ELEG/CISC 867
Advanced Machine Learning

Xiugang Wu

Dept of Electrical & Computer Engineering

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

Feb 14, 2019

Agenda

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

What is machine learning?

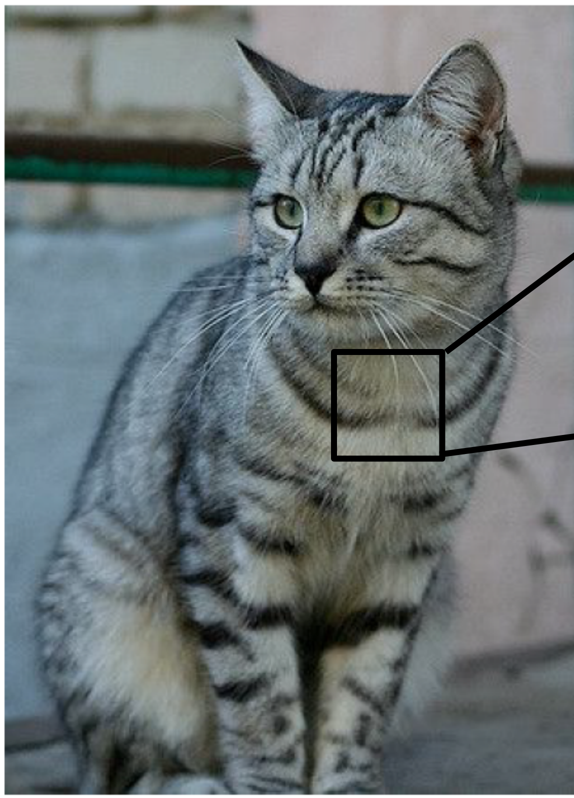
Example from the course cs231n at Stanford.

Image Classification

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

cat

The Problem: Semantic Gap



```
[[105 112 108 111 104  99 106  99  96 103 112 119 104  97  93  87]
 [ 91  98 102 106 104  79  98 103  99 105 123 136 110 105  94  85]
 [ 76  85  90 105 128 105  87  96  95  99 115 112 106 103  99  85]
 [ 99  81  81  93 120 131 127 100  95  98 102  99  96  93 101  94]
 [106  91  61  64  69  91  88  85 101 107 109  98  75  84  96  95]
 [114 108  85  55  55  69  64  54  64  87 112 129  98  74  84  91]
 [133 137 147 103  65  81  80  65  52  54  74  84 102  93  85  82]
 [128 137 144 140 109  95  86  70  62  65  63  63  60  73  86 101]
 [125 133 148 137 119 121 117  94  65  79  80  65  54  64  72  98]
 [127 125 131 147 133 127 126 131 111  96  89  75  61  64  72  84]
 [115 114 109 123 150 148 131 118 113 109 100  92  74  65  72  78]
 [ 89  93  90  97 108 147 131 118 113 114 113 109 106  95  77  80]
 [ 63  77  86  81  77  79 102 123 117 115 117 125 125 130 115  87]
 [ 62  65  82  89  78  71  80 101 124 126 119 101 107 114 131 119]
 [ 63  65  75  88  89  71  62  81 120 138 135 105  81  98 110 118]
 [ 87  65  71  87 106  95  69  45  76 130 126 107  92  94 105 112]
 [118  97  82  86 117 123 116  66  41  51  95  93  89  95 102 107]
 [164 146 112  80  82 120 124 104  76  48  45  66  88 101 102 109]
 [157 170 157 120  93  86 114 132 112  97  69  55  70  82  99  94]
 [130 128 134 161 139 100 109 118 121 134 114  87  65  53  69  86]
 [128 112  96 117 150 144 120 115 104 107 102  93  87  81  72  79]
 [123 107  96  86  83 112 153 149 122 109 104  75  80 107 112  99]
 [122 121 102  80  82  86  94 117 145 148 153 102  58  78  92 107]
 [122 164 148 103  71  56  78  83  93 103 119 139 102  61  69  84]]
```

What the computer sees:

a big grid of numbers between [0, 255]

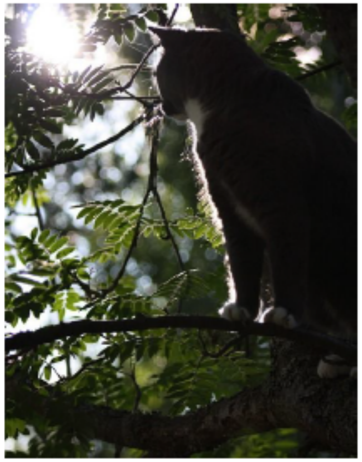
e.g. 800 x 600 x 3 (3 channels RGB)

Challenges: Viewpoint Variation

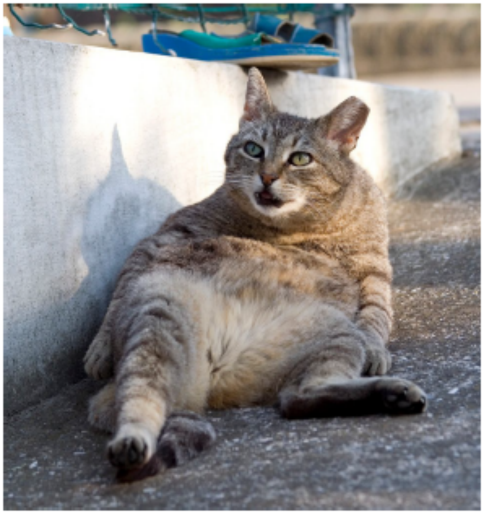
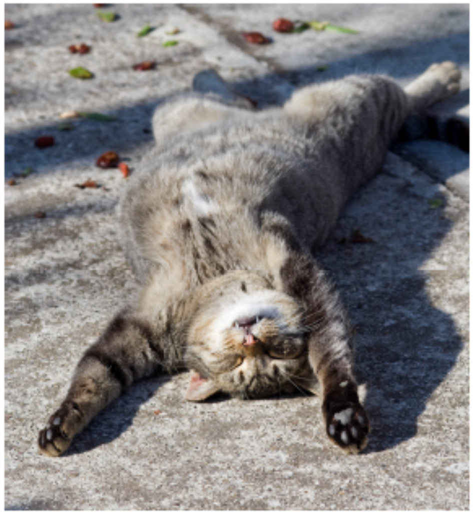
All pixels will change!

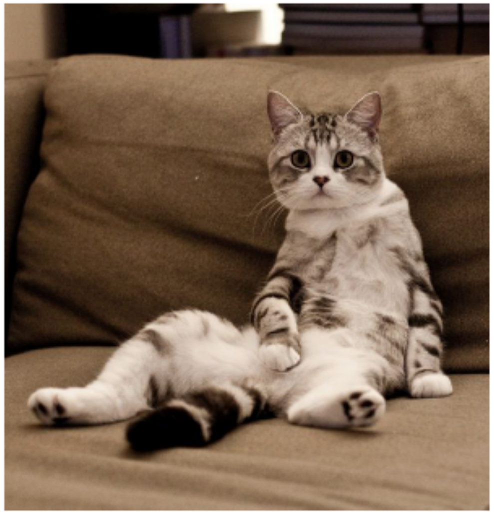
Challenges: Illumination



Challenges: Deformation



Challenges: Background Clutter

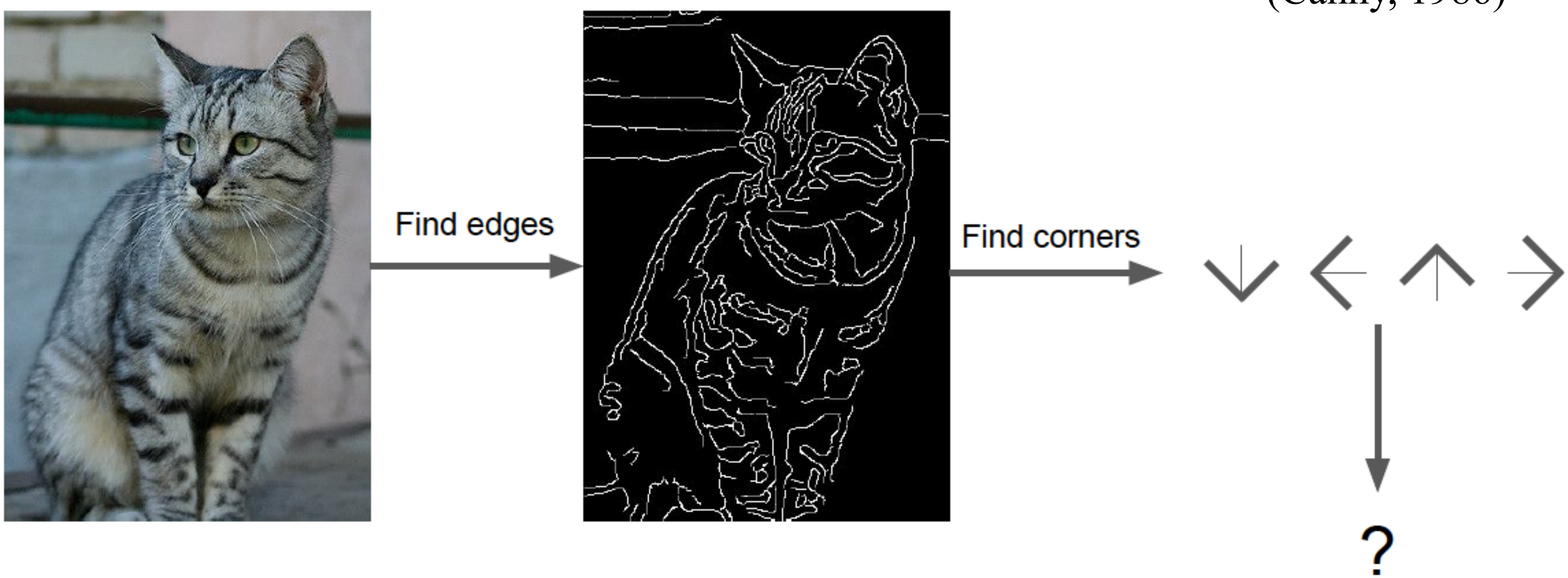
An Image Classifier

```
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)

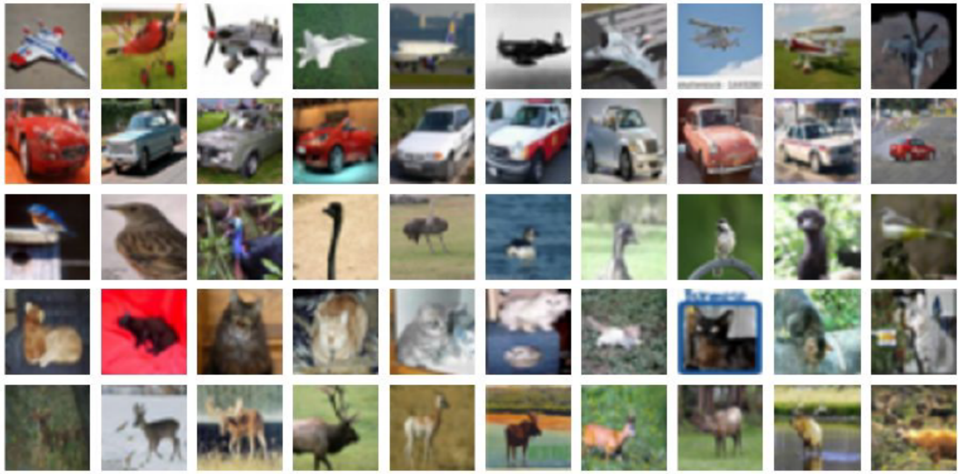


Data-Driven Approach

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

An example of training set

airplane
automobile
bird
cat
deer



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

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

$$\left| \overset{\text{test image}}{\begin{bmatrix} 56 & 32 & 10 & 18 \\ 90 & 23 & 128 & 133 \\ 24 & 26 & 178 & 200 \\ 2 & 0 & 255 & 220 \end{bmatrix}} - \overset{\text{training image}}{\begin{bmatrix} 10 & 20 & 24 & 17 \\ 8 & 10 & 89 & 100 \\ 12 & 16 & 178 & 170 \\ 4 & 32 & 233 & 112 \end{bmatrix}} \right| = \overset{\text{pixel-wise absolute value differences}}{\begin{bmatrix} 46 & 12 & 14 & 1 \\ 82 & 13 & 39 & 33 \\ 12 & 10 & 0 & 30 \\ 2 & 32 & 22 & 108 \end{bmatrix}} \xrightarrow{\text{add}} 456$$

Example on Dataset CIFAR10

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

airplane
automobile
bird
cat
deer
dog
frog
horse
ship
truck

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 is 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

Learning:

Inference:

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:

$$\xrightarrow{\{(X_i, Y_i)\}_{i=1}^{n}} \boxed{\text{Learner}} \xrightarrow{f}$$

Inference:

$$\xrightarrow{X} \boxed{\text{Inferrer } f} \xrightarrow{\hat{Y} = f(X)}$$

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

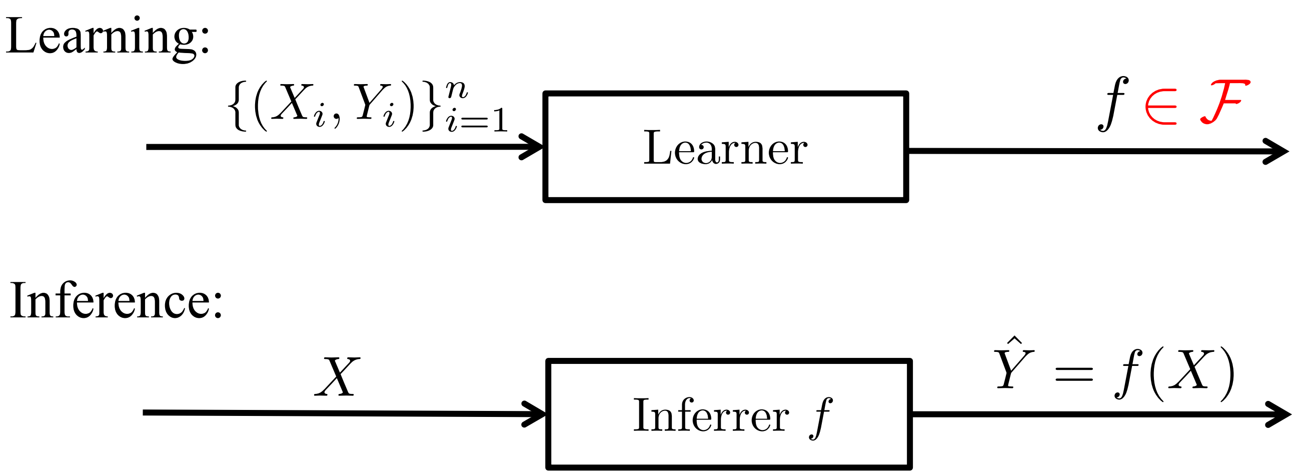
What can we learn from this course?

Top-Down Approach

$$\text{Theory} \Rightarrow \text{Algorithm} \Rightarrow \text{Implementation}$$

- Part I: Foundation
- Part II: From Theory to Algorithms
- Part III: Advanced Topics

Part I: Foundation



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

- No free lunch theorem.

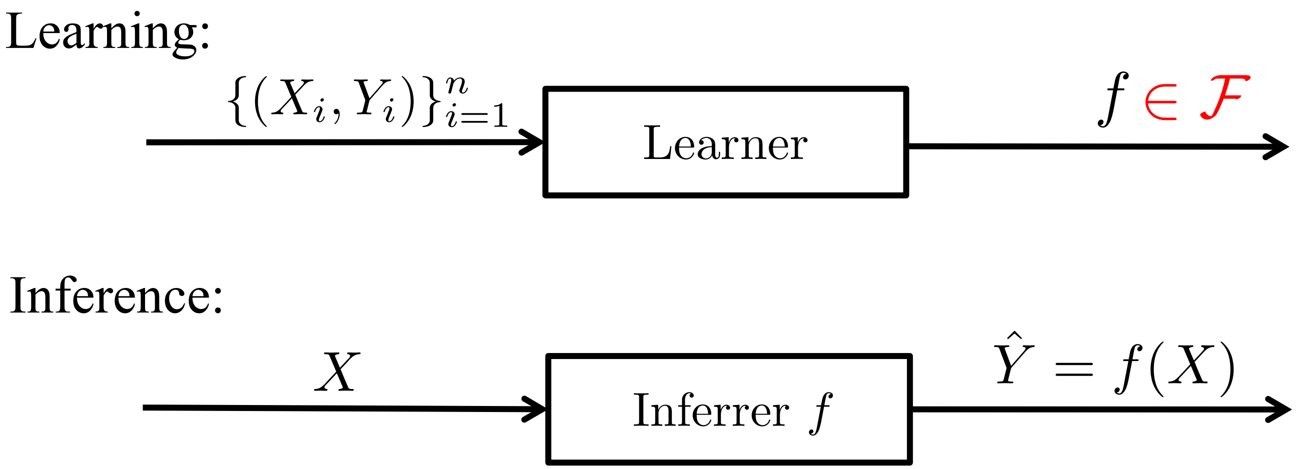
What kind of $\mathcal{F}$ is good, i.e. learnable? What is not learnable?

- PAC (Probably Approximately Correct) learning framework.

How much data do we need to learn?

- VC theory.

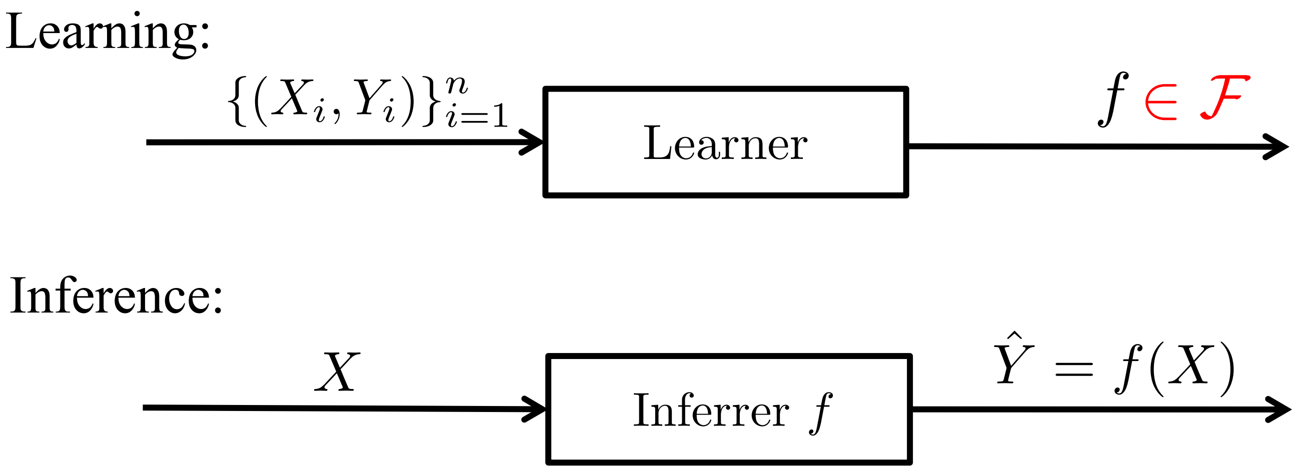
Part II: From Theory to Algorithms



Various $\mathcal{F}$ lead to various learning models.

- Linear predictor and boosting
- Support vector machine
- Decision trees

Part III: Advanced Topics



Can we learn even if f is not restricted?

- Yes, minimax learning.

Can we learn in real time when training data is progressively given?

- Yes, online learning.

Why is deep neural network working?

- An information theoretic perspective.

Other forms of learning?

- Unsupervised learning, reinforcement learning...

Logistics

Lecture and Office Hour

- Lecture

  - TR 9:30−10:45 AM

  - Gore 308

- Office hour

  - TR 11:00-12:00 AM

  - Evans 314

- Course website

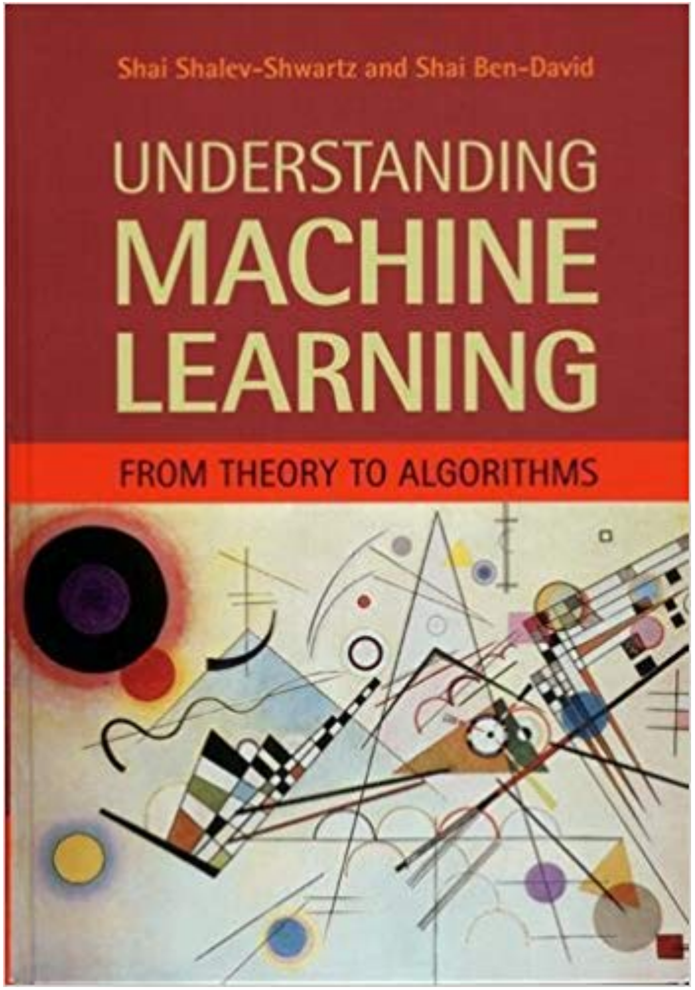
  https://www.eecis.udel.edu/~xwu/class/ELEG867/

Prerequisite

- Undergraduate-level probability theory and linear algebra; mathematical maturity in general

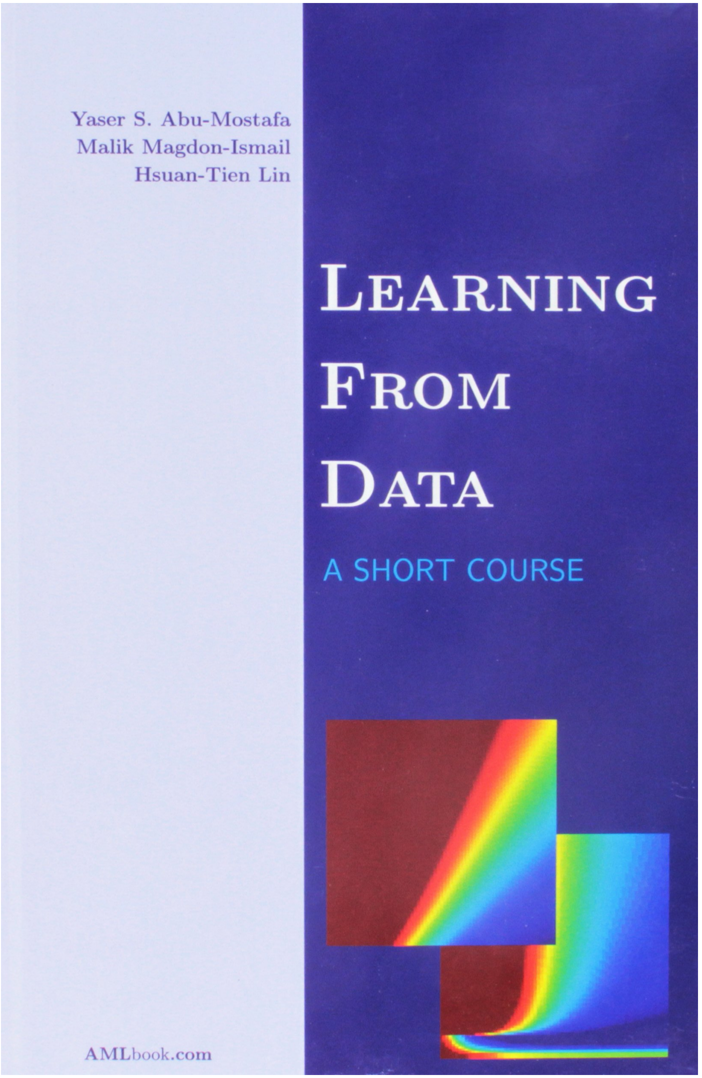
- Previous exposure to machine learning is preferred but not mandatory

Textbook



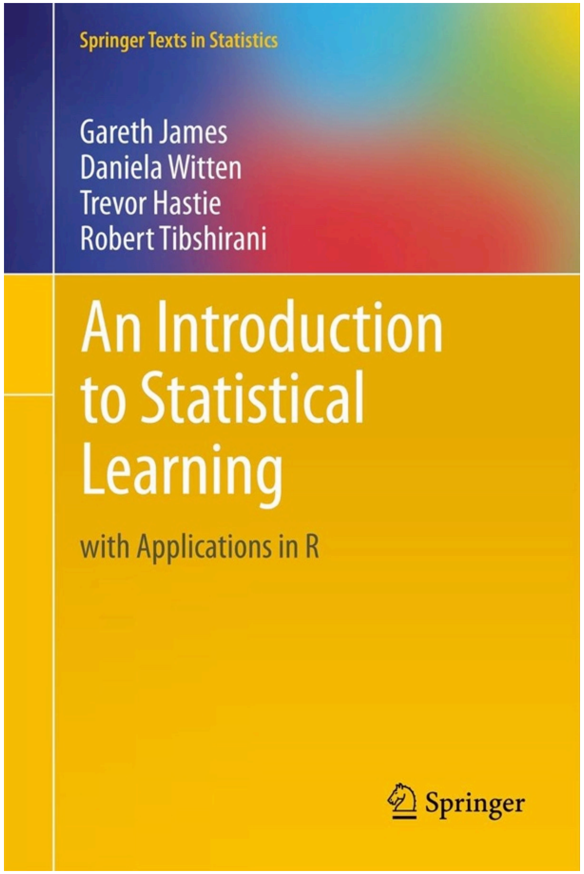
- Free pdf version online
- Lecture notes on course website

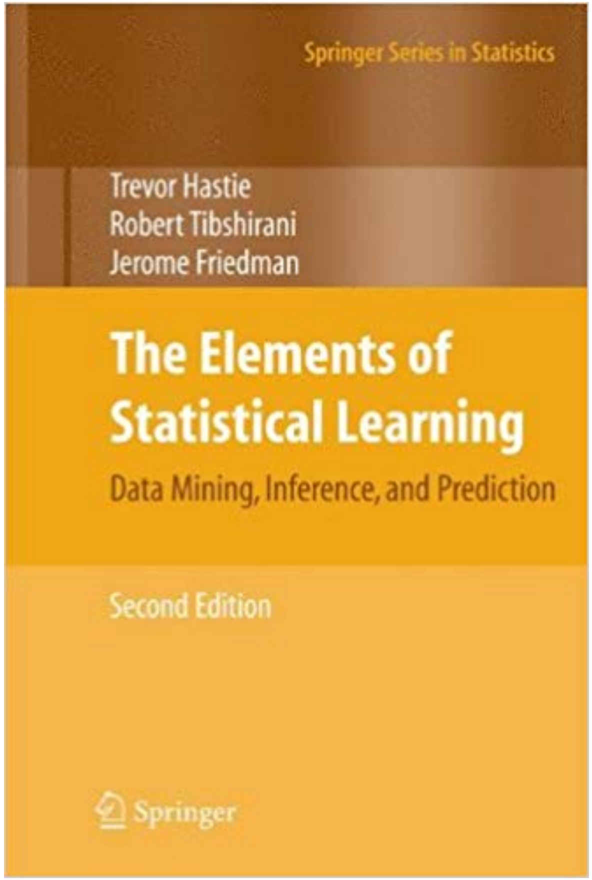
Other Recommended Books



- Text for "Statistical learning", by Prof. Gonzalo Arce

Other Recommended Books





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

Grading

- Attendance: 10 points

  - Six random samples (after the class is stabilized)

  - Two points for each attendance (5 out of 6 gives you max 10 pts)

- Project (Presentation and Report): 40 points

  - Can be either theoretical or experimental

  - Topics will be posted towards midterm

- Final: 50 points + 10 bonus points

  - Closed book with one letter-size aid-sheet allowed

  - Homeworks (three in total, one for each part) are not graded but crucial for after-exam happiness

Questions?