloom
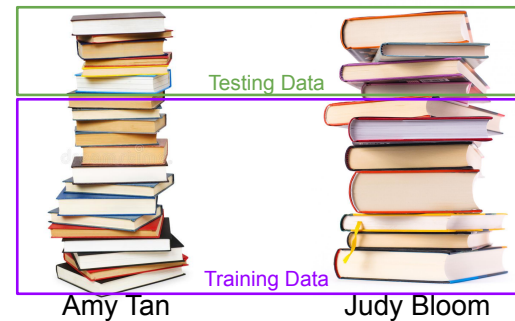
Who is the author?

Data Representation with Features

- A feature is a question you ask of every example in your data (in training and testing)

- A feature value is the answer to that question for a particular example in the data

- Typically have LOTS of features

- Machine learning algorithms specify how to use these features to build classifiers
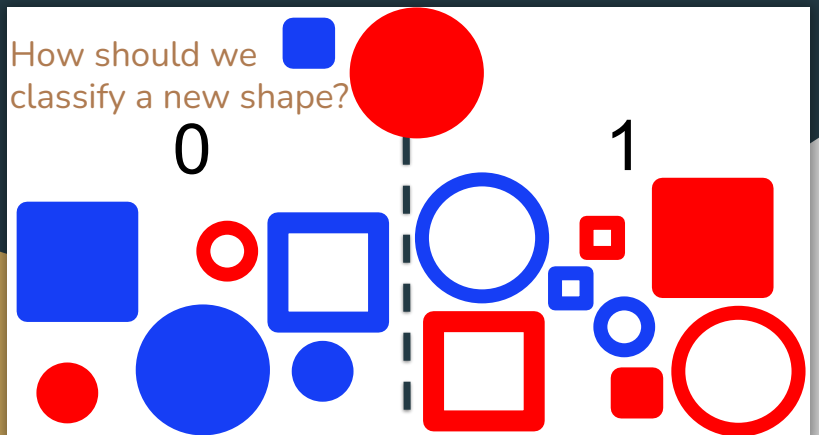
Common Danger: Overfitting

Z

TS

- TS's shirt is black in all training examples

- Classifier may learn that "black shirt" is a good feature for recognizing her

- What happens if she is not wearing black clothing, or if someone else is wearing black clothing?

Classification Problem

0   1
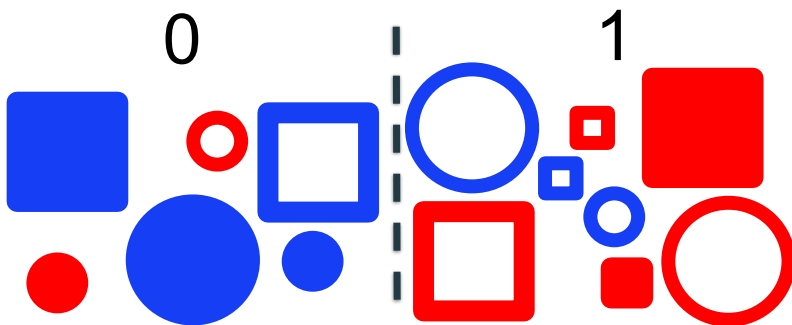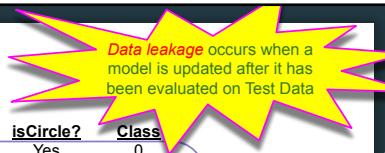
How should we classify a new shape?

0   1

Classification Problem

isLarge?	isRed?	isFilled?	isCircle?	Class
Yes	No	No	Yes	1
Yes	No	Yes	Yes	0
No	No	No	Yes	1
Yes	No	No	No	0
No	No	No	No	1
No	No	Yes	Yes	0
Yes	No	Yes	No	0
No	Yes	Yes	No	1
Yes	Yes	No	No	1
No	Yes	No	Yes	0
Yes	Yes	Yes	No	1
No	Yes	Yes	Yes	0
Yes	Yes	No	Yes	1
No	Yes	No	No	1

Randomly Permute Rows

isLarge?	isRed?	isFilled?	isCircle?	Class
Yes	No	Yes	Yes	0
No	Yes	No	No	1
Yes	No	No	Yes	1
Yes	No	Yes	No	0
Yes	Yes	No	No	1
No	Yes	Yes	Yes	0
No	No	No	No	1
No	No	No	Yes	1
No	Yes	Yes	No	1
No	Yes	No	Yes	0
Yes	Yes	Yes	No	1
Yes	Yes	No	Yes	1
Yes	No	No	No	0
No	No	Yes	Yes	0

Training and Test Data

Data leakage occurs when a model is updated after it has been evaluated on Test Data

isLarge?	isRed?	isFilled?	isCircle?	Class	
Yes	No	Yes	Yes	0	Training data
No	Yes	No	No	1	
Yes	No	No	Yes	1	
Yes	No	Yes	No	0	
Yes	Yes	No	No	1	
No	Yes	Yes	Yes	0	
No	No	No	No	1	
No	No	No	Yes	1	
No	Yes	Yes	No	1	
No	Yes	No	Yes	0	
Yes	Yes	Yes	No	1	Test data
Yes	Yes	No	Yes	1	
Yes	No	No	No	0	
No	No	Yes	Yes	0	

X   y

This Course

- Algorithms for learning supervised classifiers
- Some unsupervised learning
- Social and ethical issues
- Designing good features for different domains
- How to evaluate your classifier
- How to overcome overfitting
- Processing large amounts of data