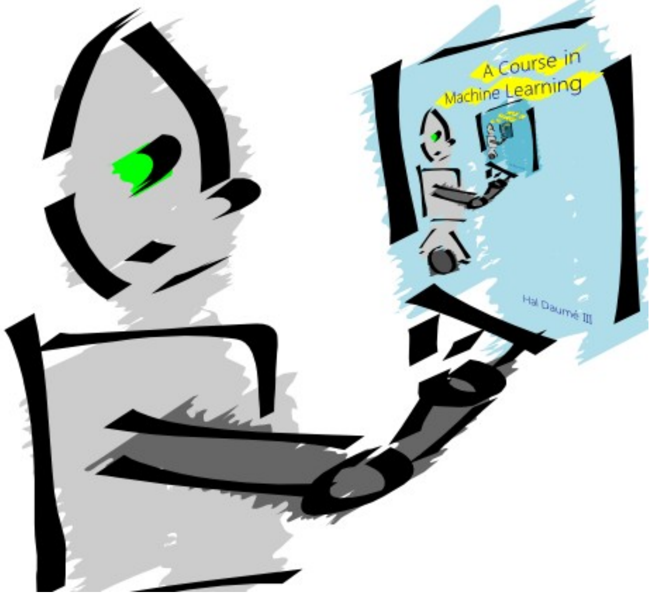
INTRODUCTION TO MACHINE LEARNING

David Kauchak, Joseph C. Osborn
CS 51A – Fall 2019

Machine Learning is…

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

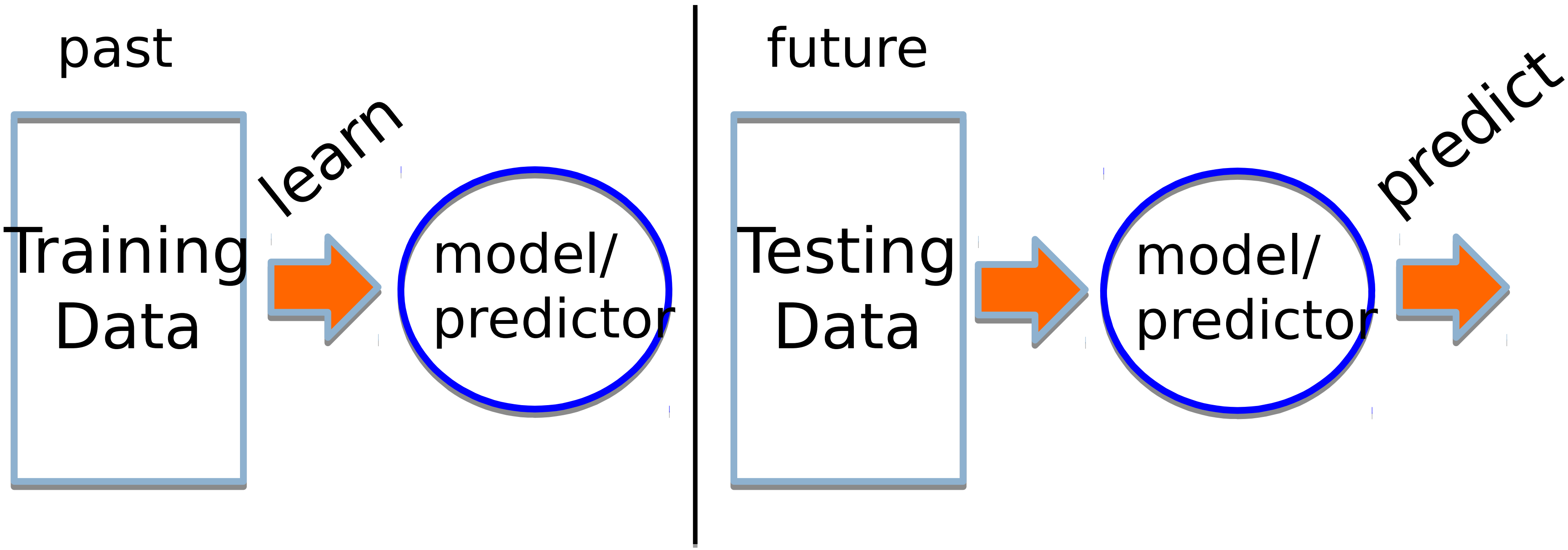
-- Hal Daume III



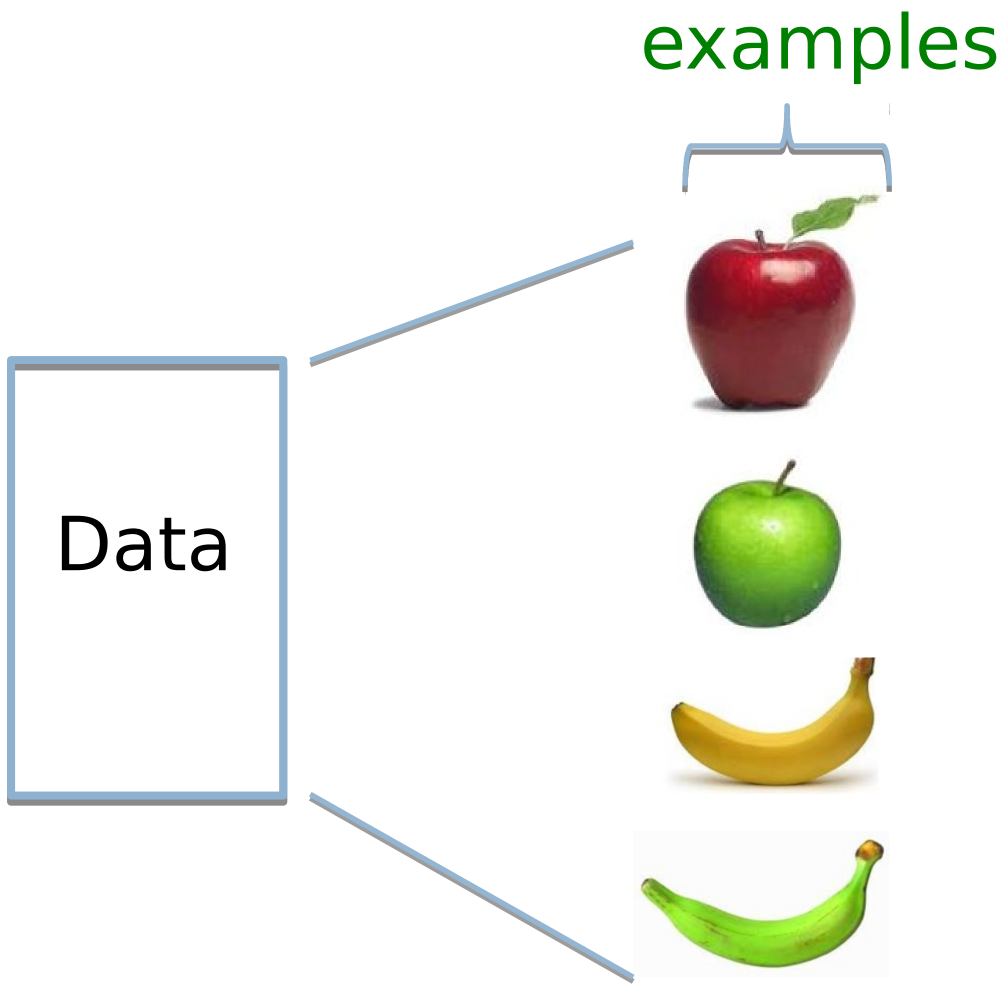
Machine Learning is…

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

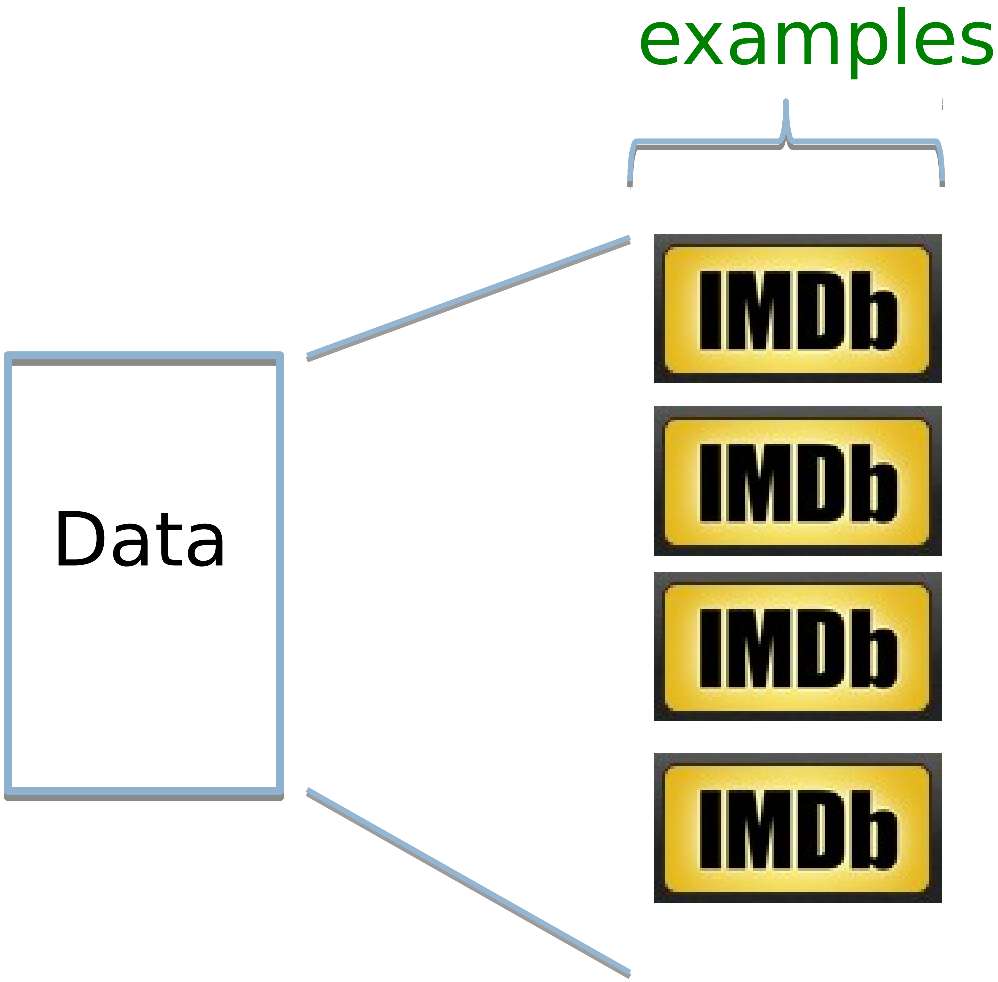
-- Hal Daume III



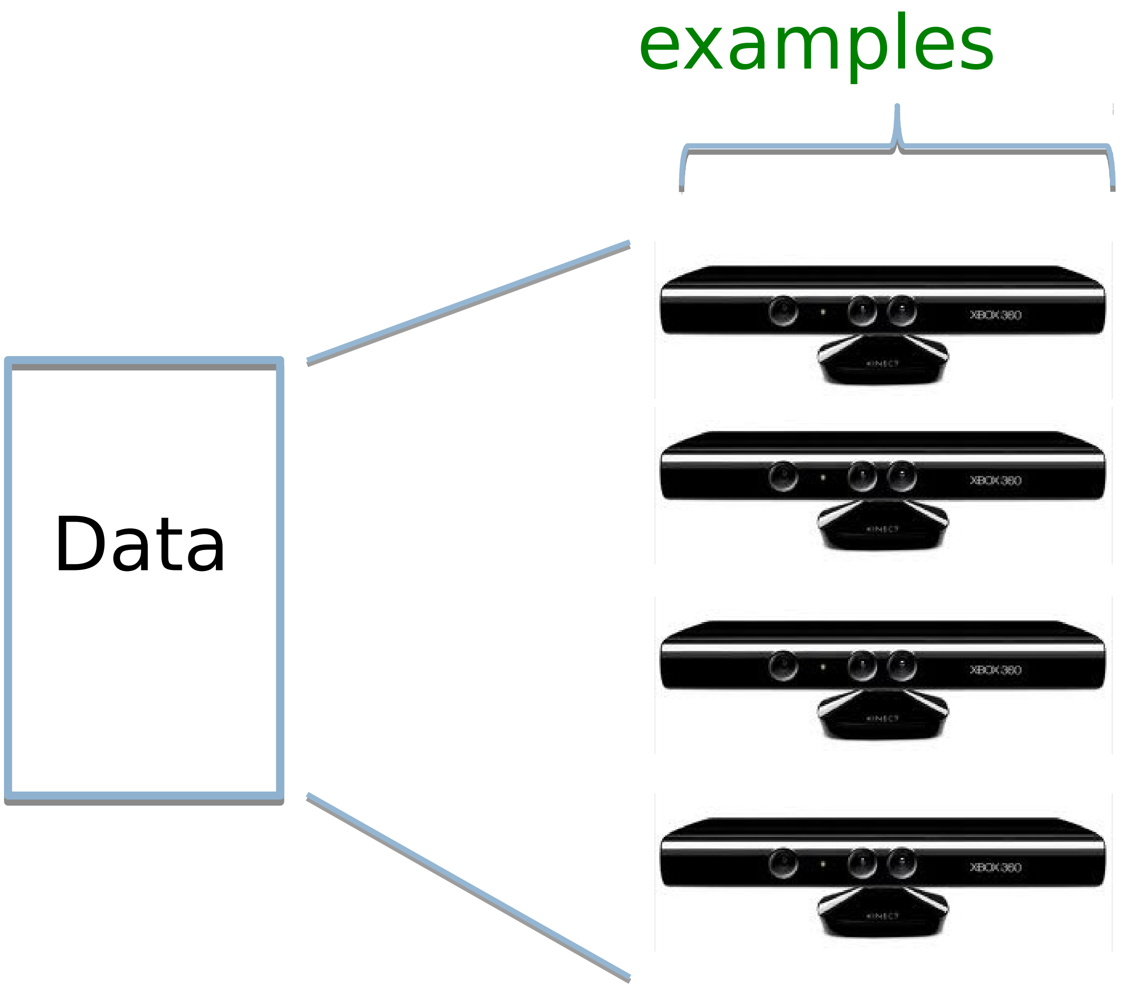
Data



Data



Data



Data

examples

Data

Supervised learning

examples

label

label$_1$

label$_3$

label$_4$

label$_5$

labeled examples

Supervised learning: given labeled examples

Supervised learning

label

label_1

label_3

label_4

label_5

model/
predictor

Supervised learning: given labeled examples

Supervised learning

model/
predictor

predicted label

Supervised learning: learn to predict new example

Supervised learning: classification

label

apple

apple

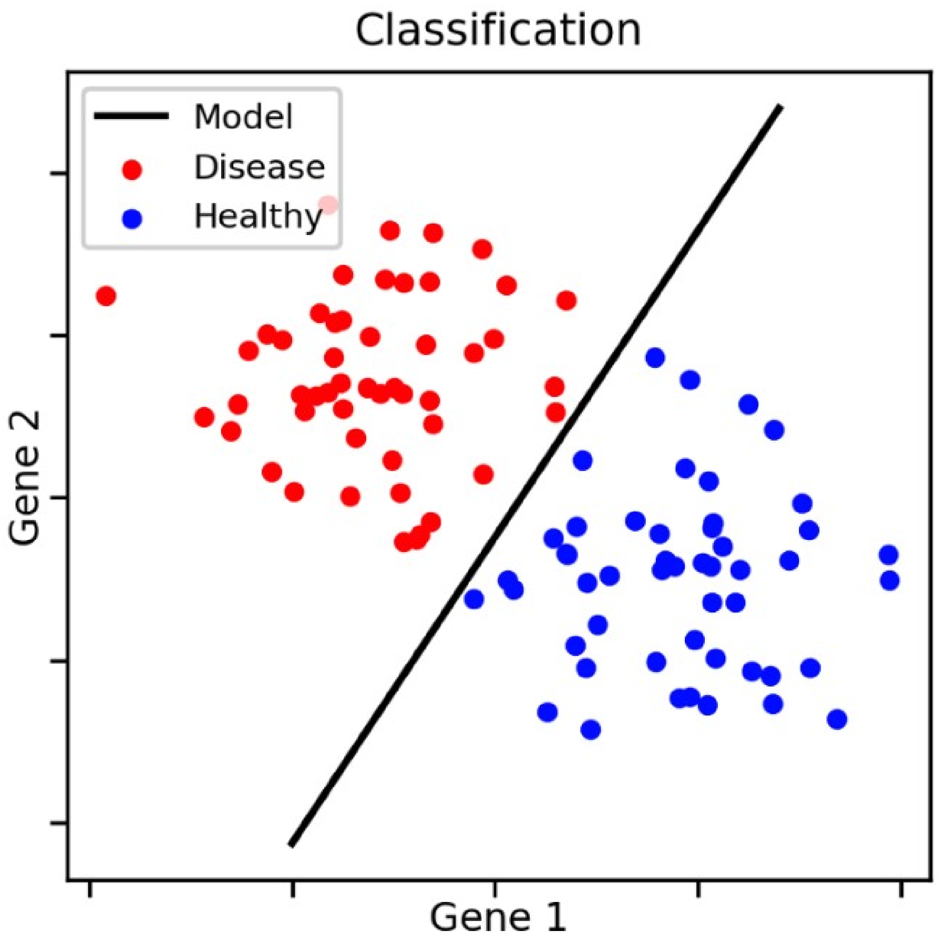
Classification: a finite set of labels

banana

banana

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

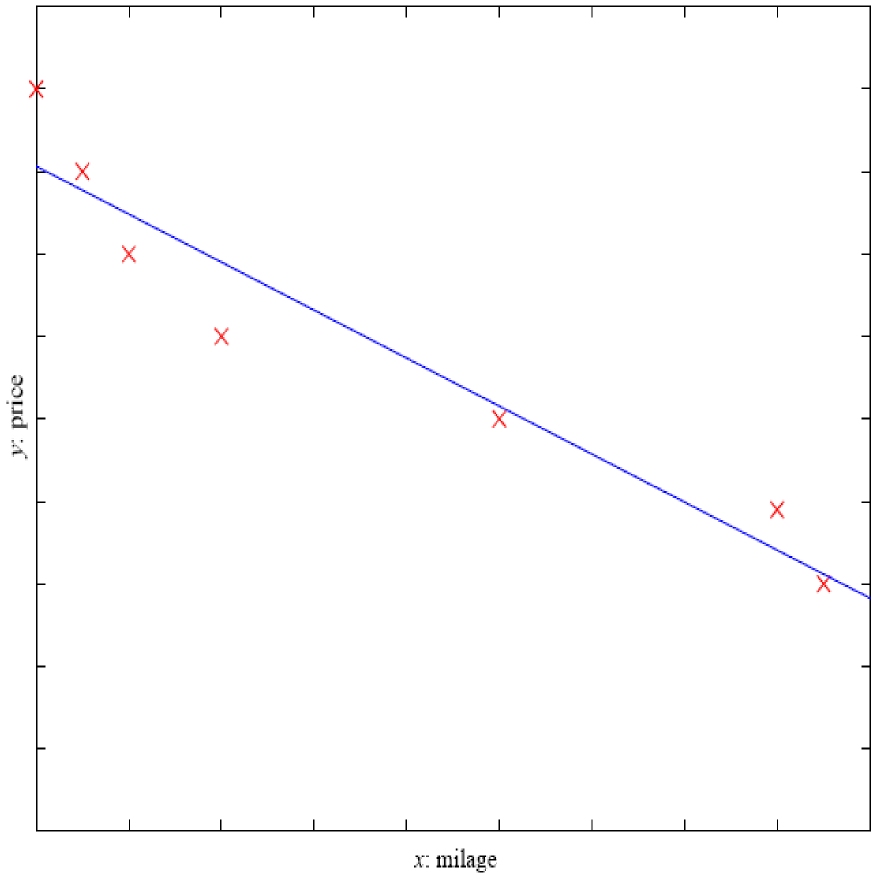
Supervised learning: given labeled examples

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

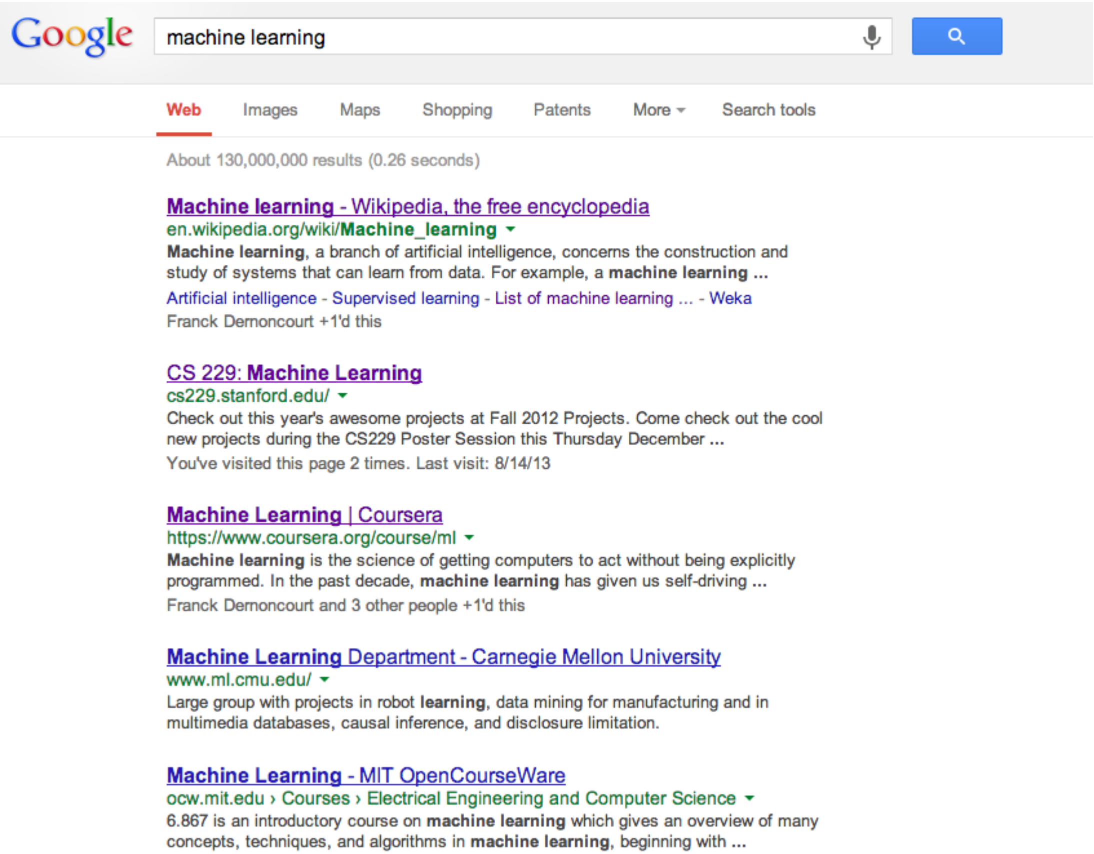
Ranking example

Given a query and

a set of web pages,

rank them according

to relevance



Ranking Applications

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

iTunes

flight search (search in general)

reranking N-best output lists

Unsupervised learning





Unupervised 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

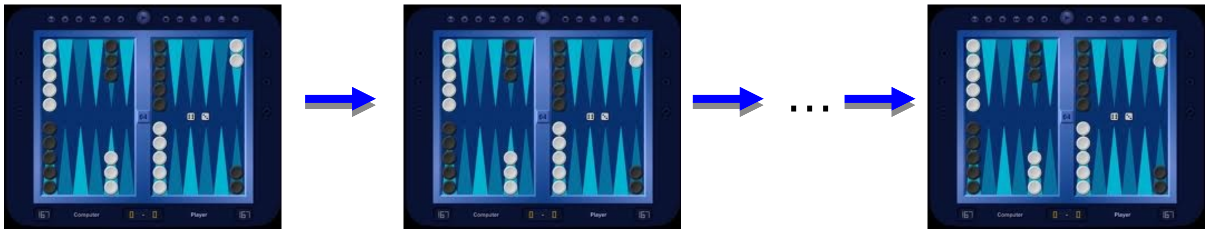
left, right, straight, left, left, left, straight	**GOOD**
left, straight, straight, left, right, straight, straight	**BAD**

left, right, straight, left, left, left, straight	**18.5**
left, straight, straight, left, right, straight, straight	**-3**

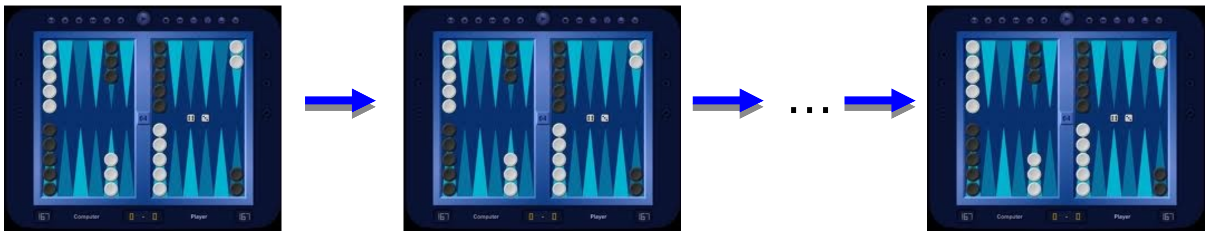
Given a *sequence* of examples/states and a *reward* after completing that sequence, learn to predict the action to take in for an individual example/state

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

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

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

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

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

How our algorithms actually “view” the data

Features are the questions we can ask about the examples

Features

examples

features

red, round, leaf, 3oz, …

green, round, no leaf, 4oz, …

yellow, curved, no leaf, 8oz, …

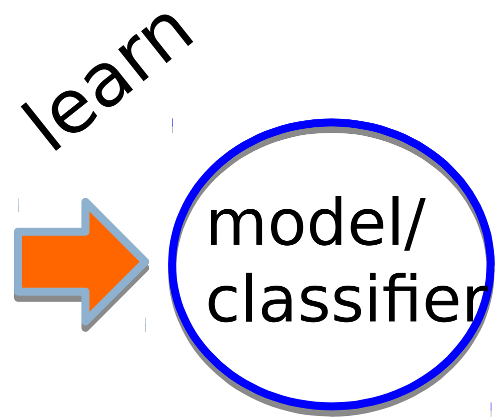
green, curved, no leaf, 7oz, …

How our algorithms actually “view” the data

Features are the questions we can ask about the examples

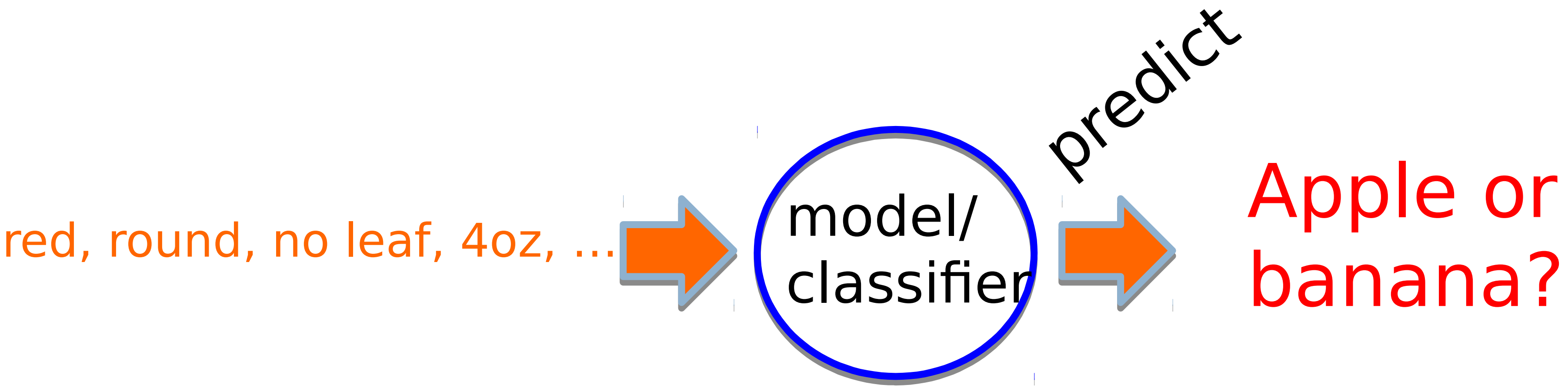
Classification revisited

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



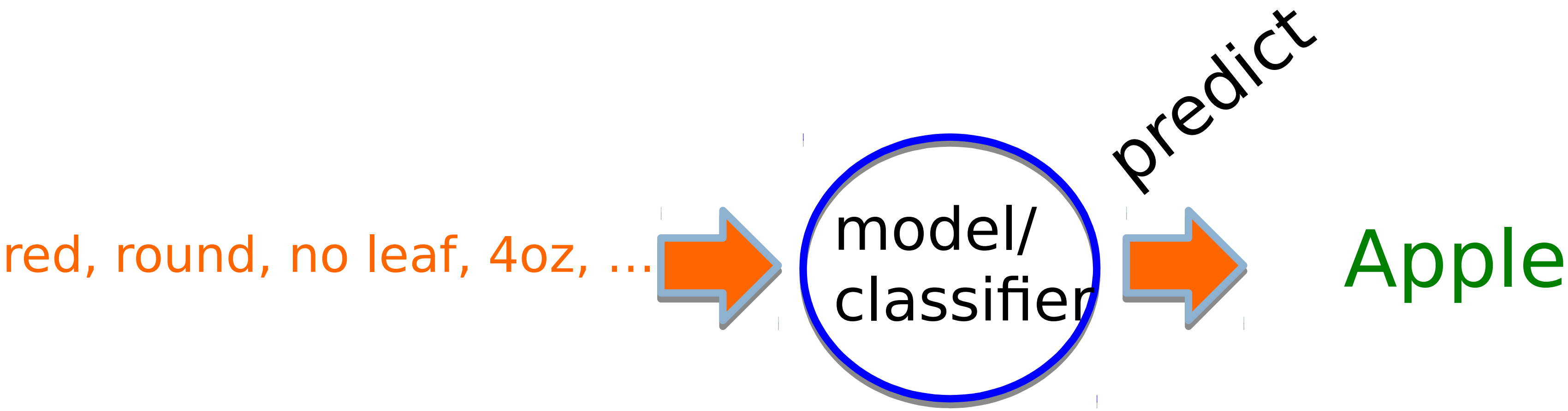
During learning/training/induction, learn a model of what distinguishes apples and bananas *based on the features*

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

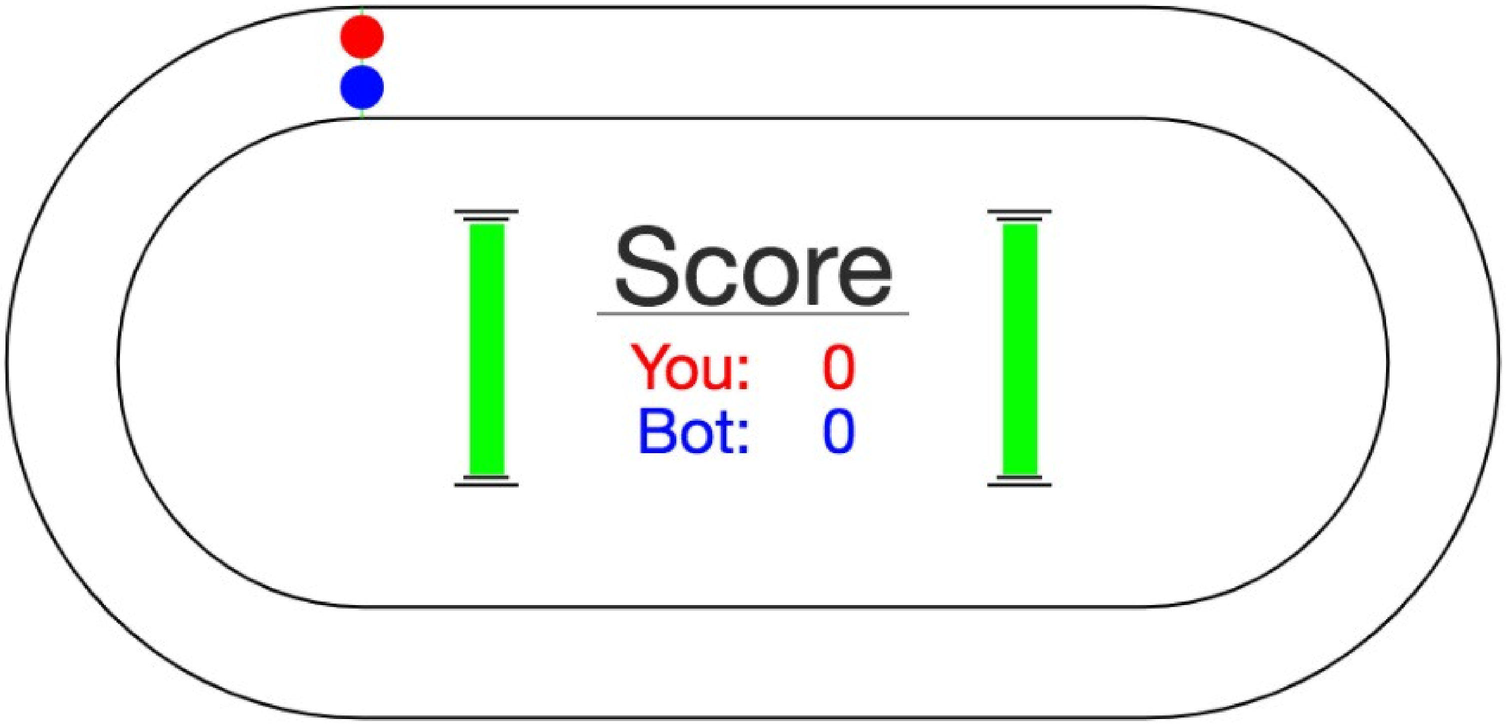
Test set

red, round, no leaf, 4oz, … ?

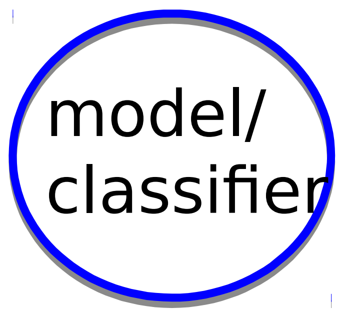
Learning is about ***generalizing*** from the training data

A simple machine learning example

http://www.mindreaderpro.appspot.com/



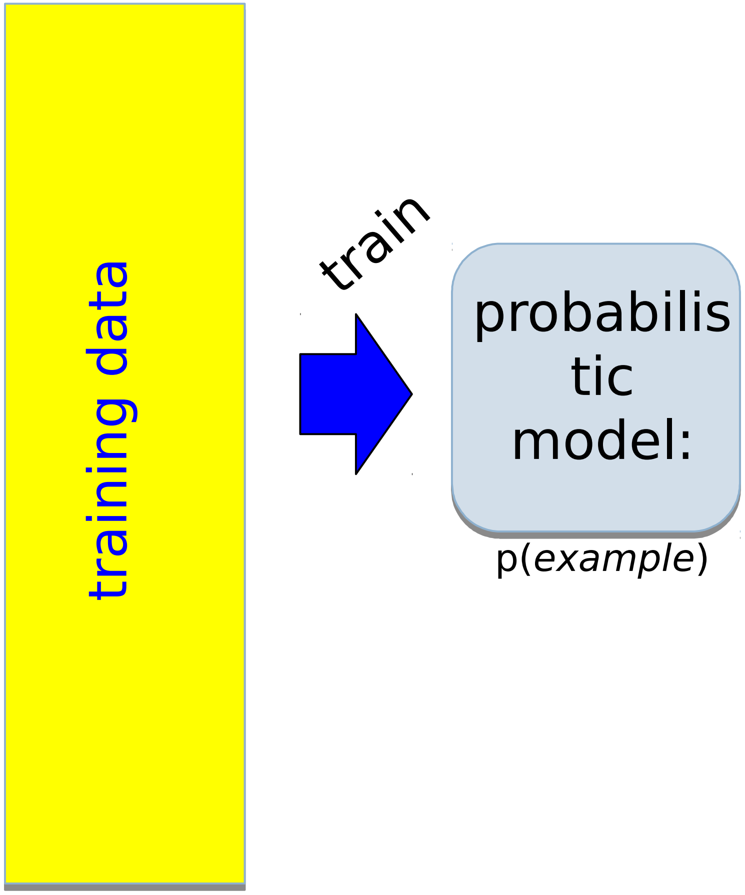
models



We have many, many different options for the model

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

Probabilistic modeling



Model the data with a probabilistic model which tells us how likely a given data example is

Probabilistic models

Example to label

yellow, curved, no leaf, 6oz

features

probabilistic model:

p(example)

apple
or
banana

Probabilistic models

For each label, ask for the probability

yellow, curved, no leaf, 6oz, banana

yellow, curved, no leaf, 6oz, apple

features

label

probabilistic model:

p(example)

Probabilistic models

Pick the label with the highest probability

yellow, curved, no leaf, 6oz, banana → **0.004**

yellow, curved, no leaf, 6oz, apple → 0.00002

probabilistic model:

p(example)

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 distribution

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 distribution

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 normal probability distribution

Conditional probability example

P(pass 51a)
P(pass) = 0.9
P(not pass) = 0.1

Unconditional probability distribution

Conditional probability example

P(pass 51a)
P(pass) = 0.9
P(not pass) = 0.1

P(pass 51a | don't study)
P(pass) = 0.5
P(not pass) = 0.5

P(pass 51a | do study)
P(pass) = 0.95
P(not pass) = 0.05

Still probability distributions over passing 51A

Conditional probability distributions

Conditional probability example

P(rain in LA)
P(rain) = 0.05
P(no rain) = 0.95

Unconditional probability distribution

Conditional probability example

P(rain in LA)
P(rain) = 0.05
P(no rain) = 0.95

P(rain in LA| January)
P(rain) = 0.2
P(no rain) = 0.8

P(rain in LA| not January)
P(pass) = 0.03
P(not pass) = 0.97

Still probability distributions over passing rain 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 joing distribution

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

What is P(51pass = 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(51Pass=true) = P(true, true) + P(true, false) = 0.89

Relationship between distributions

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

joint distribution

unconditional distribution

conditional distribution

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