ELEG 602
Advanced Machine Learning

Xiugang Wu

Dept of Electrical & Computer Engineering

https://www.eecis.udel.edu/~xwu

Fall 2021

Agenda

- Icebreaker -- What is machine learning?
- What can you expect to learn from this course?
- Logistics

What is machine learning?

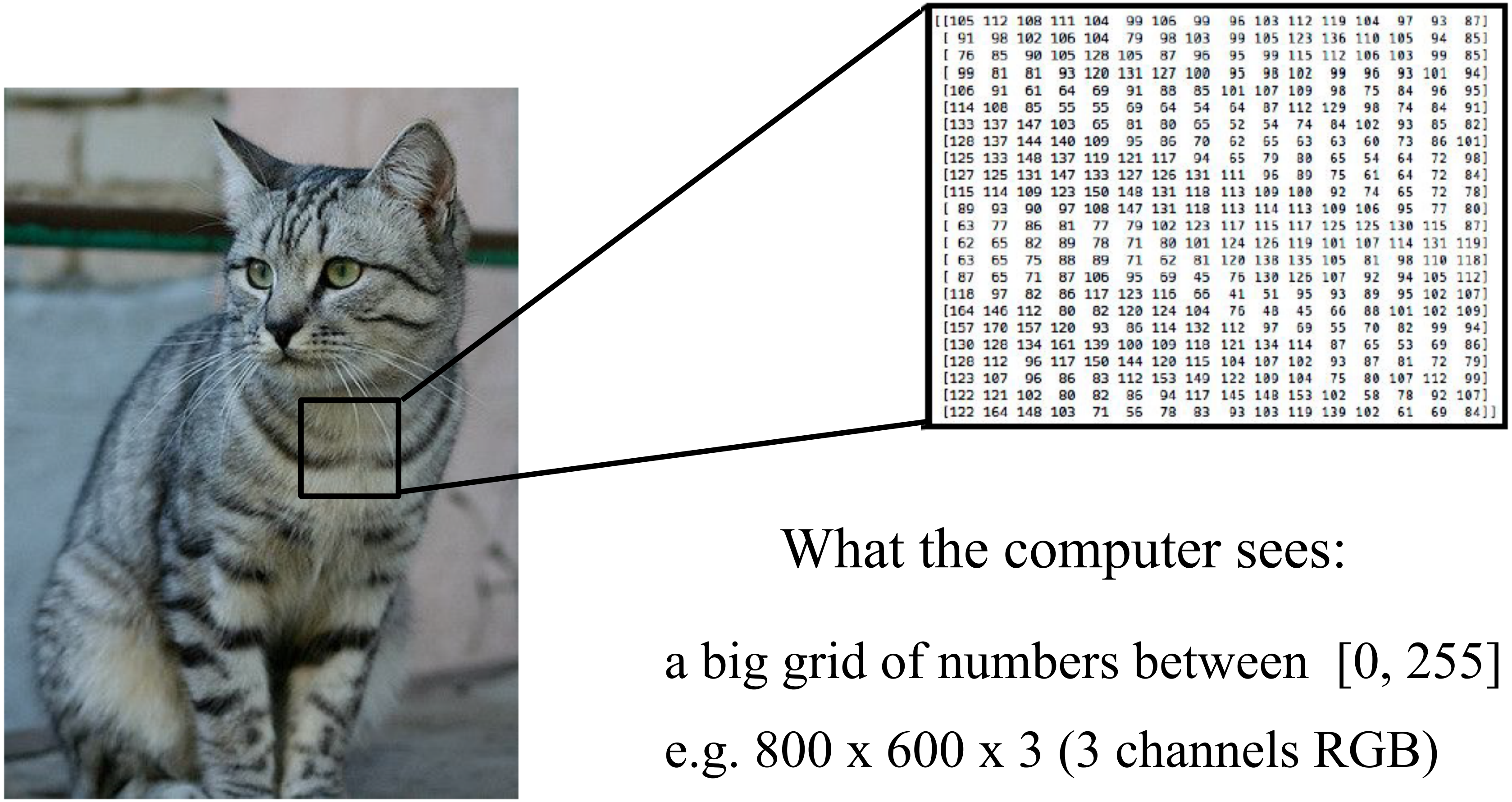
Example from the course cs231n at Stanford.

Image Classification

Given a set of discrete labels {cat, dog, frog, … }

cat

The Problem: Semantic Gap



What the computer sees:

a big grid of numbers between [0, 255]

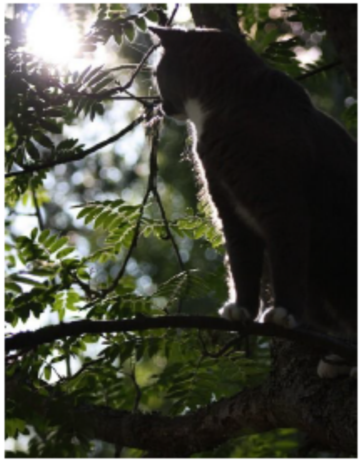
e.g. 800 x 600 x 3 (3 channels RGB)

Challenges: Viewpoint Variation

All pixels will change!

Challenges: Illumination



Challenges: Deformation

Challenges: Background Clutter

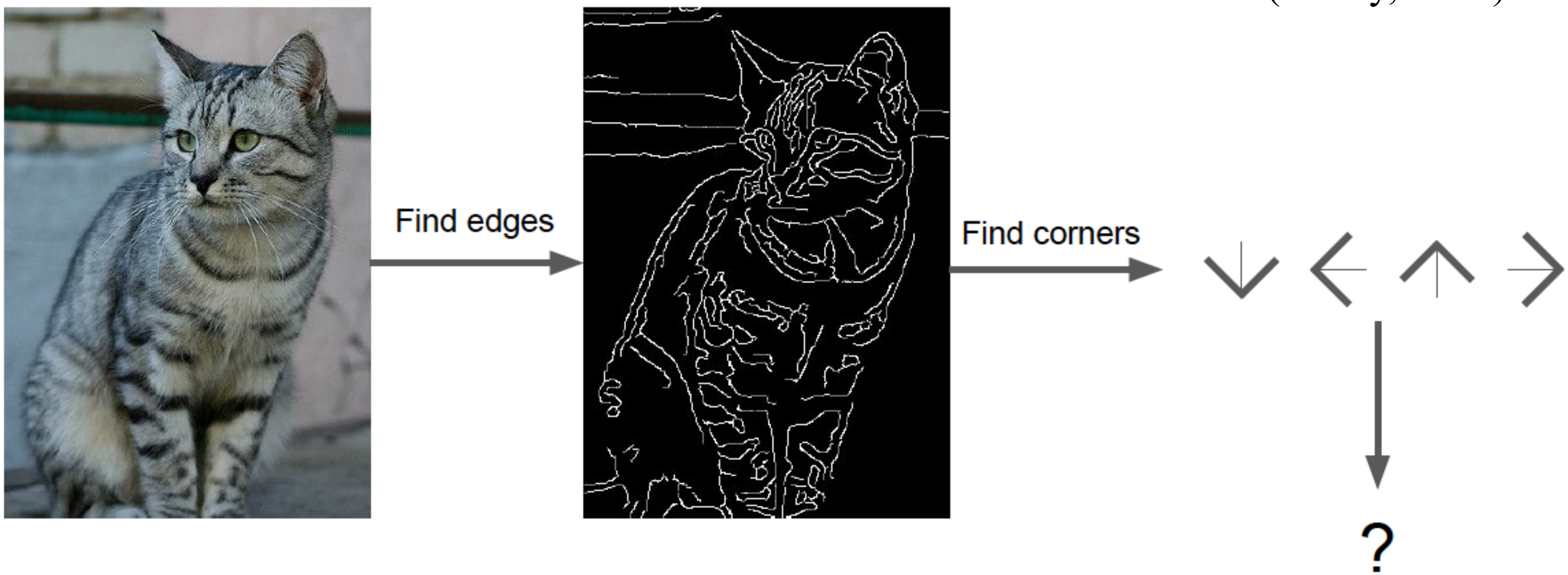
An Image Classifier

```python
def classify_image(image):
  # Some magic here?
  return class_label
```

- Unlike, e.g., sorting a list of numbers
- No obvious way to code the algorithm for classification

Rule-Based Approach

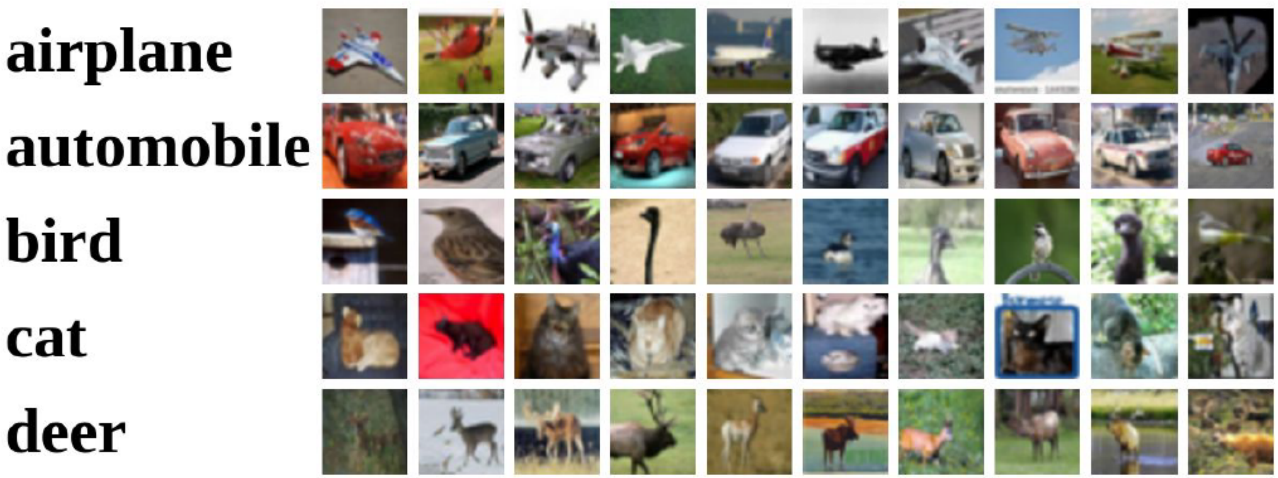
(Canny, 1986)



Data-Driven Approach

1) Data Collection: Collect a dataset of images and labels

An example of training set



Data-Driven Approach

1) Data Collection: Collect a dataset of images and labels
2) Training: Use Machine Learning to train a classifier

```
def train(images, labels):
  # Machine learning!
  return model
```

Data-Driven Approach

1) Data Collection: Collect a dataset of images and labels
2) Training: Use Machine Learning to train a classifier
3) Testing: Apply the classifier to new images

```
def predict(model, test_images):
  # Use model to predict labels
  return test_labels
```

First Classifier: Nearest Neighbor

```
def train(images, labels):
  # Machine learning!
  return model
```

→ Memorize all images/labels

```
def predict(model, test_images):
  # Use model to predict labels
  return test_labels
```

→ Output the label of the most similar training image

How to Measure Similarity

$$l_1 \text{ distance: } d(A,B) = \sum_i |A_i - B_i|$$

test image

56	32	10	18
90	23	128	133
24	26	178	200
2	0	255	220

\-

training image

10	20	24	17
8	10	89	100
12	16	178	170
4	32	233	112

=

pixel-wise absolute value differences

46	12	14	1
82	13	39	33
12	10	0	30
2	32	22	108

add → 456

Example on Dataset CIFAR10

10 classes; 50,000 training images; 10,000 testing images; size: 32 x 32

Example on Dataset CIFAR10

Test images and nearest neighbors

Complexity of Nearest Neighbor Classifier

Q: With n training examples, how fast is training and prediction ?

A: Training $O(1)$; Prediction $O(n)$

This is bad:

- We want classifiers that are fast at prediction
- Slow for training might be OK

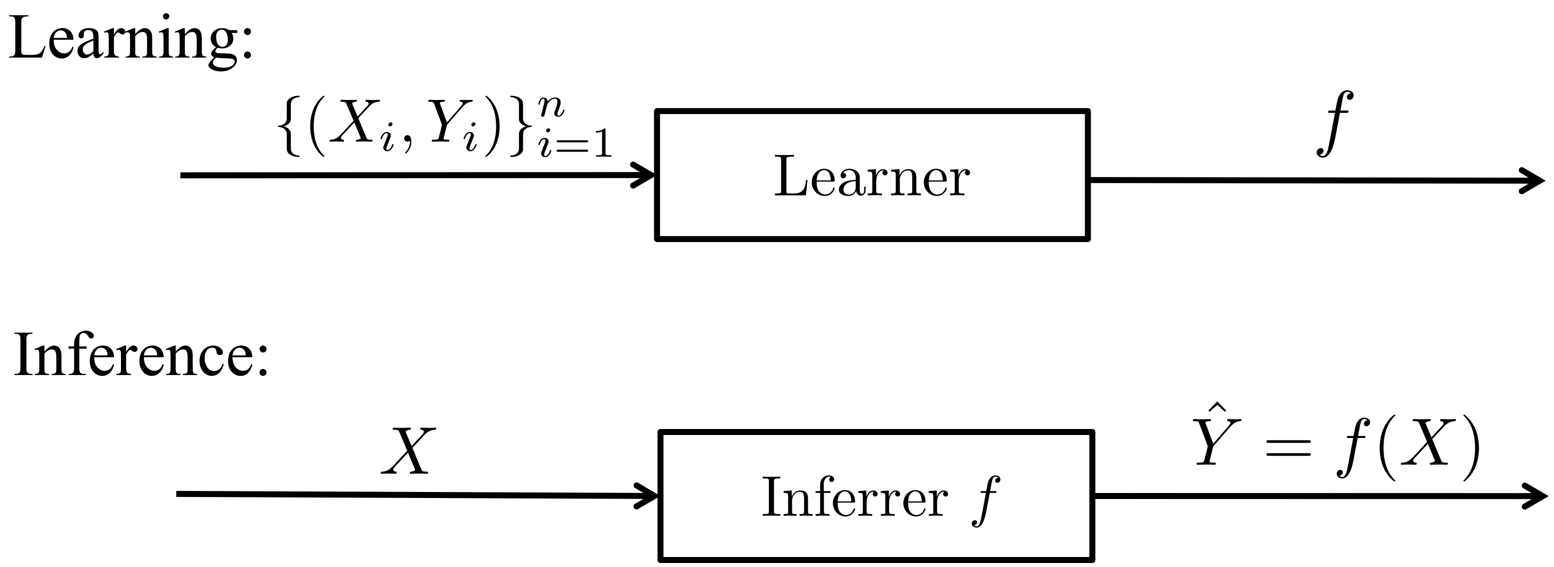
Alternatives: Support Vector Machine, Neural Network…

Summary

- Machine Learning Approach to Image Classification
  - Start with a training set of (image, label) pairs
  - Predict labels on test set

- Nearest Neighbor Classifier
  - Training: memorize all (image, label) pairs
  - Prediction: output the label of the nearest neighbor

ML Framework



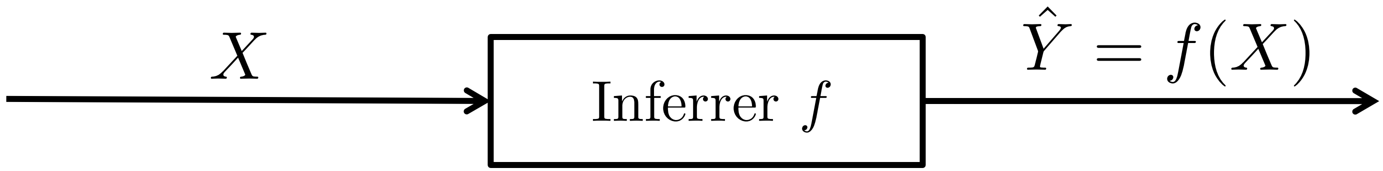
The common theme of ML is a prediction problem:

- Learn a predictor f from the training data $\{(X_i, Y_i)\}_{i=1}^n$
- Apply f at the inference stage to predict Y based on X

ML Framework

Learning:

Inference:



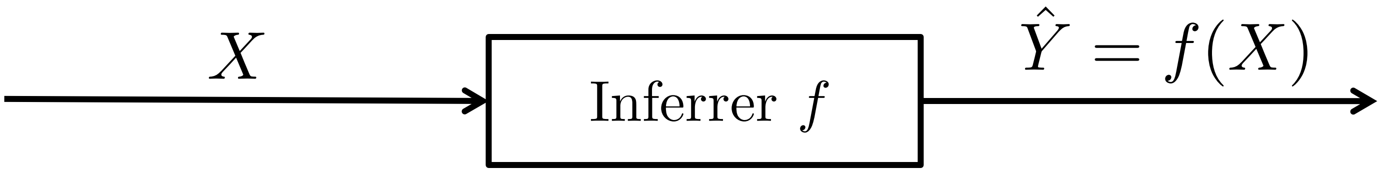
In image classification:

- $\{(X_i, Y_i)\}_{i=1}^n$ = (image, label) pairs
- f = classifier
- X = new image
- Y = true label of new image
- $\hat{Y}$ = predicted label

ML Framework

Learning:

Inference:



In regression:

- X, Y are real vectors
- f is a predictor
- $\hat{Y}$ is predicted output

Supervised vs. Unsupervised Learning

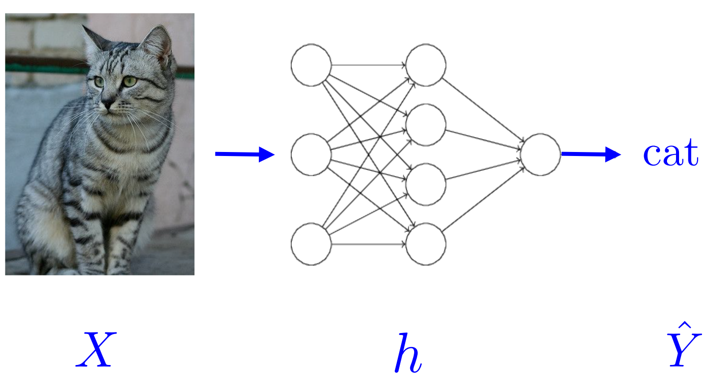
Supervised Learning

Data: $\{(x_i, y_i)\}_{i=1}^n$

Goal: Learn function h; predict $\hat{Y} = h(X)$

Example: Classification

Given $\{(\text{image}_i, \text{label}_i)\}_{i=1}^n$,
learn how to classify



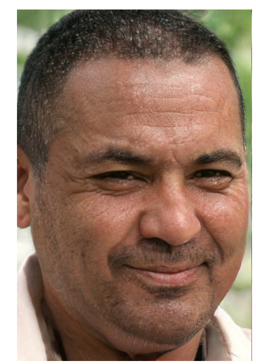
Unsupervised Learning

Data: $\{x_i\}_{i=1}^n$

Goal: Learn underlying structure of data

Example: Generative model

Given $\{\text{face image}_i\}_{i=1}^n$,
learn how to generate



By GAN (Generative Adversarial Network)!

What can we learn from this course?

Disclaimer

This course is NOT for

- Applications to specific domains (e.g. CV, NLP…)
- Implementation of ML algorithms (w/ Tensorflow, Pytorch…)

Instead, the goal of the course is to get you started in research, in particular, to make the transition from knowing how to implement towards exploring why to do this and how to do better.

Structure of the Course

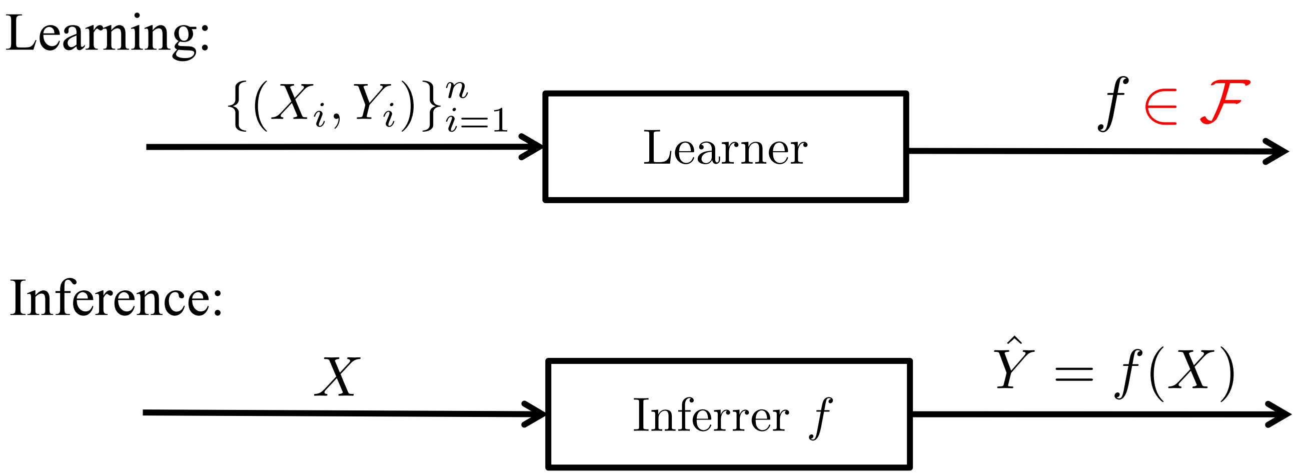
Part II: Supervised Learning

Part I: Foundations — Part III: Unsupervised Learning

Part IV: Additional Learning Models

Remark: Each part takes about 6 lectures, and includes fundamentals (4-5 lectures) and research frontiers (1-2 lectures)

Part I: Foundations



We have to restrict f to $\mathcal{F}$ to avoid overfitting. Why?

- No free lunch theorem.

What kind of $\mathcal{F}$ is good, i.e. learnable? What is not learnable?

- PAC (Probably Approximately Correct) learning framework.

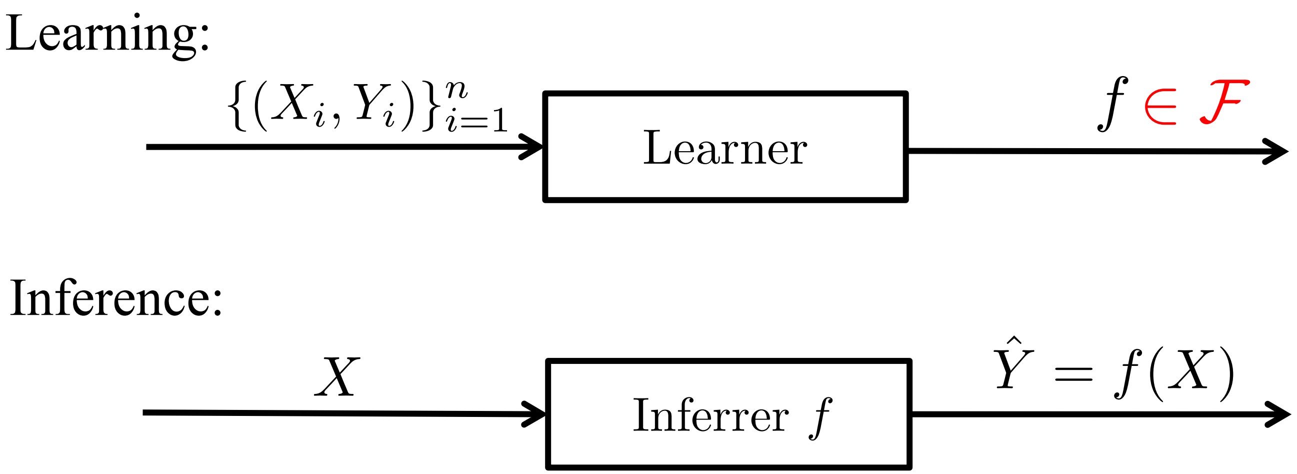
How much data do we need to learn?

- VC theory.

Research Frontiers

- PAC-Bayesian bounds, Information theoretic analysis.

Part II: Supervised Learning



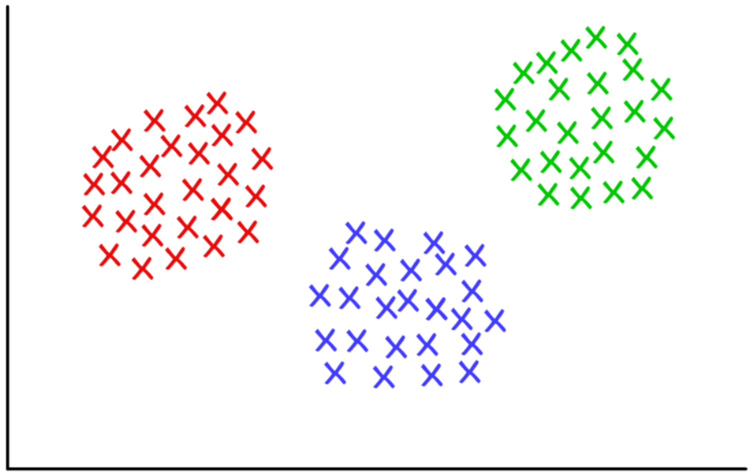
Various $\mathcal{F}$ lead to different learning models:

- Linear predictor and boosting
- Support Vector Machine
- Decision Trees

Research Frontiers

- Generalization of Neural Networks.

Part III: Unsupervised Learning



Data: $\{x_i\}_{i=1}^n$

Goal: Learn underlying structure of data

How to reduce dimensionality?
- PCA, random projection, compressed sensing.

How to cluster?
- k-means, spectral clustering.

How to generate samples that follow the same distribution?
- Generative models.

Research Frontiers
- Graph problems, Generative Adversarial Networks.

Part IV: Additional Learning Models

Can we learn even if $\mathcal{F}$ is not restricted?
- Minimax learning.

How we learn in real time when training data is progressively given?
- Online learning.

Other forms of learning?
- Reinforcement learning, Multi-armed bandits (MAB).

Research Frontiers
- Distributed learning, information theoretic analysis of MAB.

Logistics

Basic Information

- Lecture

  - TR 9:30-10:45 AM, CLB 109

- Office hour

  - TR 10:45-11:45 AM, Evans 314

- Course website

  https://www.eecis.udel.edu/~xwu/class/ELEG602/

- TA

  Yalin Liao, yalin@udel.edu

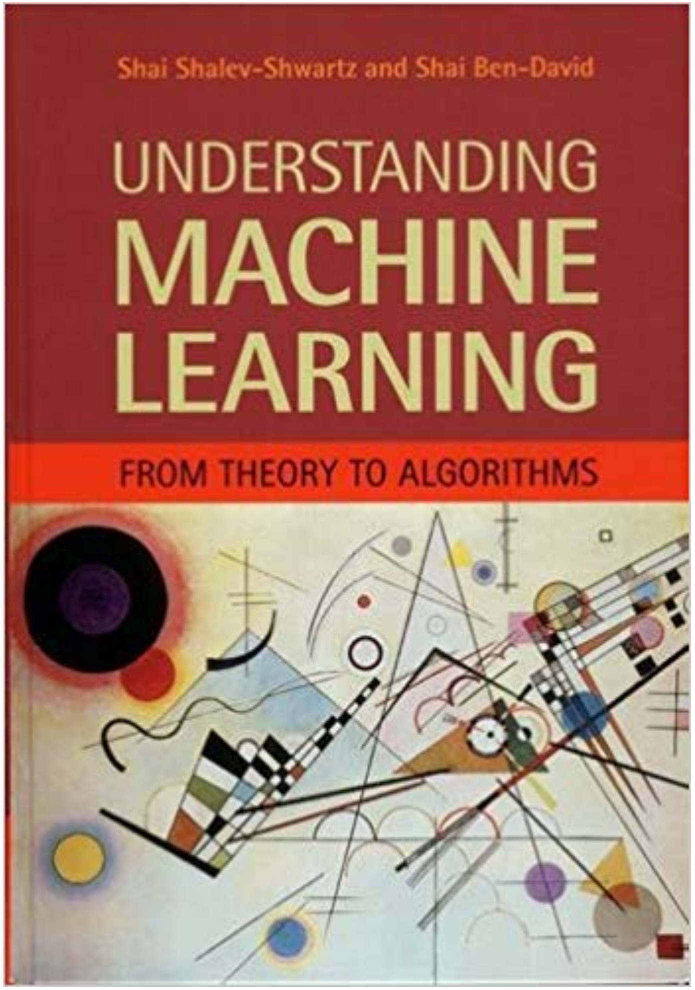
Prerequisite

- Previous exposure to ML course is strongly recommended
- Proficiency in probability theory and linear algebra
- Mathematical maturity in general

Textbook

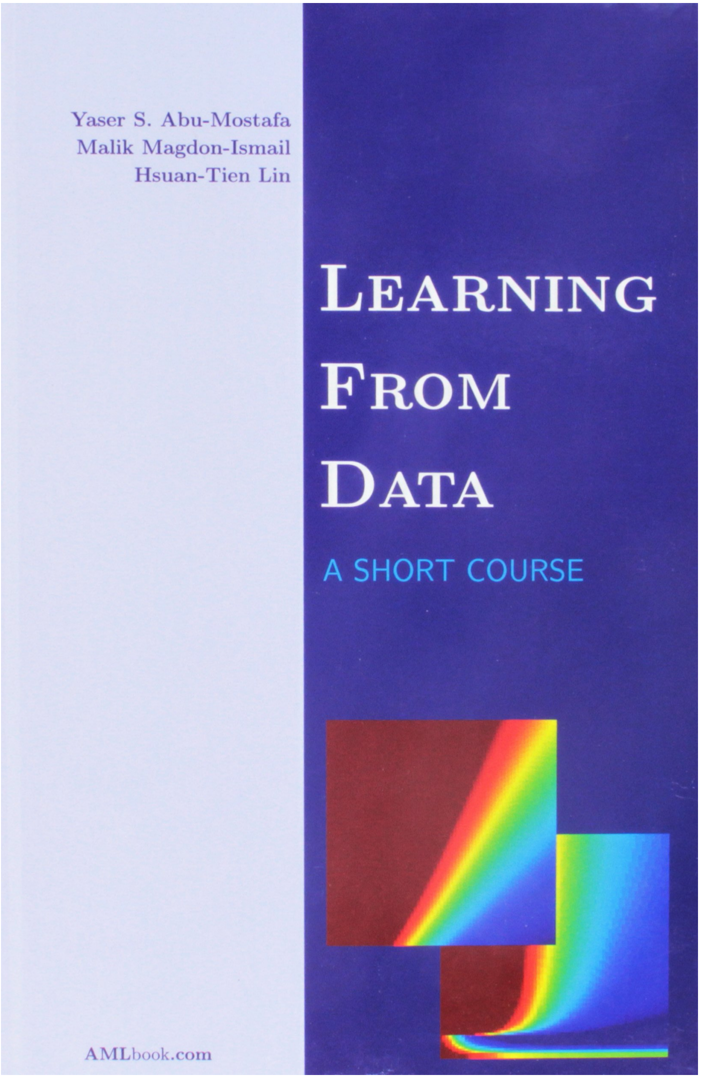
Lectures are based on slides that will be posted both on Canvas and on the course website.

Textbook



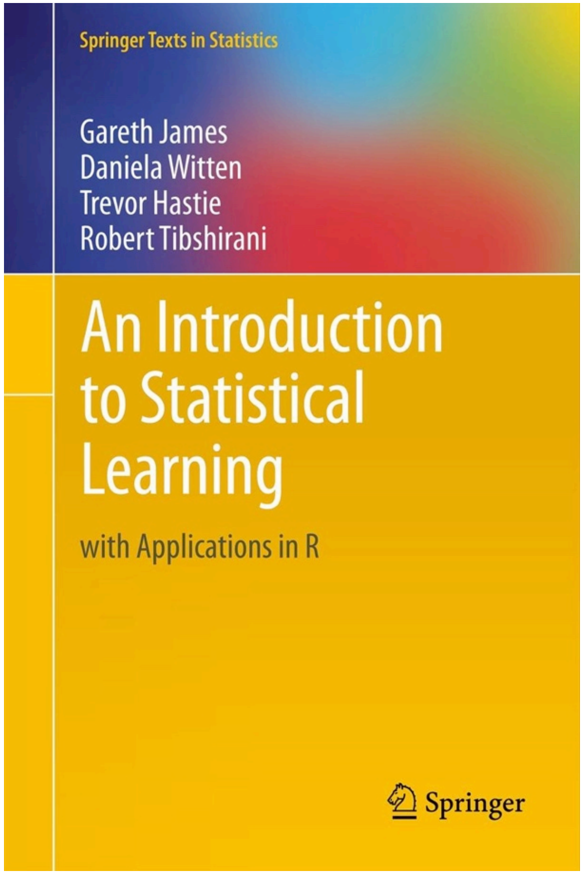
Free pdf version online

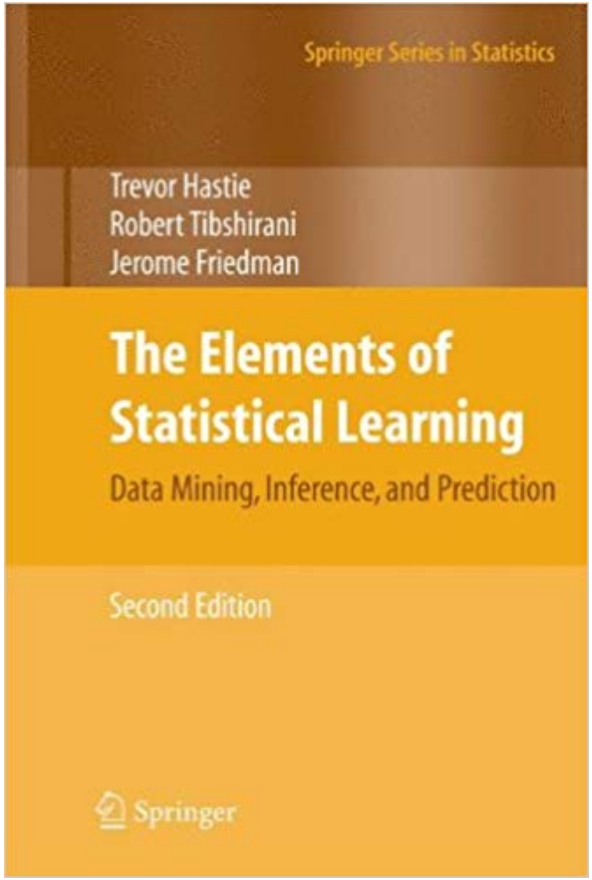
Other Recommended Books



Textbook for “Statistical learning”, taught by Prof. Gonzalo Arce

Other Recommended Books





- Classical "ISL" and "ESL"
- Both free to download

Grading

- Homework: 50 pts

  - Four HWs in total; one for each part

- Project (Presentation and Report): 50 pts + 10 bonus pts

  - You are encouraged to form groups of size 2-3 people
  - Choose one topic related to research frontiers
  - Can be either theoretical or experimental
  - Presentation in the last lecture
  - Reserve your topic with the TA early!

Questions?