https://xkcd.com/894/

1

INTRODUCTION TO MACHINE LEARNING

David Kauchak
CS 158 – Fall 2025

2

Introductions

Dr. | Prof | Professor
Dave | Kauchak

Pronouns: he/him/his

3

Why are you here?

What is Machine Learning?

Why are you taking this course?

What topics would you like to see covered?

4

Machine Learning is…

Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence.

WIKIPEDIA
The Free Encyclopedia

5

Machine Learning is…

Machine learning is programming computers to optimize a performance criterion using example data or past experience.
-- Ethem Alpaydin

The goal of machine learning is to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest.
-- Kevin P. Murphy

The field of pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions.
-- Christopher M. Bishop

6

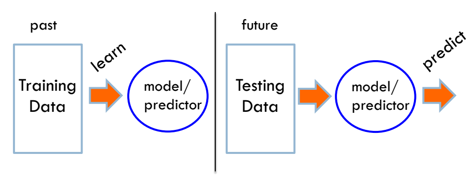
Machine Learning is…

Machine learning is about predicting the future based on the past.
-- Hal Daume III

7

Machine Learning is…

Machine learning is about predicting the future based on the past.
-- Hal Daume III



8

Machine Learning, aka

data mining: data analysis, not prediction, though often involves some shared techniques

inference and/or *estimation* in statistics

pattern recognition in engineering

signal processing in electrical engineering

induction

optimization

9

Goals of the course: learn about…

Different machine learning problems

Common techniques/tools used
- theoretical understanding
- practical implementation

Proper experimentation and evaluation

10

Goals of the course

Be able to laugh at these signs
(or at least know why one might…)

11

Goals of the course

90s: neural networks

early 2000s: support vector machines

after that: probabilistic models (aka graphical models)

currently: neural networks, deep learning

Why mention this now?

12

Course expectations

Plan to stay busy!

Applied class, so lots of programming

Machine learning involves math

13

Where we've been!

Our ML suite:

29 classes

2951 lines of code

14

Where we've been!

Our ML suite:

- Supports 7 classifiers
  - Decision Tree
  - Perceptron
  - Average Perceptron
  - Gradient descent
    - 2 loss functions
    - 2 regularization methods
  - K-NN
  - Naïve Bayes
  - 2 layer neural network
- Supports two types of data normalization
  - feature normalization
  - example normalization
- Supports two types of meta-classifiers
  - OVA
  - AVA

15

Administrative

Course page:
http://www.cs.pomona.edu/classes/cs158/

Assignments
- Weekly
- Mostly programming (Java, mostly)
- Some written/write-up
- Generally due Sunday evenings

Two "midterm" exams and one final (all time limited take home)

Late Policy

Collaboration

ChatGPT (and similar tools)

16

Other things to note

Videos before class

Lots of class participation!

Read the book (it's good)

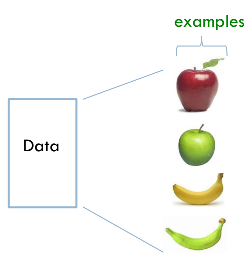
17

Machine learning problems

What high-level machine learning problems have you seen or heard of before?

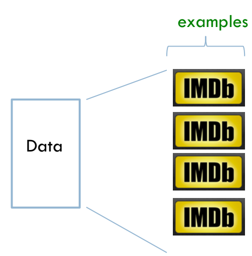
18

Data

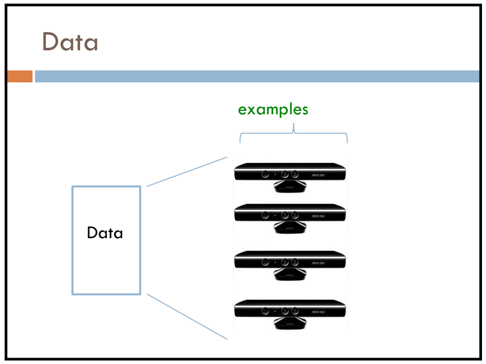


19

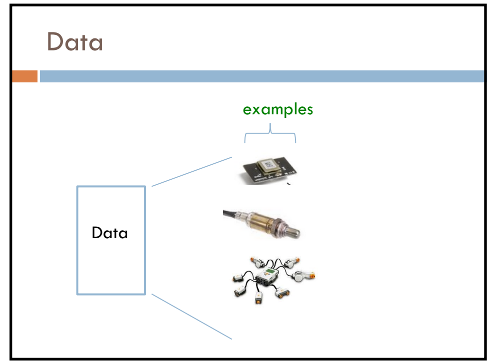
Data



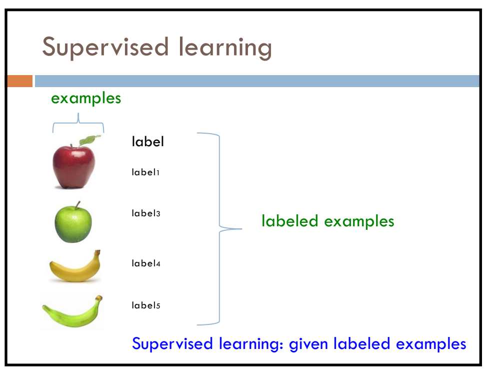
20


Data
examples
Data

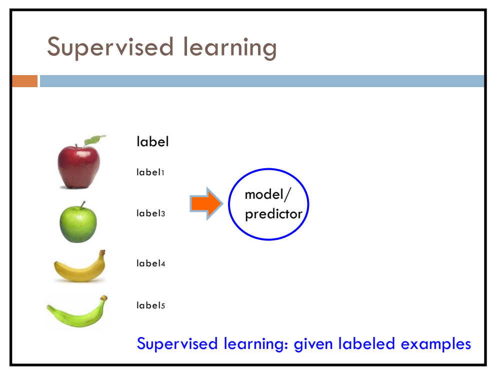
21


Data
examples
Data

22


Supervised learning
examples
label
label$_1$
label$_3$
label$_4$
label$_5$
labeled examples
Supervised learning: given labeled examples

23


Supervised learning
label
label$_1$
label$_3$
label$_4$
label$_5$
model/
predictor
Supervised learning: given labeled examples

24

Supervised learning

model/ predictor

predicted label

Supervised learning: learn to predict new example

25

Supervised learning: classification

label

apple

apple

banana

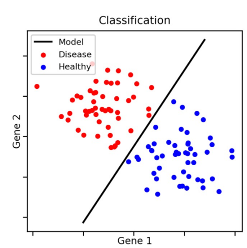
banana

Classification: a finite set of labels

Supervised learning: given labeled examples

26

Classification Example



27

Classification Applications

Face recognition

Character recognition

Spam detection

Medical diagnosis: From symptoms to illnesses

Biometrics: Recognition/authentication using physical and/or behavioral characteristics: Face, iris, signature, etc

…

28

Supervised learning: regression

label
-4.5
10.1
3.2
4.3

Regression: label is real-valued

Supervised learning: given labeled examples

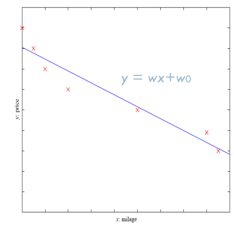
29

Regression Example

Price of a used car

x : car attributes
(e.g. mileage)

y : price



30

Regression Applications

Economics/Finance: predict the value of a stock

Epidemiology

Car/plane navigation: angle of the steering wheel, acceleration, …

Temporal trends: weather over time

…

31

Supervised learning: ranking

label
1
4
2
3

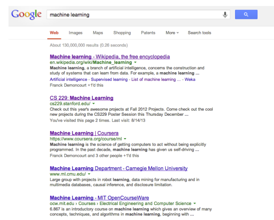
Ranking: label is a ranking

Supervised learning: given labeled examples

32

Ranking example

Given a query and a set of web pages, rank them according to relevance



33

Ranking Applications

User preference, e.g. movie ranking
iTunes

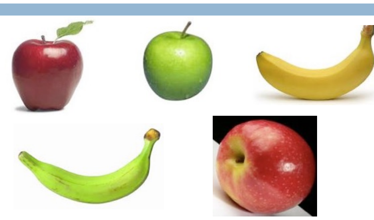
flight search (search in general)

reranking N-best output lists

…

34

Unsupervised learning



Unupervised learning: given data, i.e. examples, but no labels

35

Unsupervised learning applications

learn clusters/groups without any label

customer segmentation (i.e. grouping)

image compression

bioinformatics: learn motifs

…

36

Reinforcement learning

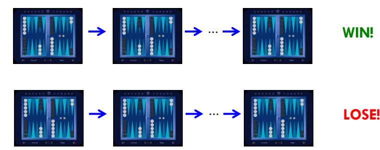
left, right, straight, left, left, left, straight	GOOD
left, straight, straight, left, right, straight, straight	BAD
left, right, straight, left, left, left, straight	18.5
left, straight, straight, left, right, straight, straight	-3

Given a *sequence* of examples/states and a *reward* after completing that sequence, learn to predict the action to take for an individual example/state

37

Reinforcement learning example

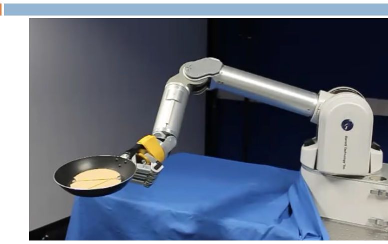
Backgammon



Given sequences of moves and whether or not the player won at the end, learn to make good moves

38

Reinforcement learning example



https://www.youtube.com/watch?v=W_gxLKSsSIE

39

Other learning variations

What data is available:

- Supervised, unsupervised, reinforcement learning
- semi-supervised, active learning, …

How are we getting the data:

- online vs. offline learning

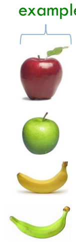
Type of model:

- generative vs. discriminative
- parametric vs. non-parametric

40

Representing examples

examples

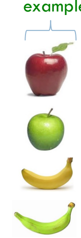


What is an example?
How is it represented?

41

Features

examples



features

$f_1, f_2, f_3, \ldots, f_n$

$f_1, f_2, f_3, \ldots, f_n$

$f_1, f_2, f_3, \ldots, f_n$

$f_1, f_2, f_3, \ldots, f_n$

How our algorithms actually "view" the data

Features are the questions we can ask about the examples

42

Features

examples

features

red, round, leaf, 3oz, …

green, round, no leaf, 4oz, …

yellow, curved, no leaf, 8oz, …

green, curved, no leaf, 7oz, …

How our algorithms actually "view" the data

Features are the questions we can ask about the examples

43

Classification revisited

examples	label
red, round, leaf, 3oz, …	apple
green, round, no leaf, 4oz, …	apple
yellow, curved, no leaf, 8oz, …	banana
green, curved, no leaf, 7oz, …	banana

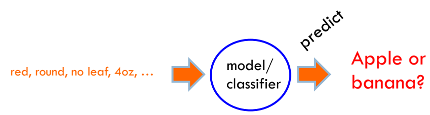
learn

model/ classifier

During learning/training/induction, learn a model of what distinguishes apples and bananas *based on the features*

44

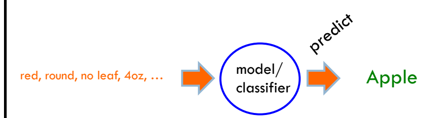
Classification revisited



The model can then classify a new example *based on the features*

45

Classification revisited



Why?

The model can then classify a new example *based on the features*

46

Classification revisited

Training data / Test set

examples	label	
red, round, leaf, 3oz, …	apple	
green, round, no leaf, 4oz, …	apple	red, round, no leaf, 4oz, … ?
yellow, curved, no leaf, 4oz, …	banana	
green, curved, no leaf, 5oz, …	banana	

47

Classification revisited

Training data / Test set

examples	label	
red, round, leaf, 3oz, …	apple	
green, round, no leaf, 4oz, …	apple	red, round, no leaf, 4oz, … ?
yellow, curved, no leaf, 4oz, …	banana	
green, curved, no leaf, 5oz, …	banana	

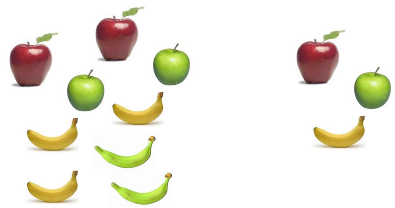
Learning is about ***generalizing*** from the training data

What does this assume about the training and test set?

48

Past predicts future

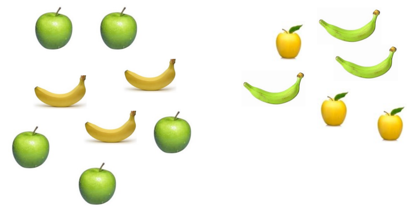
Training data | Test set



49

Past predicts future

Training data | Test set

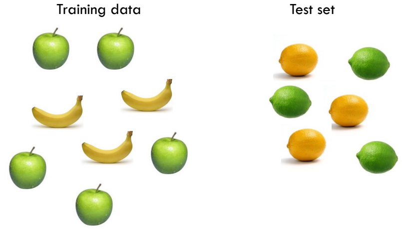


Not always the case, but we'll often assume it is!

50

Past predicts future

Training data | Test set



Not always the case, but we'll often assume it is!

51

More technically…

We are going to use the *probabilistic model* of learning

There is some probability distribution over example/label pairs called the *data generating distribution*

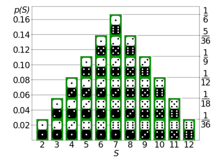
Both the training data **and** the test set are generated based on this distribution

What is a probability distribution?

52

Probability distribution

Describes how likely (i.e., probable) certain events are

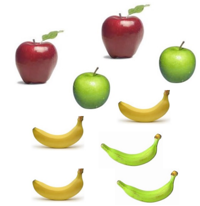


- Describes probabilities for all possible events
- Probabilities are between 0 and 1 (inclusive)
- Sum of probabilities over all events is 1

53

Probability distribution

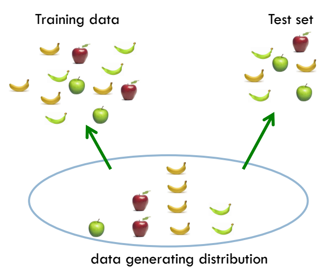
Training data



High probability	Low probability
round apples	curved apples
curved bananas	red bananas
apples with leaves	yellow apples
…	…

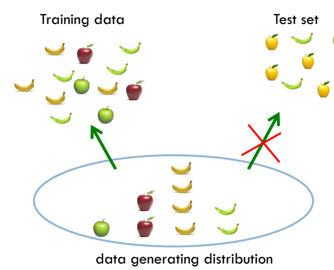
54

data generating distribution



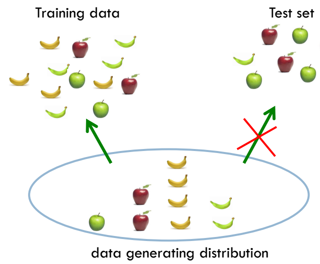
55

data generating distribution



56

data generating distribution



57