

INTRODUCTION TO MACHINE LEARNING

Dave Kauchak, Alexandra Papoutsaki, Zilong Ye
CS 51A – Spring 2022

Machine Learning is...

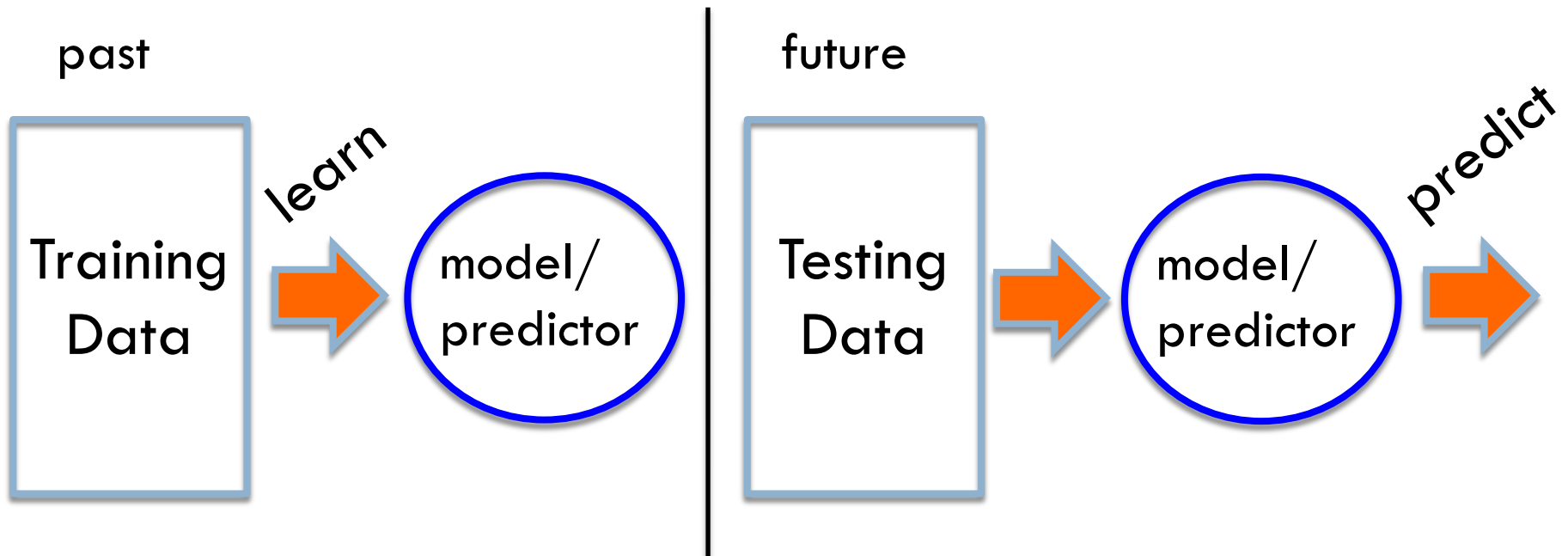
Machine learning is about predicting the future based on the past.

-- Hal Daume III

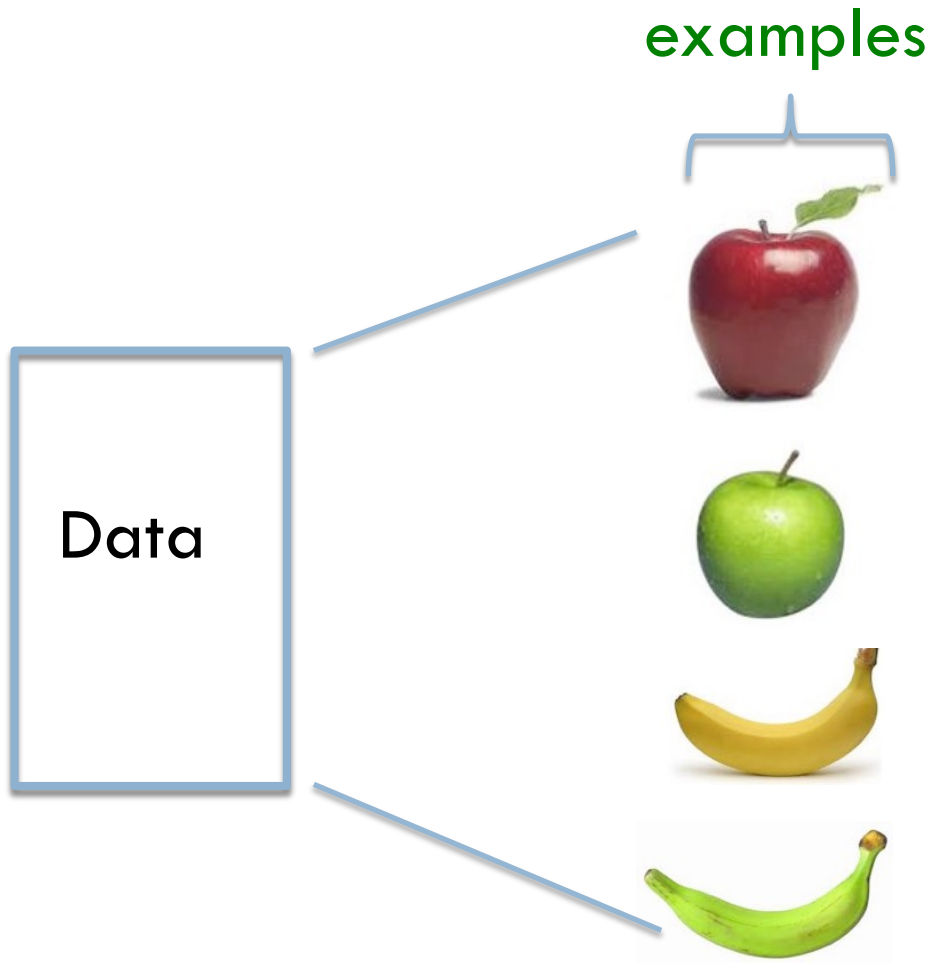


Machine Learning is...

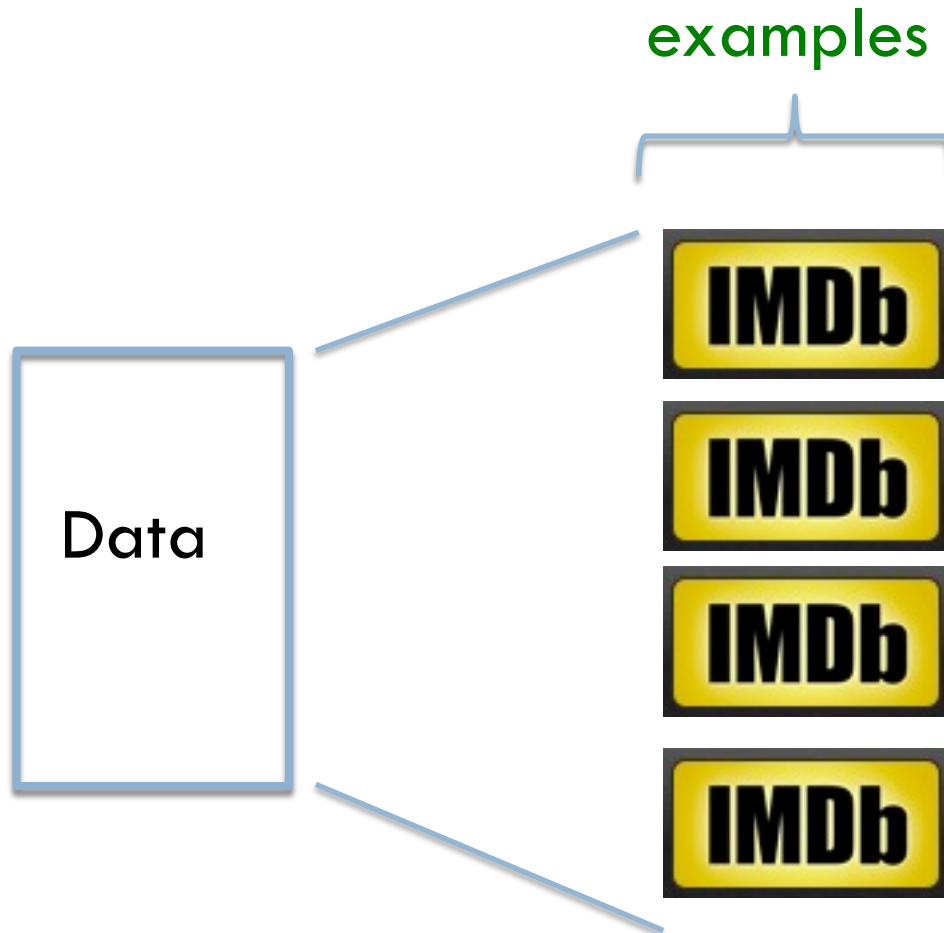
Machine learning is about predicting the future based on the past.
-- Hal Daume III



Data



Data



Data

examples

Data



Data

examples

Data



Supervised learning

examples



label

label₁

label₃

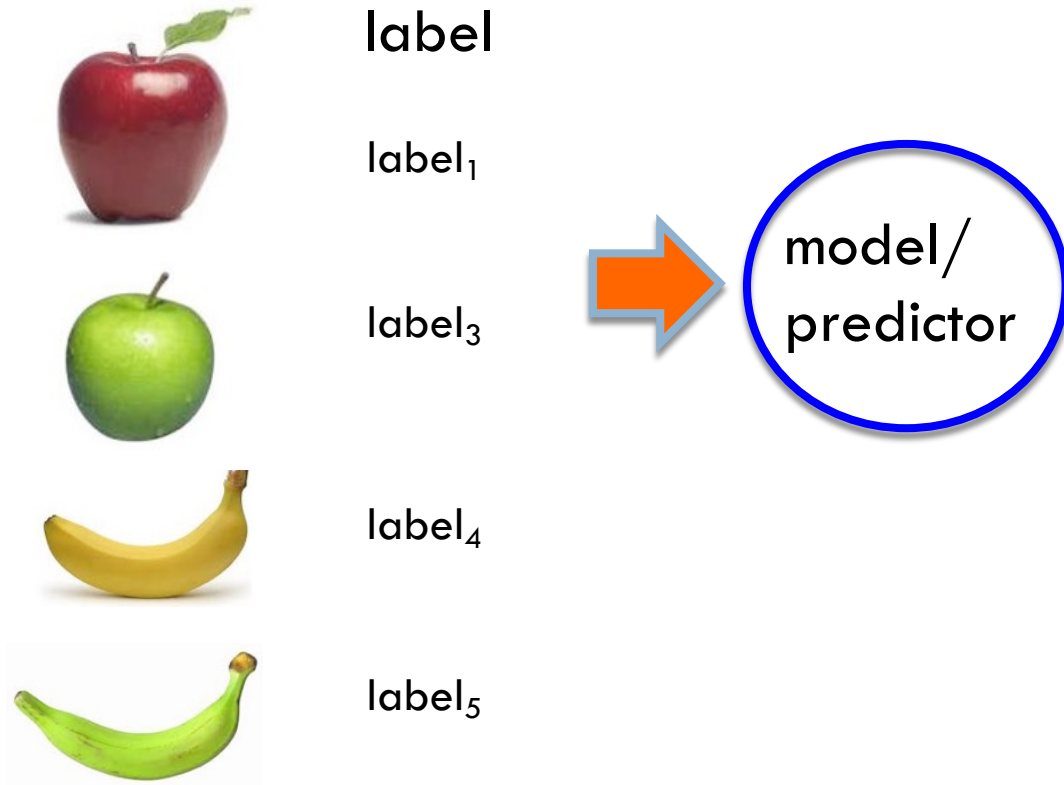
label₄

label₅

labeled examples

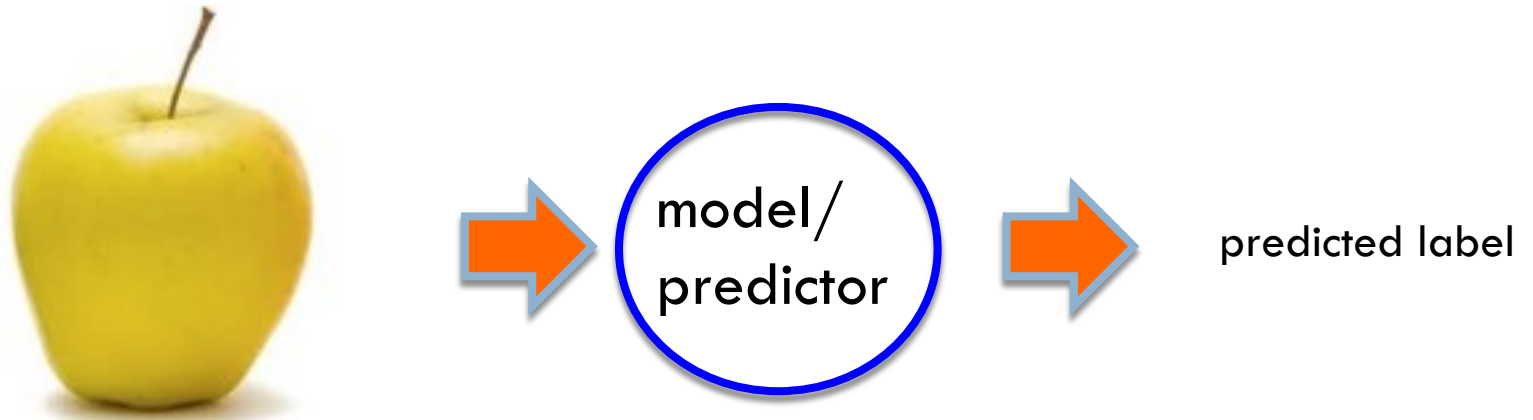
Supervised learning: given labeled examples

Supervised learning



Supervised learning: given labeled examples

Supervised learning



Supervised learning: learn to predict new example

Supervised learning: classification



label

apple



apple



banana

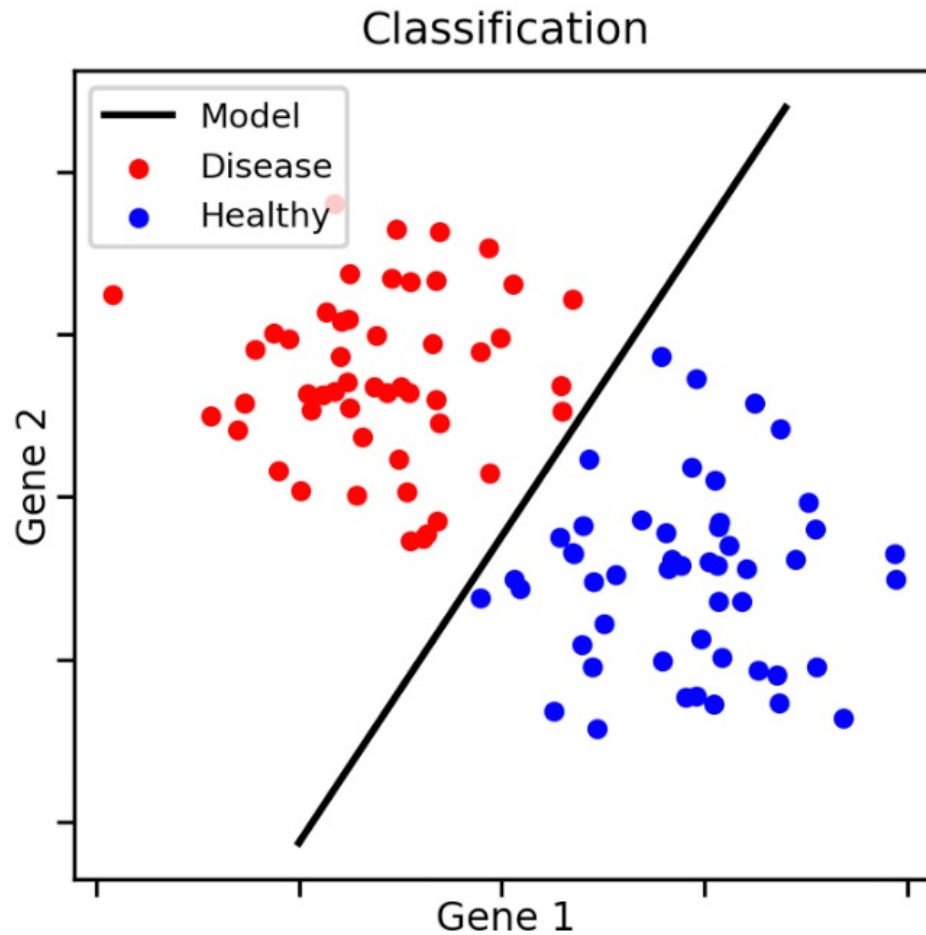


banana

Classification: a finite set of labels

Supervised learning: given labeled examples

Classification Example



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

Supervised learning: regression

| | label |
|---|-------|
|  | -4.5 |
|  | 10.1 |
|  | 3.2 |
|  | 4.3 |

Regression: label is real-valued

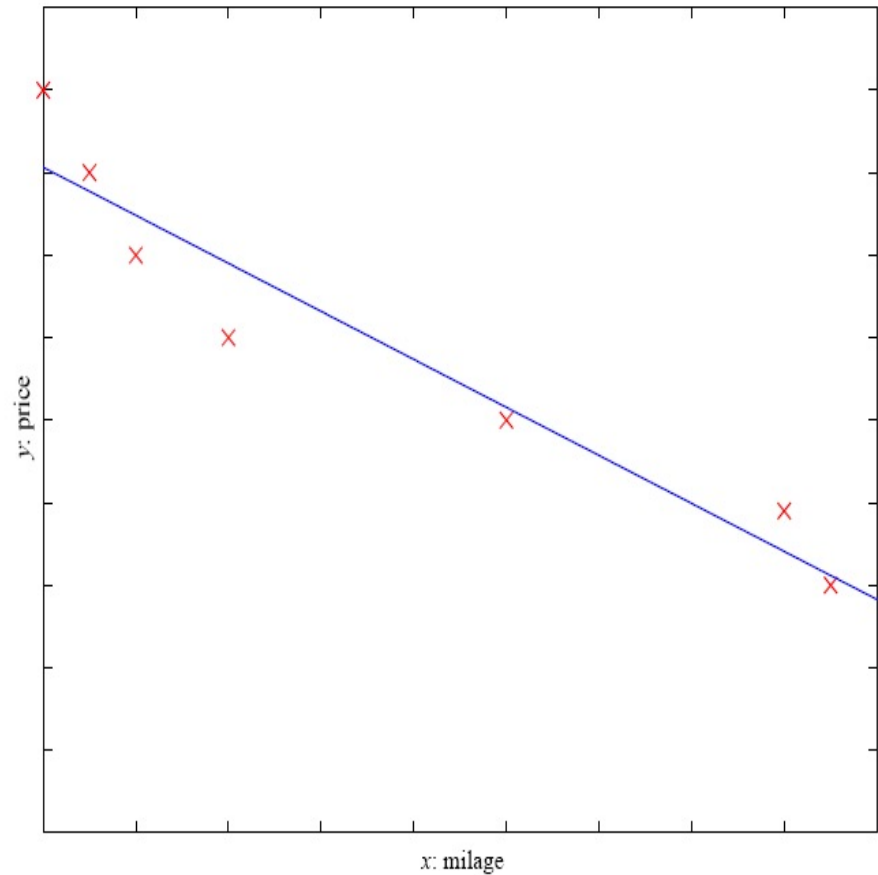
Supervised learning: given labeled examples

Regression Example

Price of a used car

x : car attributes
(e.g., mileage)

y : price



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

Supervised learning: ranking



label

1



4



2



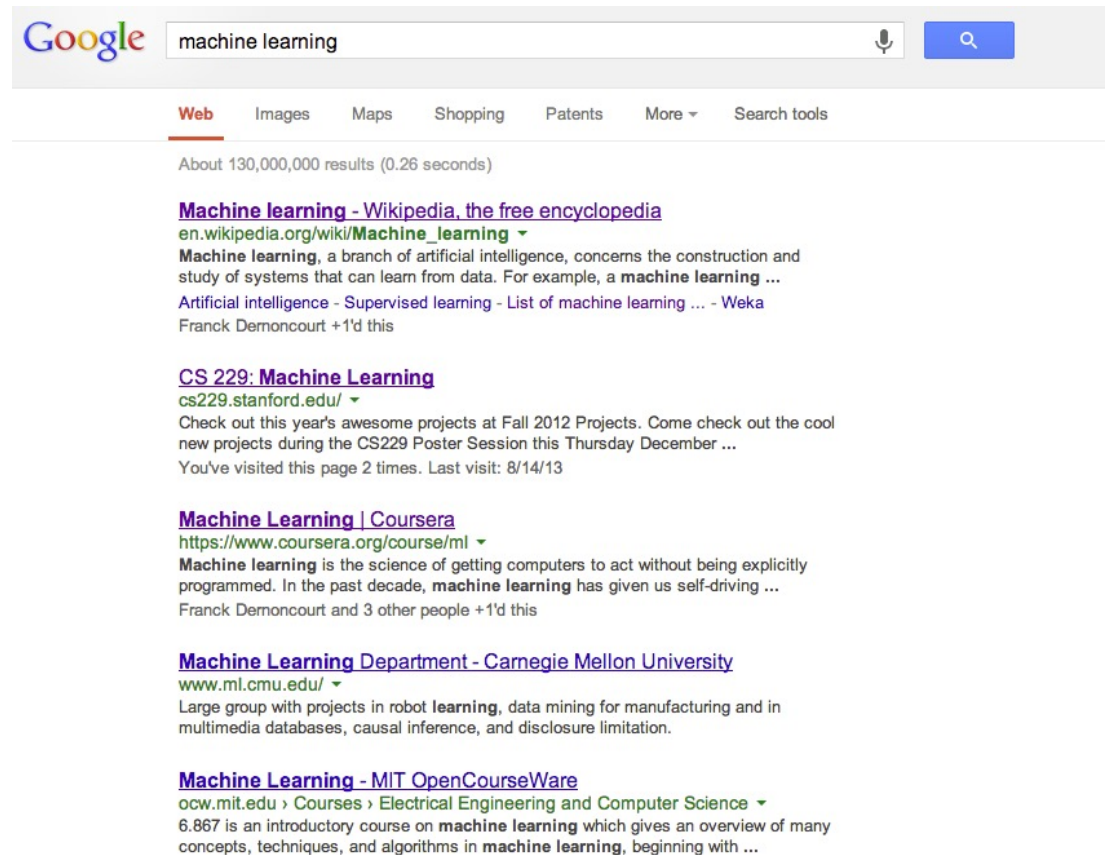
3

Ranking: label is a ranking

Supervised learning: given labeled examples

Ranking example

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



The image shows a screenshot of a Google search results page for the query "machine learning". The search bar at the top contains the text "machine learning" and a search button. Below the search bar, there are navigation tabs for "Web", "Images", "Maps", "Shopping", "Patents", "More", and "Search tools". The "Web" tab is selected. The search results show "About 130,000,000 results (0.26 seconds)". The first result is "Machine learning - Wikipedia, the free encyclopedia" with a snippet: "Machine learning, a branch of artificial intelligence, concerns the construction and study of systems that can learn from data. For example, a machine learning ...". The second result is "CS 229: Machine Learning" from Stanford University with a snippet: "Check out this year's awesome projects at Fall 2012 Projects. Come check out the cool new projects during the CS229 Poster Session this Thursday December ...". The third result is "Machine Learning | Coursera" with a snippet: "Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has given us self-driving ...". The fourth result is "Machine Learning Department - Carnegie Mellon University" with a snippet: "Large group with projects in robot learning, data mining for manufacturing and in multimedia databases, causal inference, and disclosure limitation.". The fifth result is "Machine Learning - MIT OpenCourseWare" with a snippet: "6.867 is an introductory course on machine learning which gives an overview of many concepts, techniques, and algorithms in machine learning, beginning with ...".

Google machine learning

Web Images Maps Shopping Patents More Search tools

About 130,000,000 results (0.26 seconds)

[Machine learning - Wikipedia, the free encyclopedia](#)
en.wikipedia.org/wiki/Machine_learning
Machine learning, a branch of artificial intelligence, concerns the construction and study of systems that can learn from data. For example, a machine learning ...
Artificial intelligence - Supervised learning - List of machine learning ... - Weka
Franck Demoncourt +1'd this

[CS 229: Machine Learning](#)
cs229.stanford.edu/
Check out this year's awesome projects at Fall 2012 Projects. Come check out the cool new projects during the CS229 Poster Session this Thursday December ...
You've visited this page 2 times. Last visit: 8/14/13

[Machine Learning | Coursera](#)
https://www.coursera.org/course/ml
Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has given us self-driving ...
Franck Demoncourt and 3 other people +1'd this

[Machine Learning Department - Carnegie Mellon University](#)
www.ml.cmu.edu/
Large group with projects in robot learning, data mining for manufacturing and in multimedia databases, causal inference, and disclosure limitation.

[Machine Learning - MIT OpenCourseWare](#)
ocw.mit.edu › Courses › Electrical Engineering and Computer Science
6.867 is an introductory course on machine learning which gives an overview of many concepts, techniques, and algorithms in machine learning, beginning with ...

Ranking Applications

User preference, e.g., Netflix “My List” -- movie queue ranking

Spotify

flight search (search in general)

reranking N-best output lists

Unsupervised learning



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

Unsupervised learning applications

learn clusters/groups without any label

customer segmentation (i.e. grouping)

image compression

bioinformatics: learn motifs

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

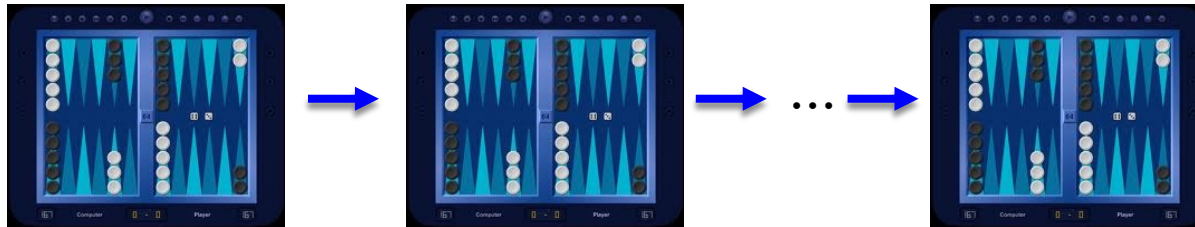
left, straight, straight, left, right, straight, straight

-3

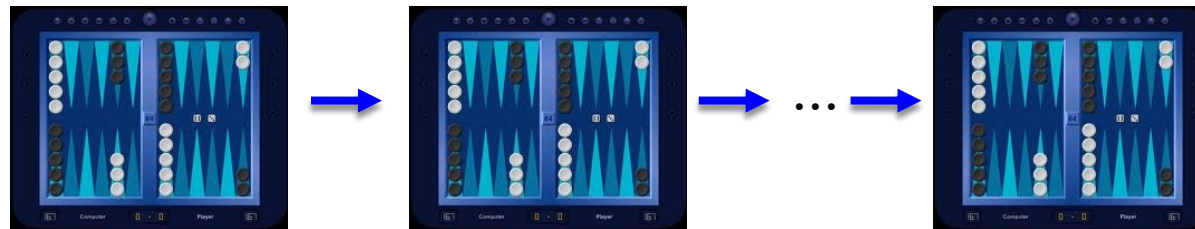
Given a *sequence* of examples/states and a *reward* after completing that sequence, learn to predict the action to take in for an individual example/state

Reinforcement learning example

Backgammon



WIN!



LOSE!

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

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

Representing examples

examples



What is an example?
How is it represented?

Features

examples



features

$f_1, f_2, f_3, \dots, f_n$

$f_1, f_2, f_3, \dots, f_n$

$f_1, f_2, f_3, \dots, f_n$

$f_1, f_2, f_3, \dots, f_n$

How our algorithms actually “view” the data

Features are the questions we can ask about the examples

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

Classification revisited

examples

red, round, leaf, 3oz, ...

green, round, no leaf, 4oz, ...

yellow, curved, no leaf, 8oz, ...

green, curved, no leaf, 7oz, ...

label

apple

apple

banana

banana

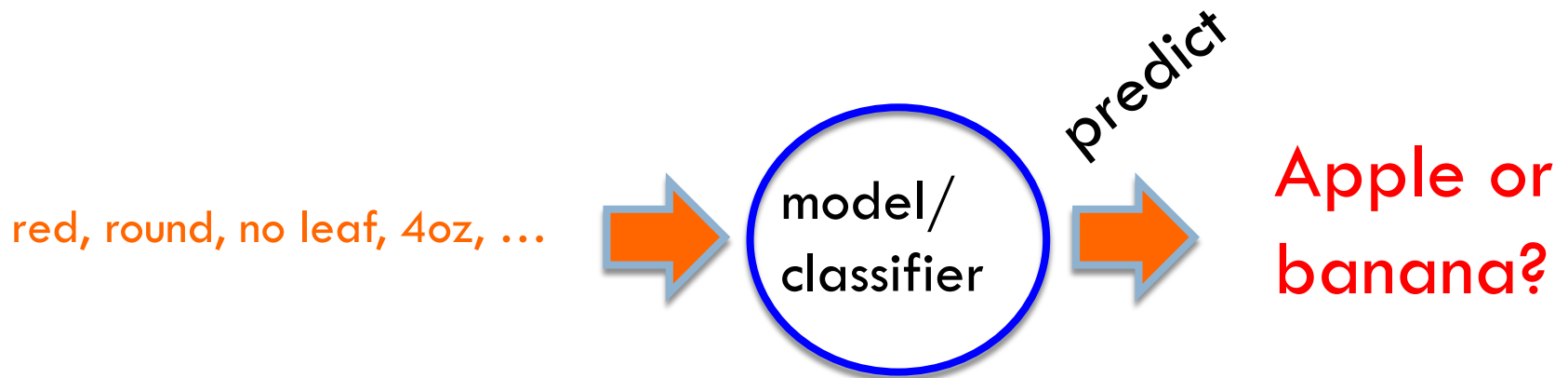
learn



model/
classifier

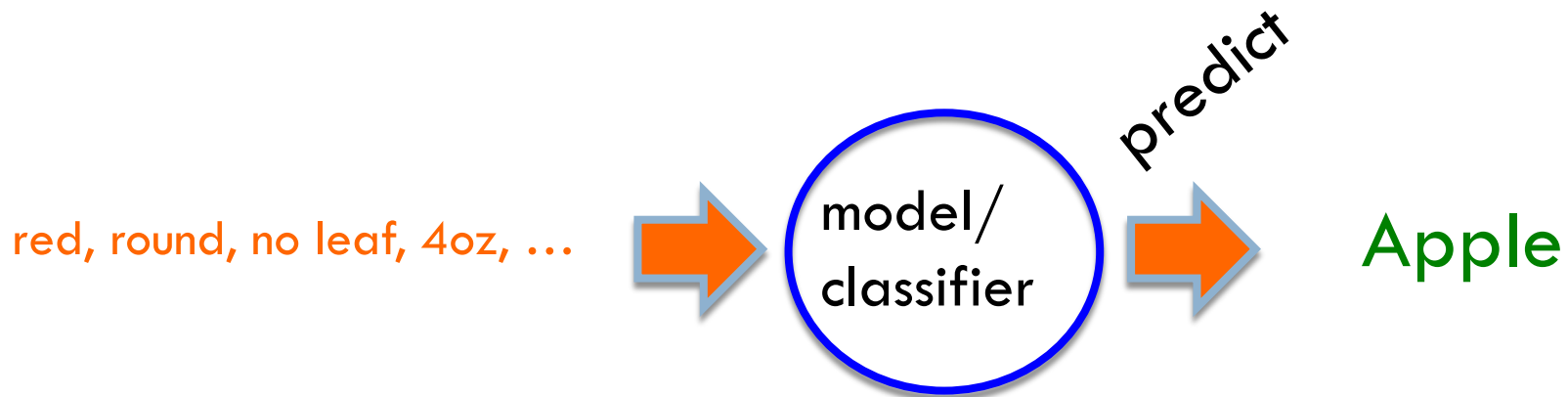
During learning/training/induction, learn a model of what distinguishes apples and bananas *based on the features*

Classification revisited



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

Classification revisited



Why?

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

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

Classification revisited

Training data

examples

red, round, leaf, 3oz, ...

green, round, no leaf, 4oz, ...

yellow, curved, no leaf, 4oz, ...

green, curved, no leaf, 5oz, ...

label

apple

apple

banana

banana

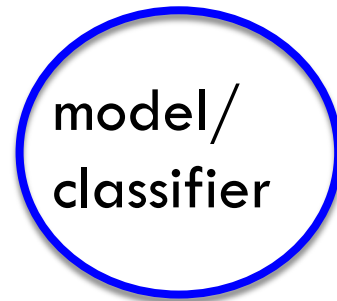
Test set

red, round, no leaf, 4oz, ...

?

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

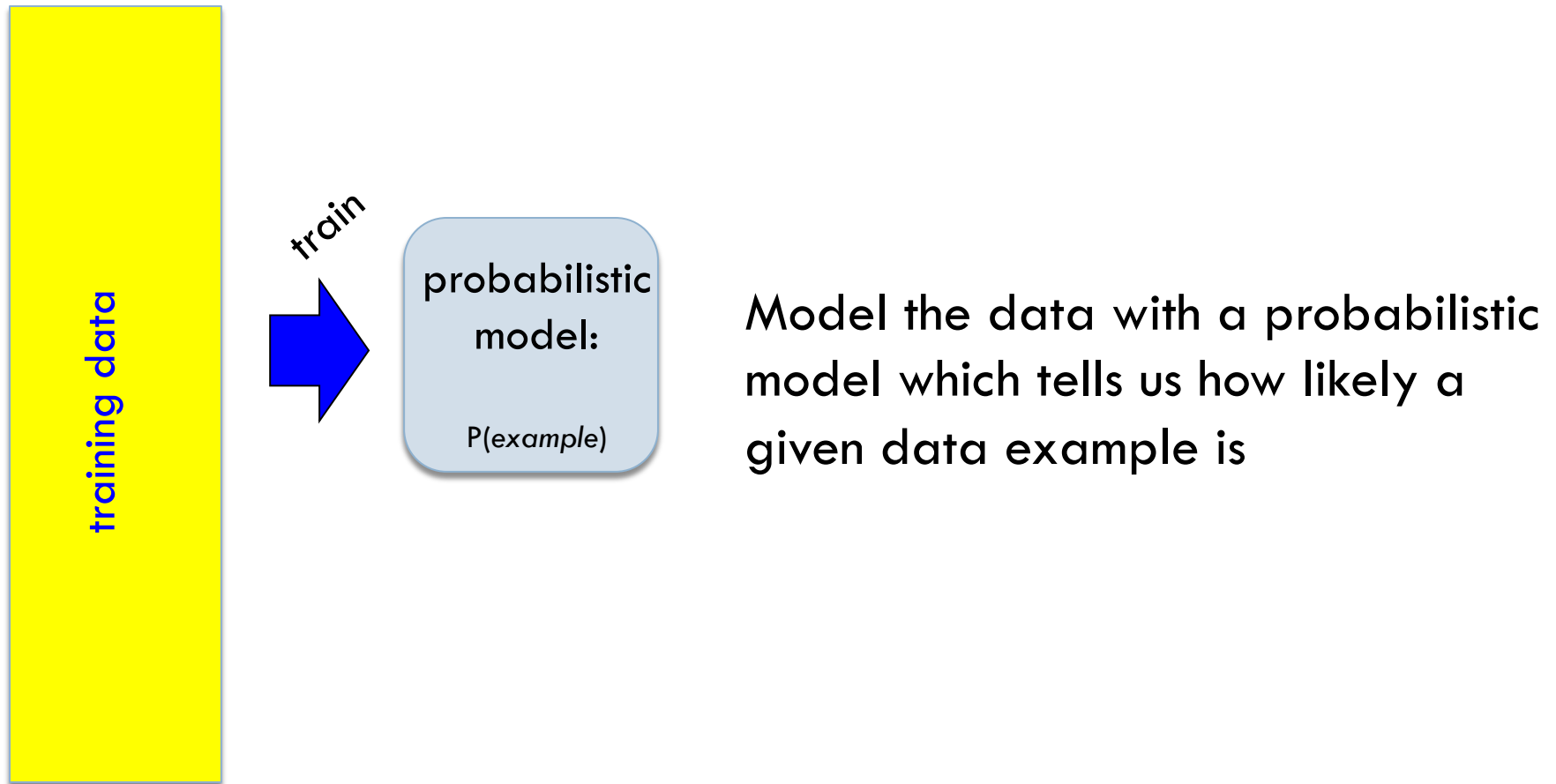
models



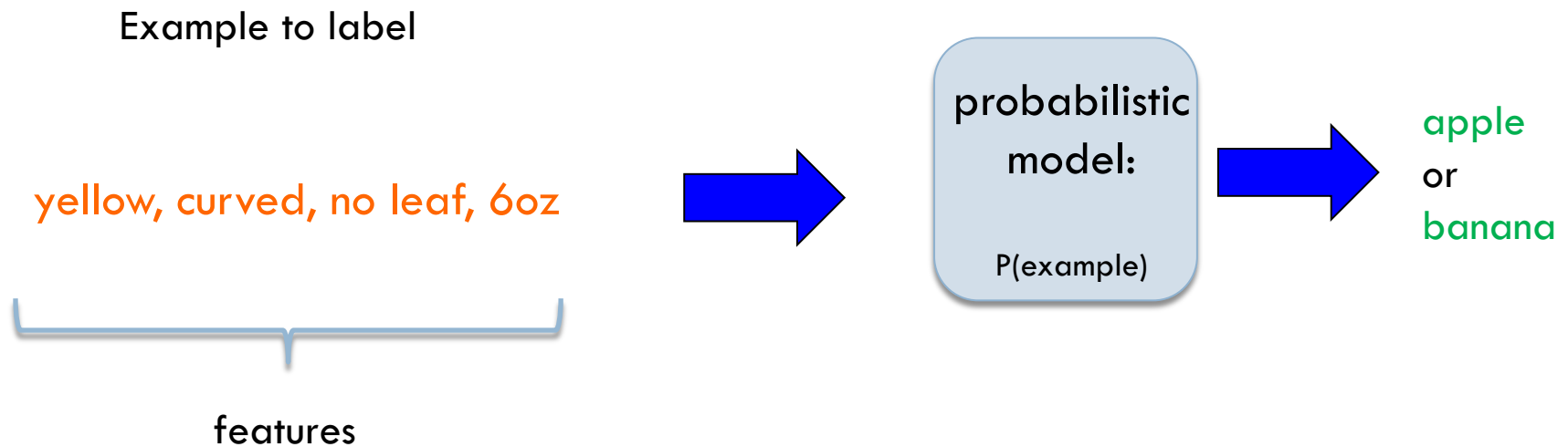
We have many, many different options for the model

They have different characteristics and perform differently (accuracy, speed, etc.)

Probabilistic modeling



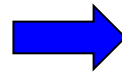
Probabilistic models



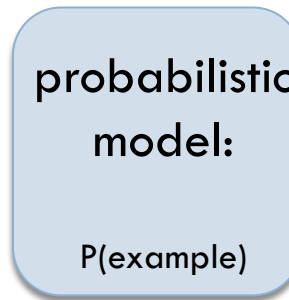
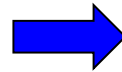
Probabilistic models

For each label, ask for the probability

yellow, curved, no leaf, 6oz, banana



yellow, curved, no leaf, 6oz, apple



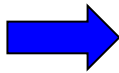
features

label

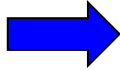
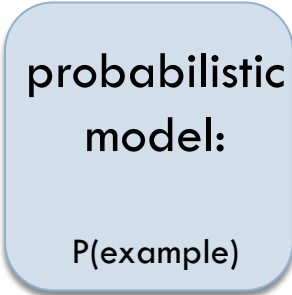
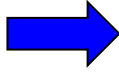
Probabilistic models

Pick the label with the highest probability

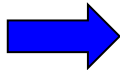
yellow, curved, no leaf, 6oz, banana



yellow, curved, no leaf, 6oz, apple



0.004



0.00002



features

label

Probability basics

A **probability distribution** gives the probabilities of all possible values of an event

For example, say we flip a coin three times. We can define the probability of the number of time the coin came up heads.

| P(num heads) |
|---------------------|
| P(3) = ? |
| P(2) = ? |
| P(1) = ? |
| P(0) = ? |

Probability distributions

What are the possible outcomes of three flips (hint, there are eight of them)?

T T T
T T H
T H T
T H H
H T T
H T H
H H T
H H H

Probability distributions

Assuming the coin is fair, what are our probabilities?

$$\text{probability} = \frac{\text{number of times it happens}}{\text{total number of cases}}$$

T T T
T T H
T H T
T H H
H T T
H T H
H H T
H H H

| P(num heads) |
|--------------|
| P(3) = ? |
| P(2) = ? |
| P(1) = ? |
| P(0) = ? |

Probability distributions

Assuming the coin is fair, what are our probabilities?

$$\text{probability} = \frac{\text{number of times it happens}}{\text{total number of cases}}$$

T T T
T T H
T H T
T H H
H T T
H T H
H H T
H H H

| P(num heads) |
|--------------|
| P(3) = ? |
| P(2) = ? |
| P(1) = ? |
| P(0) = ? |

Probability distributions

Assuming the coin is fair, what are our probabilities?

$$\text{probability} = \frac{\text{number of times it happens}}{\text{total number of cases}}$$

T T T
T T H
T H T
T H H
H T T
H T H
H H T
H H H

| P(num heads) |
|--------------|
| P(3) = 1/8 |
| P(2) = ? |
| P(1) = ? |
| P(0) = ? |

Probability distributions

Assuming the coin is fair, what are our probabilities?

$$\text{probability} = \frac{\text{number of times it happens}}{\text{total number of cases}}$$

T T T
T T H
T H T
T H H
H T T
H T H
H H T
H H H

| P(num heads) |
|--------------|
| P(3) = 1/8 |
| P(2) = ? |
| P(1) = ? |
| P(0) = ? |

Probability distributions

Assuming the coin is fair, what are our probabilities?

$$\text{probability} = \frac{\text{number of times it happens}}{\text{total number of cases}}$$

T T T
T T H
T H T
T H H
H T T
H T H
H H T
H H H

| P(num heads) |
|--------------|
| P(3) = 1/8 |
| P(2) = 3/8 |
| P(1) = ? |
| P(0) = ? |

Probability distributions

Assuming the coin is fair, what are our probabilities?

$$\text{probability} = \frac{\text{number of times it happens}}{\text{total number of cases}}$$

T T T
T T H
T H T
T H H
H T T
H T H
H H T
H H H

| P(num heads) |
|--------------|
| P(3) = 1/8 |
| P(2) = 3/8 |
| P(1) = 3/8 |
| P(0) = 1/8 |

Probability distributions

A probability distribution assigns probability values to *all possible values*

Probabilities are between 0 and 1, inclusive

The sum of all probabilities in a distribution must be 1

| P(num heads) |
|---------------------|
| $P(3) = 1/8$ |
| $P(2) = 3/8$ |
| $P(1) = 3/8$ |
| $P(0) = 1/8$ |

Probability distributions

A probability distribution assigns probability values to *all possible values*

Probabilities are between 0 and 1, inclusive

The sum of all probabilities in a distribution must be 1

| P |
|--------------|
| $P(3) = 1/2$ |
| $P(2) = 1/2$ |
| $P(1) = 1/2$ |
| $P(0) = 1/2$ |

| P |
|-------------|
| $P(3) = -1$ |
| $P(2) = 2$ |
| $P(1) = 0$ |
| $P(0) = 0$ |

Some example probability distributions

probability of heads

(distribution options: heads, tails)

probability of passing class

(distribution options: pass, fail)

probability of rain today

(distribution options: rain or no rain)

probability of getting an 'A'

(distribution options: A, B, C, D, F)

Conditional probability distributions

Sometimes we may know extra information about the world that may change our probability distribution

$P(X | Y)$ captures this (read “probability of X given Y ”)

- ▣ Given some information (Y) what does our probability distribution look like
- ▣ Note that this is still just a typical probability distribution

Conditional probability example

P(pass 51a)

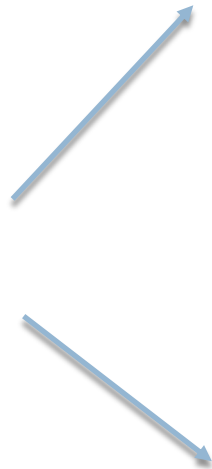
$P(\text{pass}) = 0.9$

$P(\text{not pass}) = 0.1$

Unconditional probability distribution

Conditional probability example

| |
|----------------------------|
| P(pass 51a) |
| $P(\text{pass}) = 0.9$ |
| $P(\text{not pass}) = 0.1$ |



| |
|----------------------------------|
| P(pass 51a don't study) |
| $P(\text{pass}) = 0.5$ |
| $P(\text{not pass}) = 0.5$ |

| |
|-------------------------------|
| P(pass 51a do study) |
| $P(\text{pass}) = 0.95$ |
| $P(\text{not pass}) = 0.05$ |

Still probability distributions over passing 51A

Conditional probability distributions

Conditional probability example

P(rain in LA)

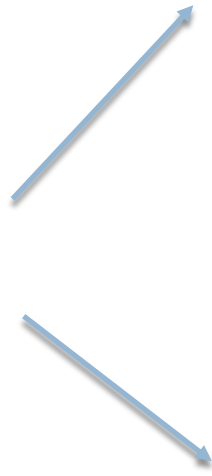
$P(\text{rain}) = 0.05$

$P(\text{no rain}) = 0.95$

Unconditional probability distribution

Conditional probability example

| |
|----------------------------|
| P(rain in LA) |
| $P(\text{rain}) = 0.05$ |
| $P(\text{no rain}) = 0.95$ |



| |
|---------------------------------|
| P(rain in LA January) |
| $P(\text{rain}) = 0.2$ |
| $P(\text{no rain}) = 0.8$ |

| |
|-------------------------------------|
| P(rain in LA not January) |
| $P(\text{rain}) = 0.03$ |
| $P(\text{no rain}) = 0.97$ |

Still probability distributions over raining in LA

Conditional probability distributions

Joint distribution

Probability over two events: $P(X,Y)$

Has probabilities for all possible combinations over the two events

| 51Pass, EngPass | P(51Pass, EngPass) |
|------------------------|---------------------------|
| true, true | .88 |
| true, false | .01 |
| false, true | .04 |
| false, false | .07 |

Joint distribution

Still a probability distribution

All questions/probabilities that we might want to ask about these two things can be calculated from the joint distribution

| 51Pass, EngPass | P(51Pass, EngPass) |
|-----------------|--------------------|
| true, true | .88 |
| true, false | .01 |
| false, true | .04 |
| false, false | .07 |

What is $P(51 \text{ pass} = \text{true})$?

Joint distribution

| 51Pass, EngPass | P(51Pass, EngPass) |
|-----------------|--------------------|
| true, true | .88 |
| true, false | .01 |
| false, true | .04 |
| false, false | .07 |

There are two ways that a person can pass 51:
they can do it while passing or not passing English

$$P(51\text{Pass}=\text{true}) = P(\text{true, true}) + P(\text{true, false}) = 0.89$$

Relationship between distributions

$$P(X, Y) = P(Y) * P(X|Y)$$

joint distribution

unconditional distribution

conditional distribution

The diagram illustrates the relationship between three types of probability distributions. The equation $P(X, Y) = P(Y) * P(X|Y)$ is centered. Three blue arrows point from labels below to terms in the equation: one from 'joint distribution' to $P(X, Y)$, one from 'unconditional distribution' to $P(Y)$, and one from 'conditional distribution' to $P(X|Y)$.

Can think of it as describing the two events happening in two steps:

The likelihood of X and Y happening:

1. How likely it is that Y happened?
2. Given that Y happened, how likely is it that X happened?

Relationship between distributions

$$P(51Pass, EngPass) = P(EngPass) * P(51Pass|EngPass)$$

The probability of passing CS51 and English is:

1. Probability of passing English *
2. Probability of passing CS51 **given** that you passed English

Relationship between distributions

$$P(51Pass, EngPass) = P(51Pass) * P(EngPass|51Pass)$$

The probability of passing CS51 and English is:

1. Probability of passing **CS51** *
2. Probability of passing **English** **given** that you passed **CS51**

Can also view it with the other event happening first