CSE419 – Artificial Intelligence and Machine Learning 2018

PhD Furkan Gözükara, Toros University

https://github.com/FurkanGozukara/CSE419_2018

Lecture 2

Introduction to Machine Learning

Based on Asst. Prof. Dr. David Kauchak (Pomona College) Lecture Slides

Why are you here?

What is Machine Learning?

Why are you taking this course?

What topics would you like to see covered?

Machine Learning is…

Machine learning, a branch of artificial intelligence, concerns the construction and study of systems that can learn from data.

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

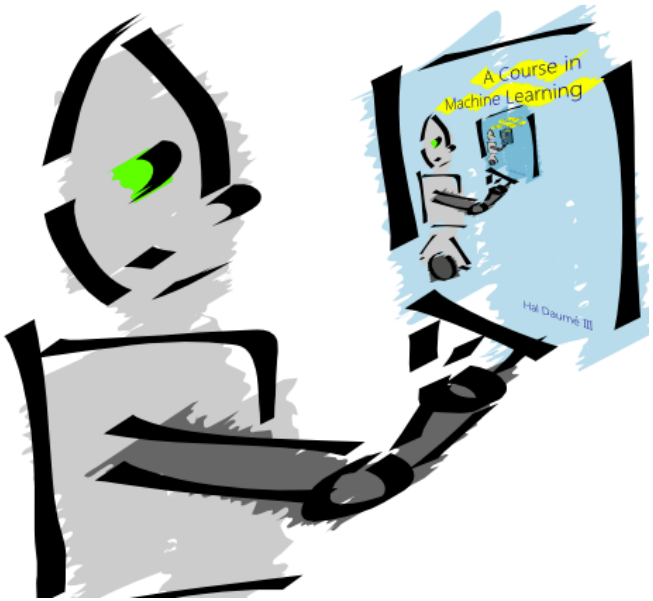
The field of pattern recognition is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions.

-- Christopher M. Bishop

Machine Learning is…

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

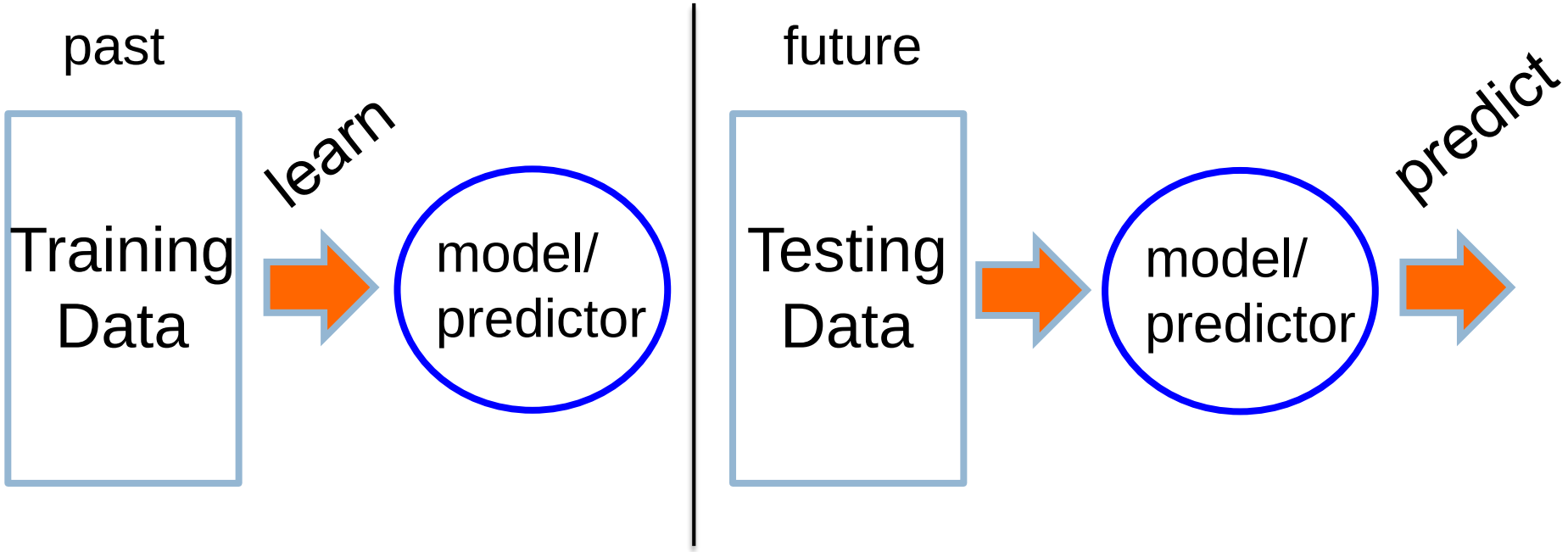
-- Hal Daume III



Machine Learning is…

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

-- Hal Daume III



Machine Learning, aka

data mining: machine learning applied to "databases", i.e. collections of data

inference and/or *estimation* in statistics

pattern recognition in engineering

signal processing in electrical engineering

induction

optimization

Goals of the course: Learn about…

Different machine learning problems

Common techniques/tools used

- theoretical understanding
- practical implementation

Proper experimentation and evaluation

Dealing with large (huge) data sets

- Parallelization frameworks
- Programming tools

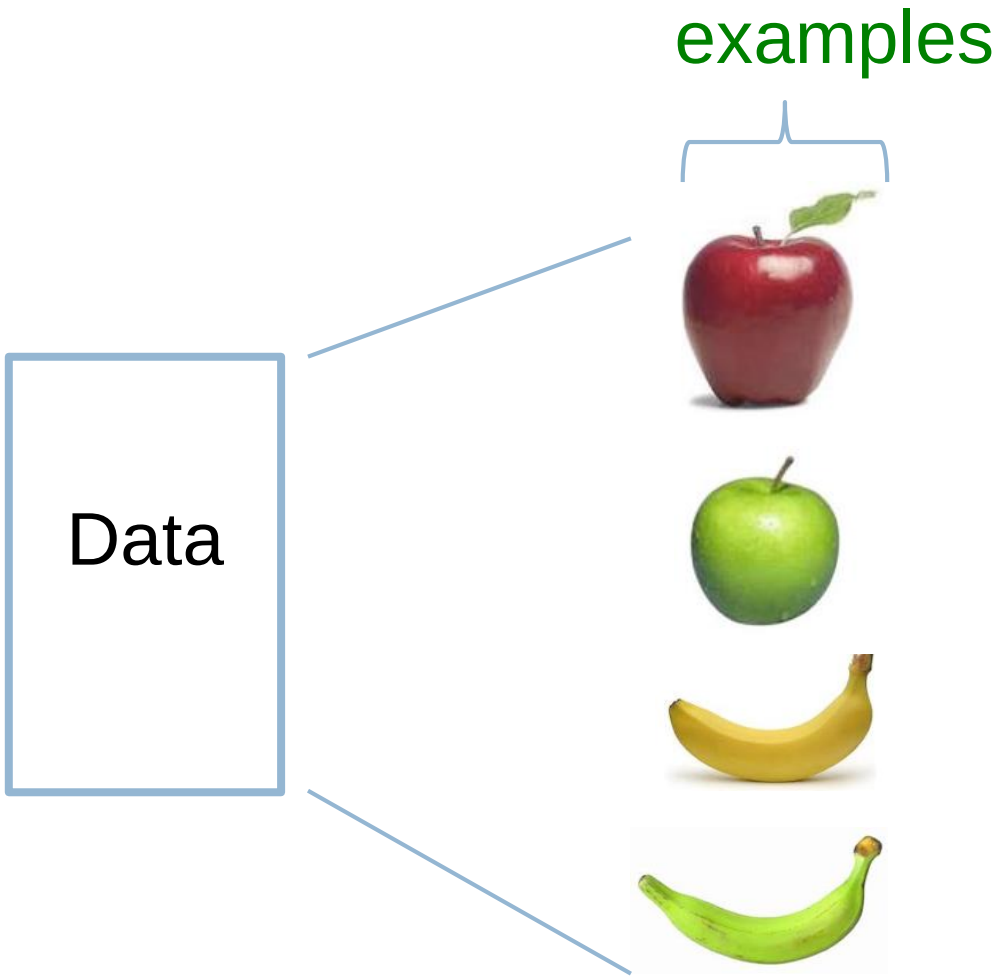
Goals of the course

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

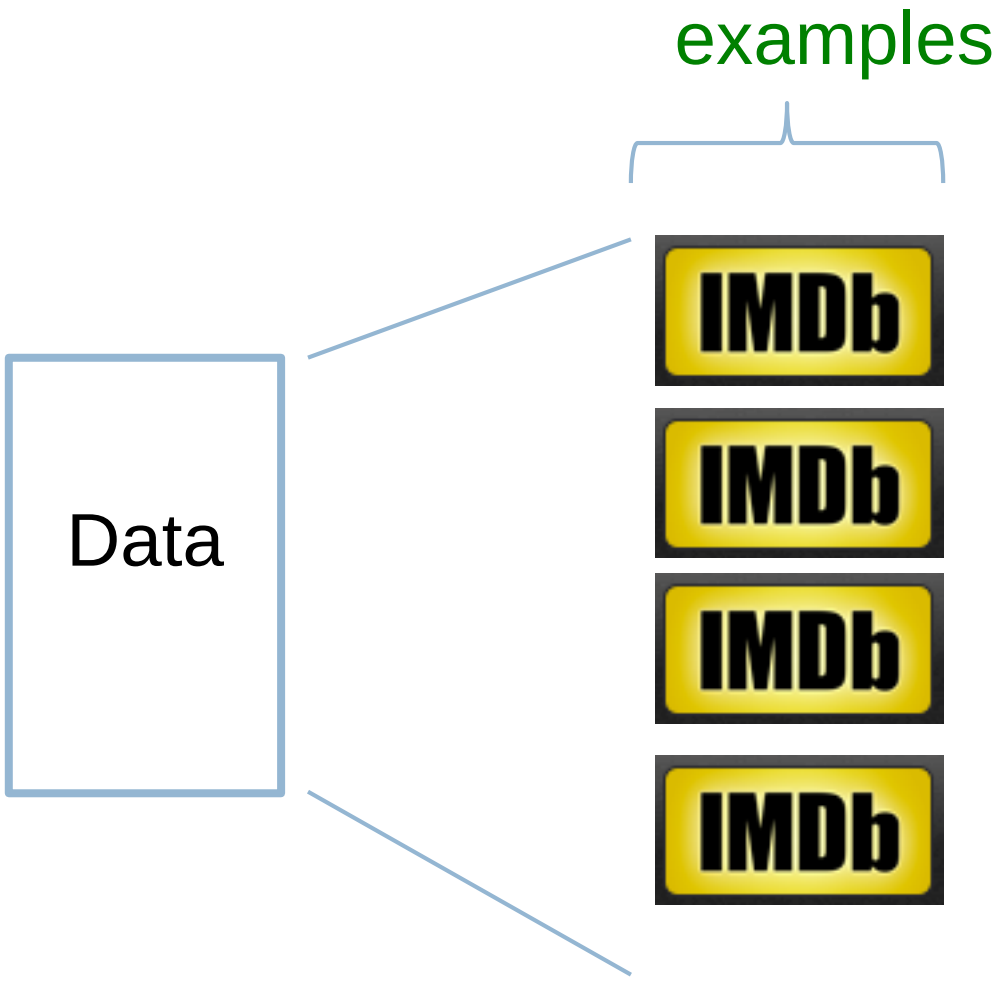
Machine learning problems

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

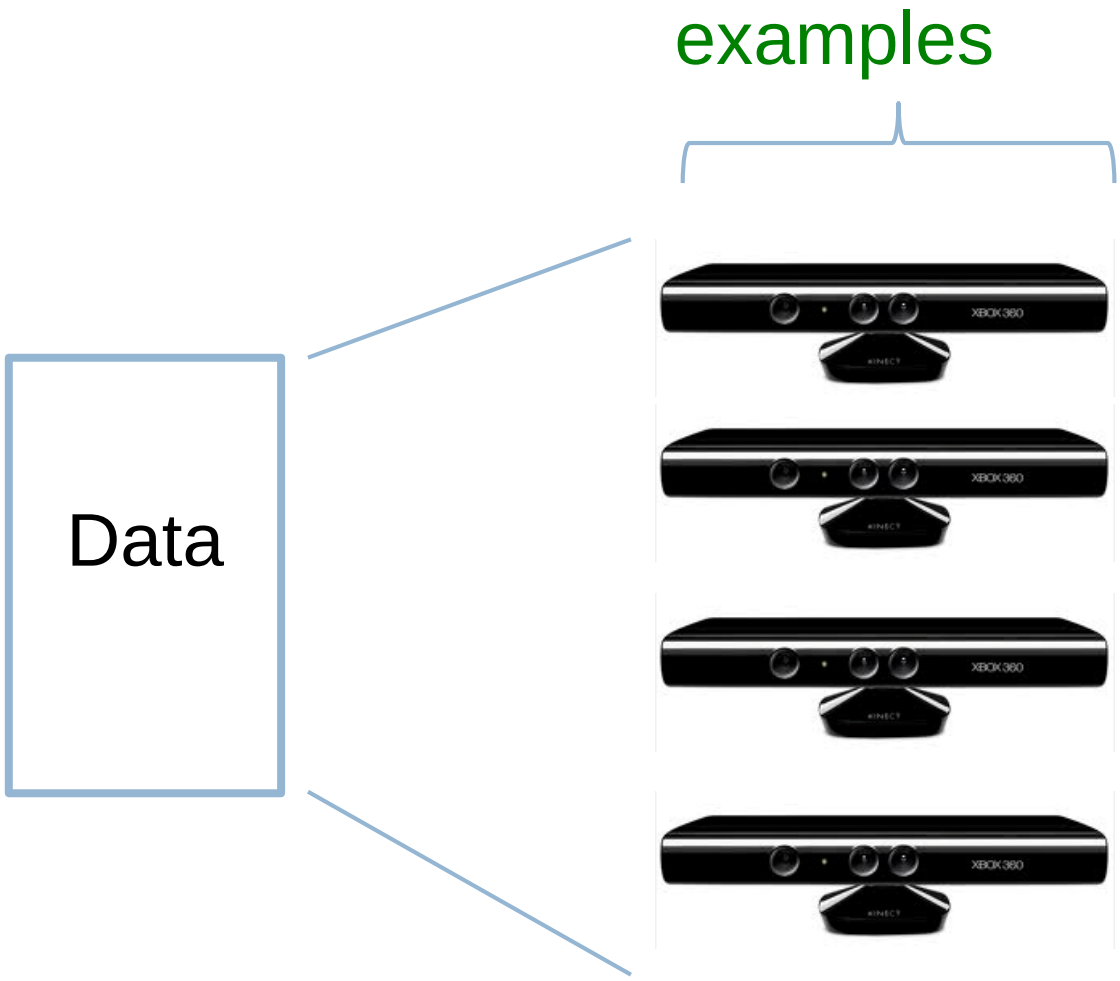
Data



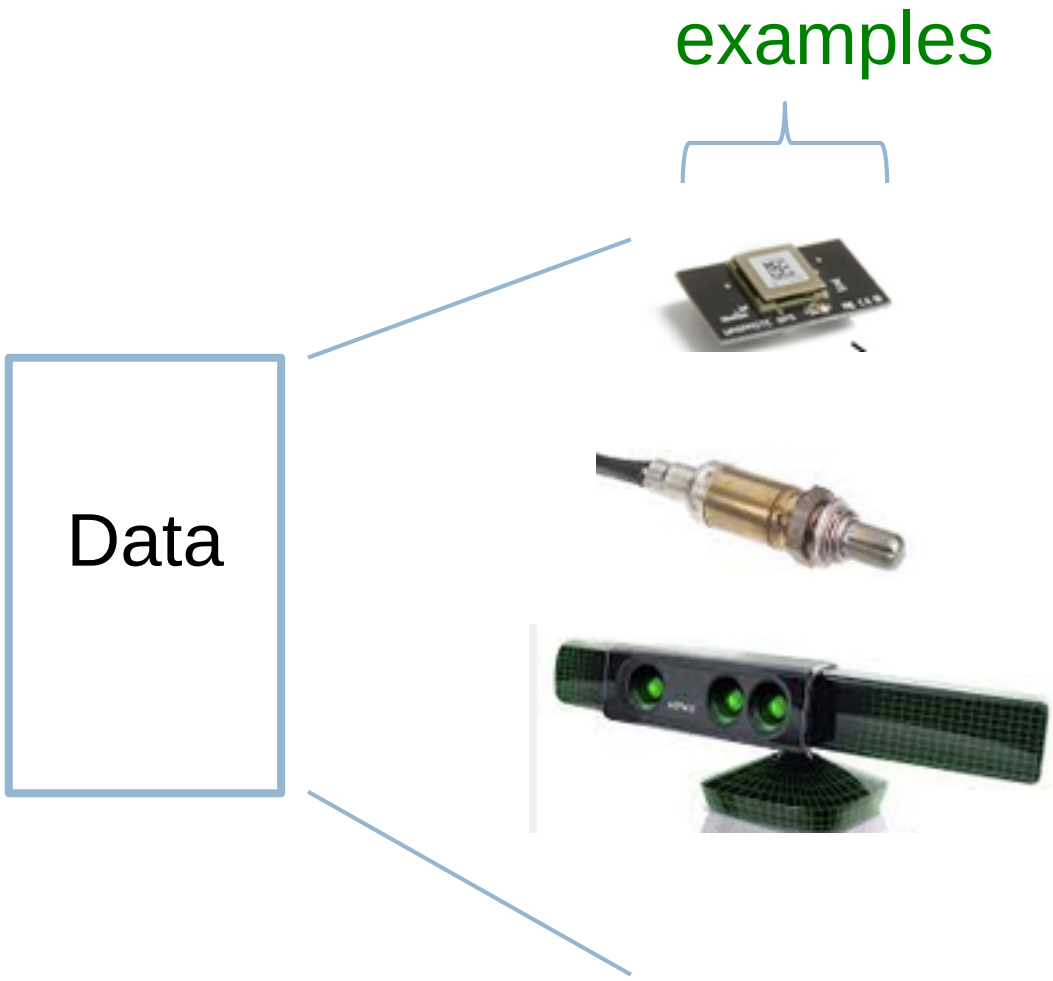
Data



Data



Data



Supervised learning

examples

label

$label_1$

$label_3$

$label_4$

$label_5$

labeled examples

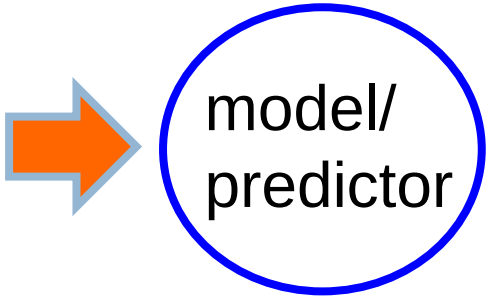
Supervised learning: given labeled examples

Supervised learning

label

$label_1$

$label_3$

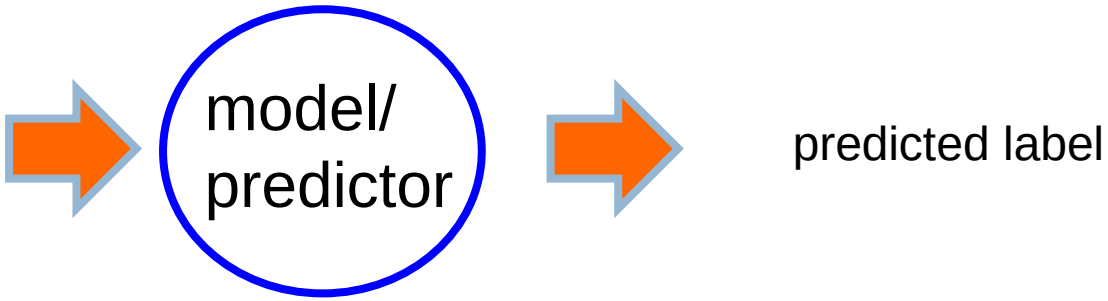


$label_4$

$label_5$

Supervised learning: given labeled examples

Supervised learning



Supervised learning: learn to predict new example

Supervised learning: classification

label

apple

apple

banana

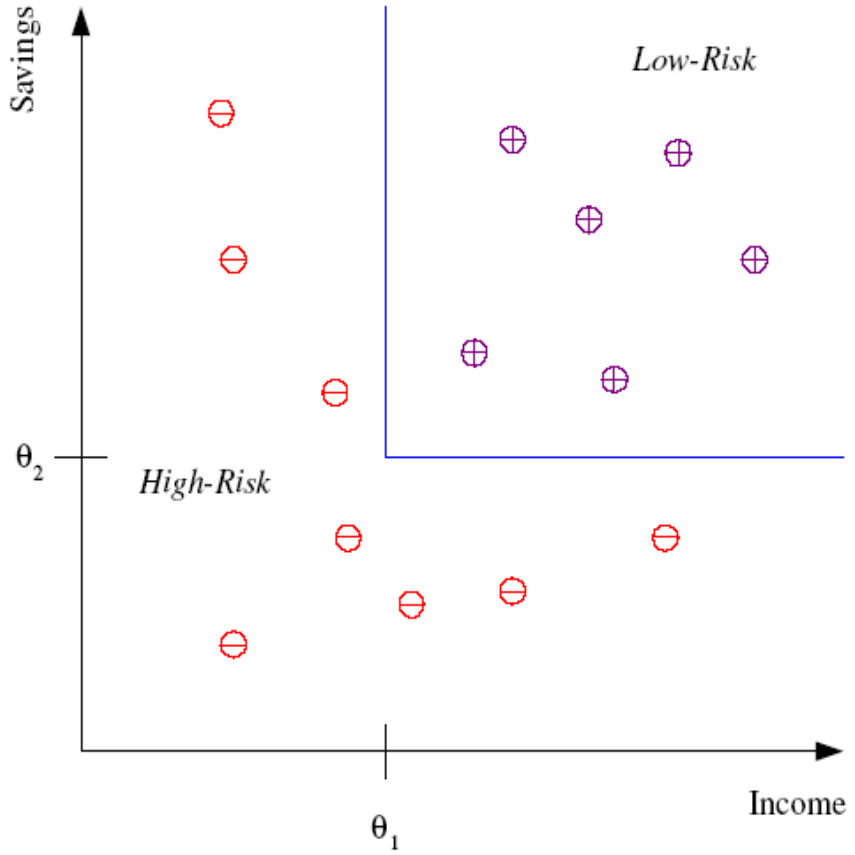
banana

Classification: a finite set of labels

Supervised learning: given labeled examples

Classification Example

Differentiate between low-risk and high-risk customers from their *income* and *savings*



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

-4.5

10.1

3.2

4.3

label

Regression: label is real-valued

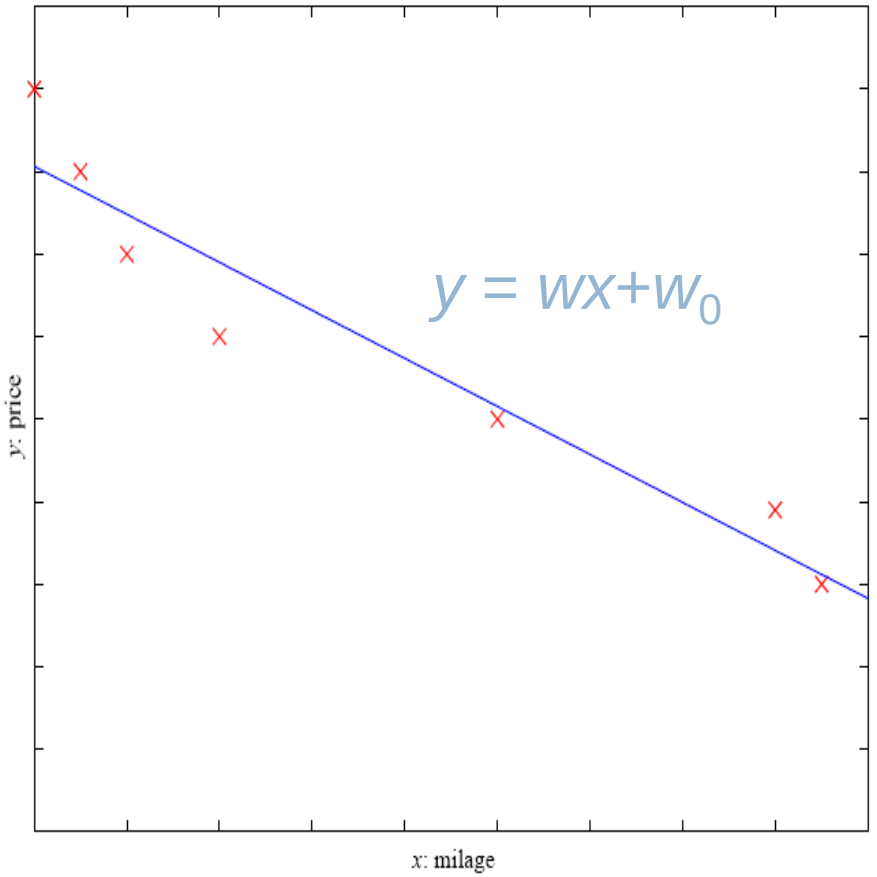
Supervised learning: given labeled examples

Regression Example

Price of a used car

x : car attributes
(e.g. mileage)

y : price



A good reading about regression >
https://stats.stackexchange.com/questions/22381/why-not-approach-classification-through-regression

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

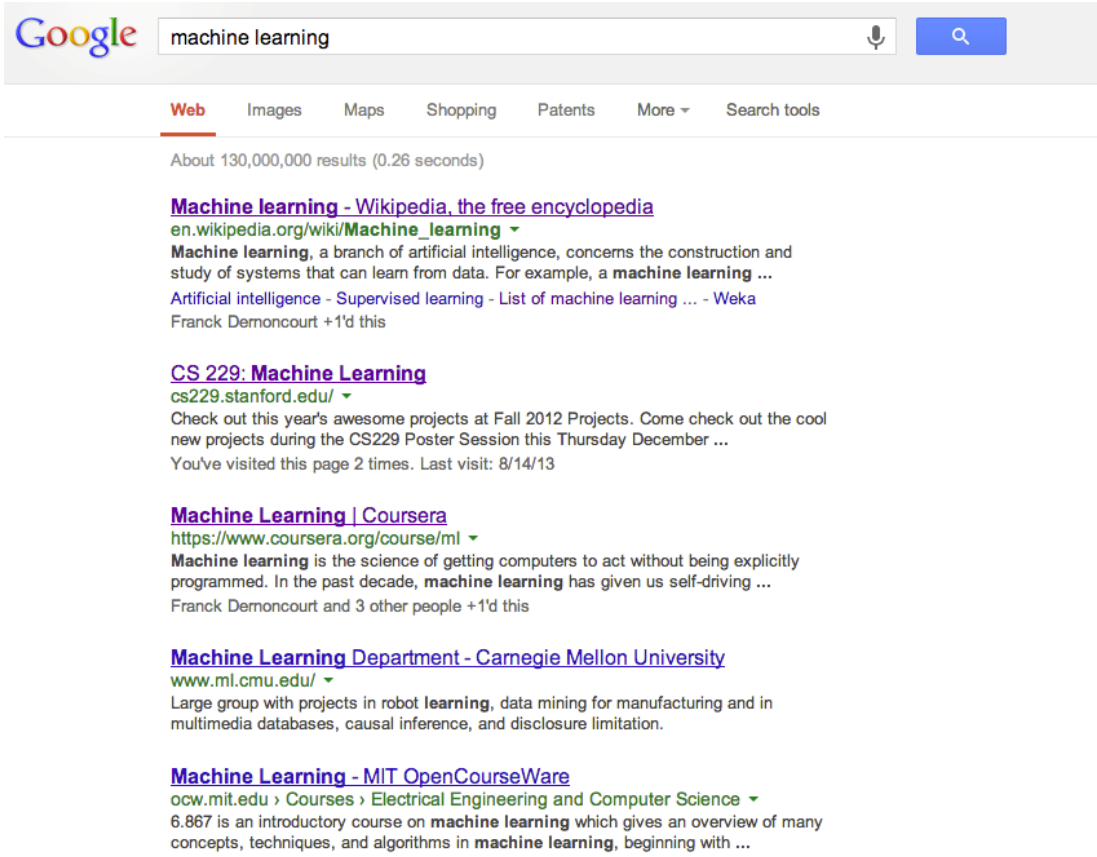
3

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)

re-ranking 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

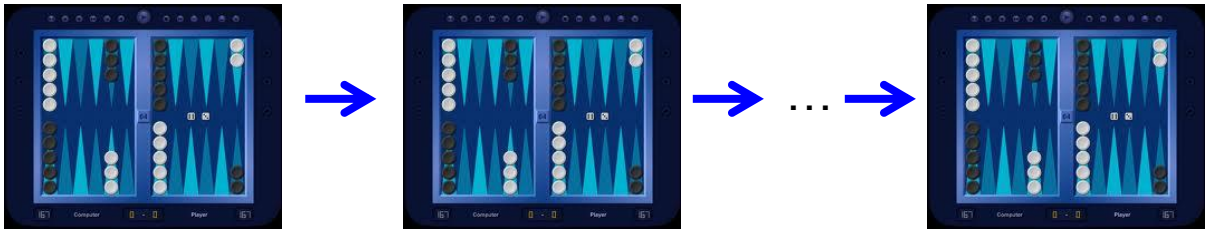
left, right, straight, left, left, left, straight	**GOOD**
left, straight, straight, left, right, straight, straight	**BAD**

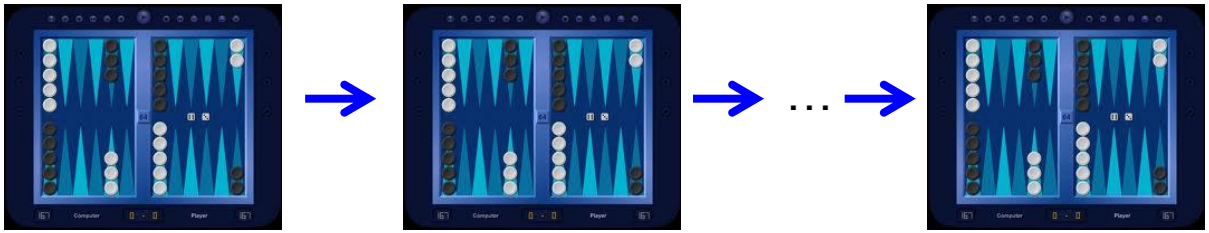
left, right, straight, left, left, left, straight	**18.5**
left, straight, straight, left, right, straight, straight	**-3**

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





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

Reinforcement learning example

http://www.youtube.com/watch?v=VCdxqn0fcnE

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