Masterclass: Introduction to Machine Learning

Kyle Swanson

Massachusetts Institute of Technology

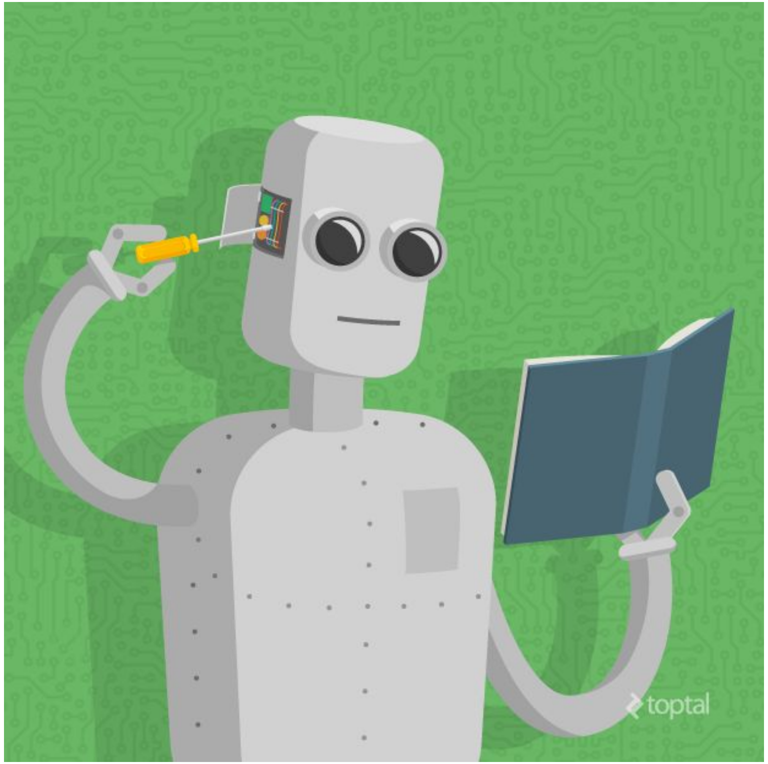
About me

- Name
  - Kyle Swanson
- Hometown
  - Bronxville, NY
- School
  - 4th year at the Massachusetts Institute of Technology (MIT)
  - Studying computer science and mathematics with a minor in music
  - Experience with machine learning, computer vision, and natural language processing
- Research interests
  - Working with Regina Barzilay in MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL)
  - Applying deep learning to medical imaging to improve cancer detection

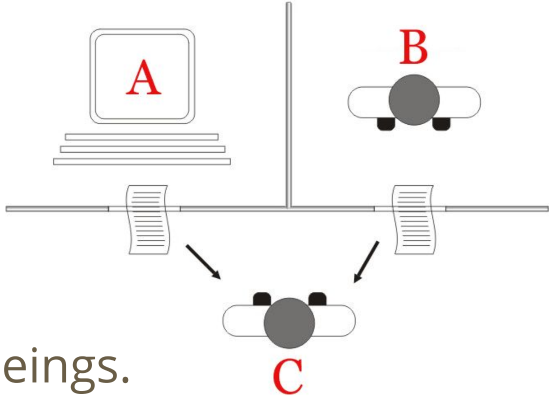

ONTARIO
QUEBEC
WASHINGTON
MONTANA
NORTH DAKOTA
MINNESOTA
NB
PE
Ottawa
Montreal
MAINE
NOVA SCOTIA
OREGON
IDAHO
SOUTH DAKOTA
WISCONSIN
MICHIGAN
Toronto
VT
NH
WYOMING
NEW YORK
MA
CT
RI
IOWA
Chicago
NEBRASKA
ILLINOIS
OHIO
PENN
Philadelphia
United States
NEVADA
UTAH
COLORADO
KANSAS
INDIANA
MISSOURI
MD
DE
NJ
WEST VIRGINIA
San Francisco
KENTUCKY
VIRGINIA
CALIFORNIA
Las Vegas
OKLAHOMA
TENNESSEE
NORTH CAROLINA
Los Angeles
ARIZONA
NEW MEXICO
ARKANSAS
San Diego
Dallas
MISSISSIPPI
SOUTH CAROLINA
ALABAMA
GEORGIA
TEXAS
LOUISIANA
Houston
Gulf of California
FLORIDA
Gulf of Mexico
Mexico
Cuba
Dominican Republic
Mexico City
Puerto Rico

Today

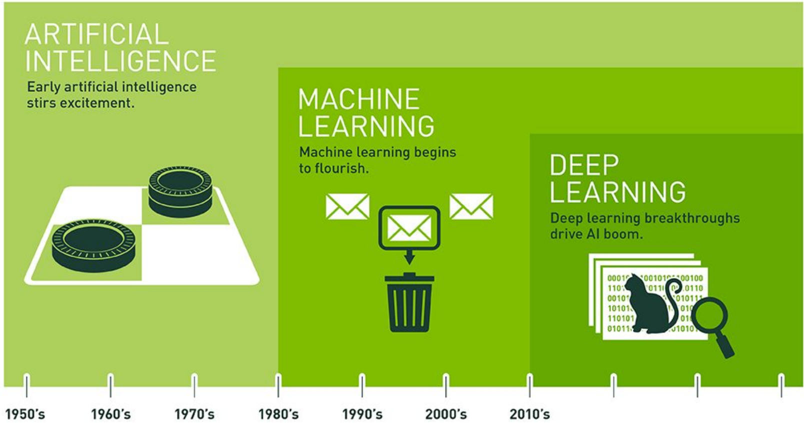
- What is machine learning?
- History
- Types of machine learning problems
- Methods of learning
- Example: Decision trees
- Machine learning applications
- How to apply machine learning



What is machine learning?



- **Artificial intelligence:** The ability of a computer to perform tasks commonly associated with intelligent beings.
  - Turing Test, 1950: *"Are there imaginable digital computers which would do well in the imitation game?"*
- **Machine learning:** "Field of study that gives computers the ability to learn without being explicitly programmed." (Arthur Samuel, 1959)
- **Deep learning:** A subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks.



History

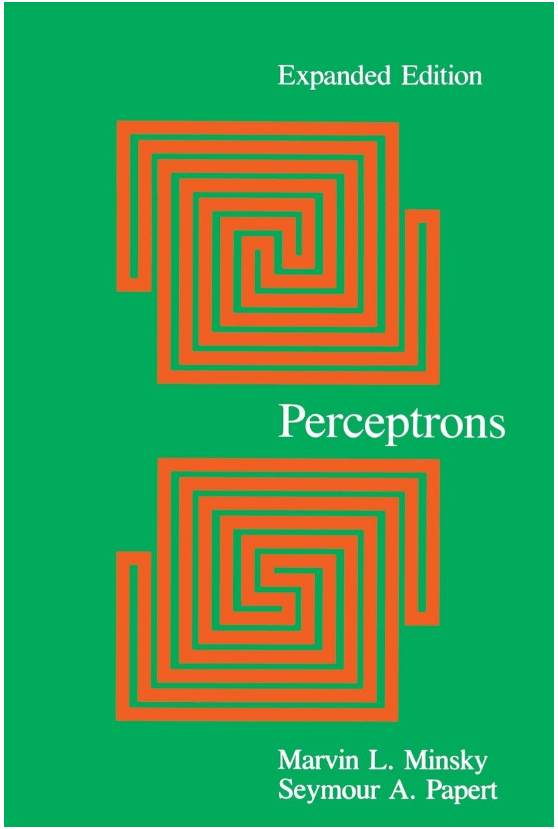
- Artificial intelligence
  - Dartmouth Summer Research Project on Artificial Intelligence, 1956
    - Organized by John McCarthy, who coined the term "artificial intelligence" in 1955
    - Attended by Marvin Minsky, Claude Shannon, other founding fathers of AI

    "An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer."

  - Boom and bust cycles
    - AI winters of 1974-1980, 1987-1993
  - AI used search and rule-based methods (video)
  - IBM's Deep Blue beats Garry Kasparov in chess, 1997

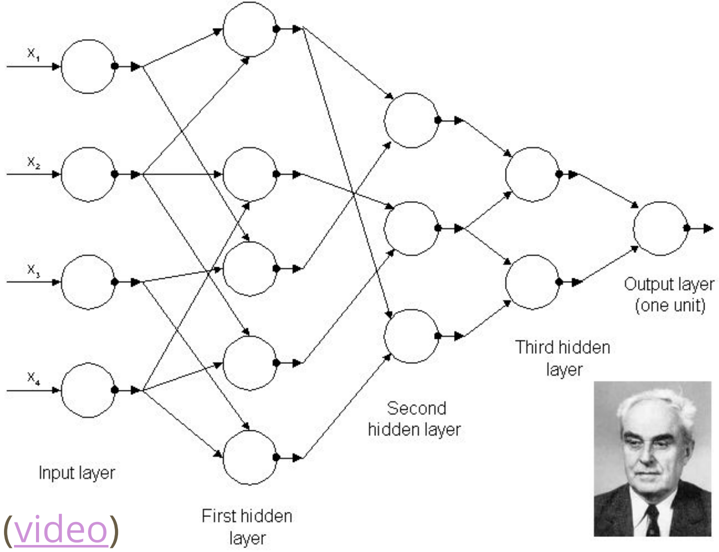
History

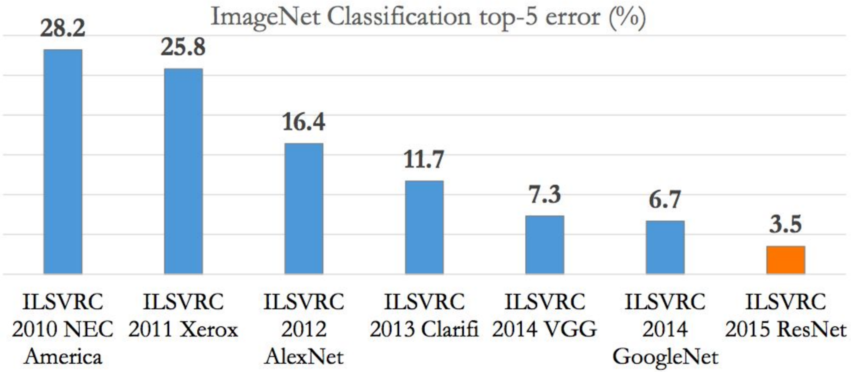
- Machine learning
  - "Machine learning" coined by Arthur Samuel, 1959
    - Developed a machine which learned to play checkers
  - Perceptron invented by Frank Rosenblatt, 1957 (video)
  - *Perceptrons* by Marvin Minsky and Seymour Papert, 1969
    - Limitations of single-layer Perceptron
    - AI winter of the 1970s

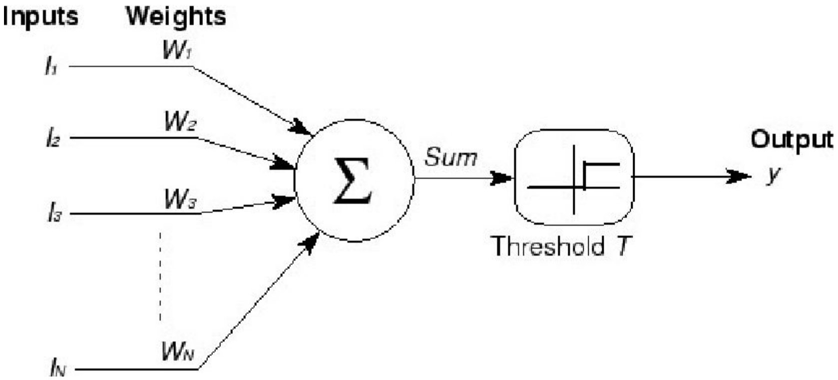


History

- Deep learning
  - Threshold Logic Units of McCulloch and Pitts, 1943
  - First deep architecture by Alexey Ivakhnenko, 1965
  - Advanced models
    - CNN for digit recognition by Yann LeCun, 1993 (video)
    - LSTM for natural language by Hochreiter and Schmidhuber, 1997
  - Success only came with GPUs and AlexNet on ILSVRC, 2012





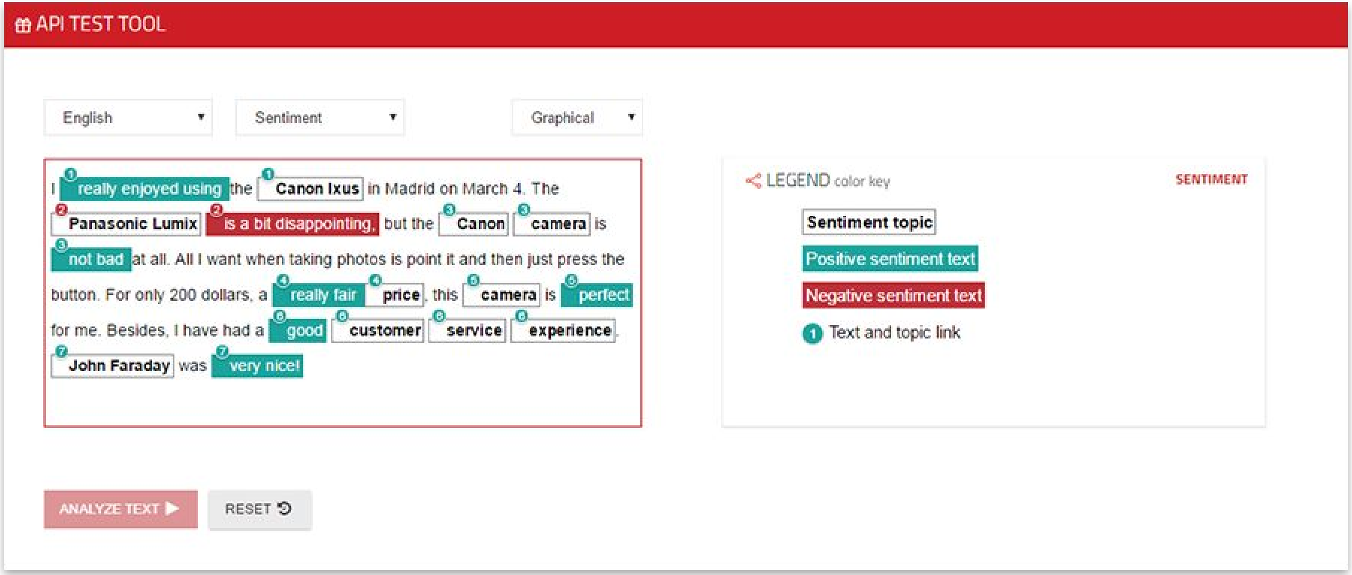


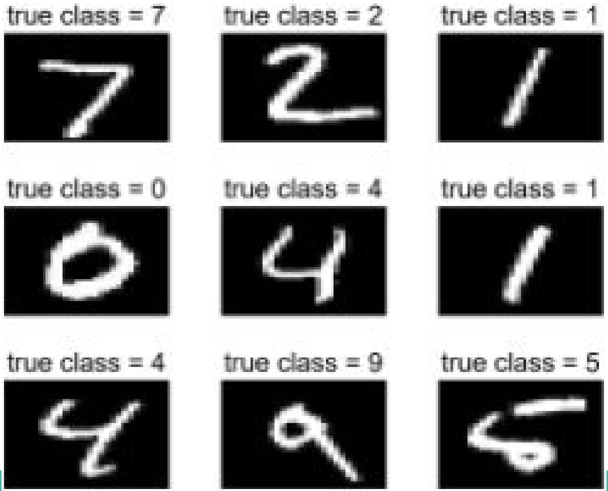
Types of problems

- Classification
    - Discrete prediction
        - Choosing one option among several
- Regression
    - Continuous prediction
        - Predicting a real number
- Generation
    - Output creation
        - Generating text or images

Classification

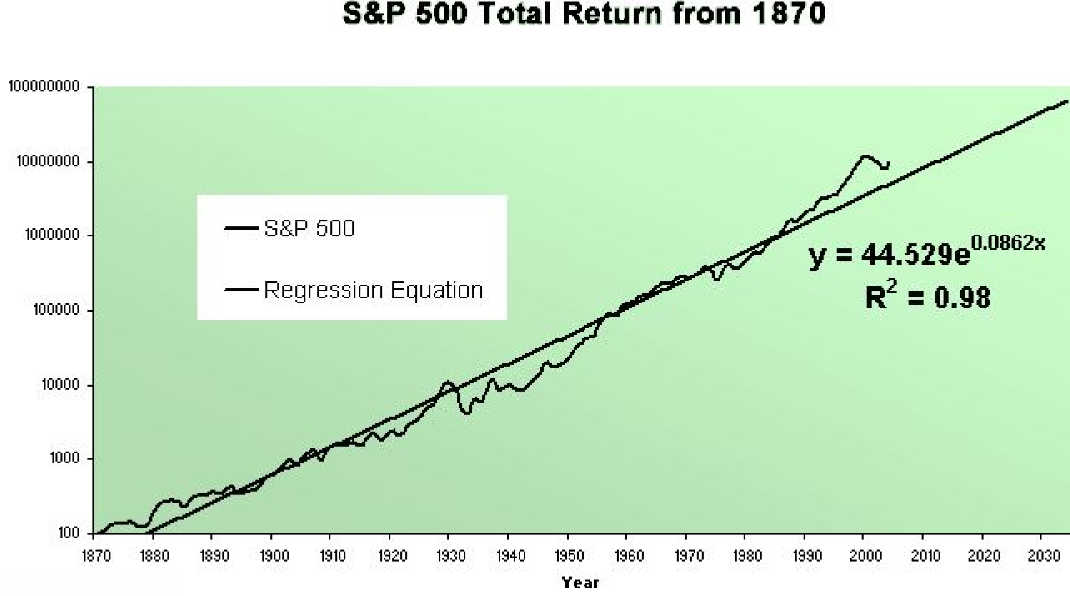
- Discrete prediction



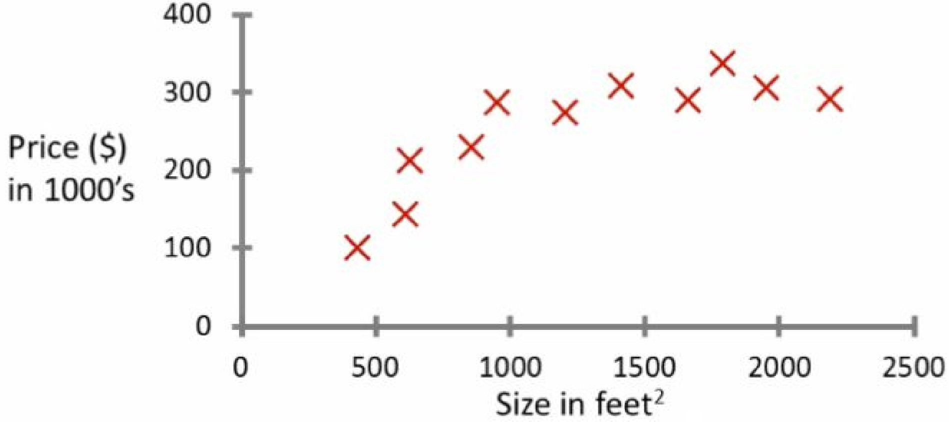


Regression

- Continuous prediction



Housing price prediction.



Generation

- Output creation

"Two pizzas sitting on top of a stove top oven"

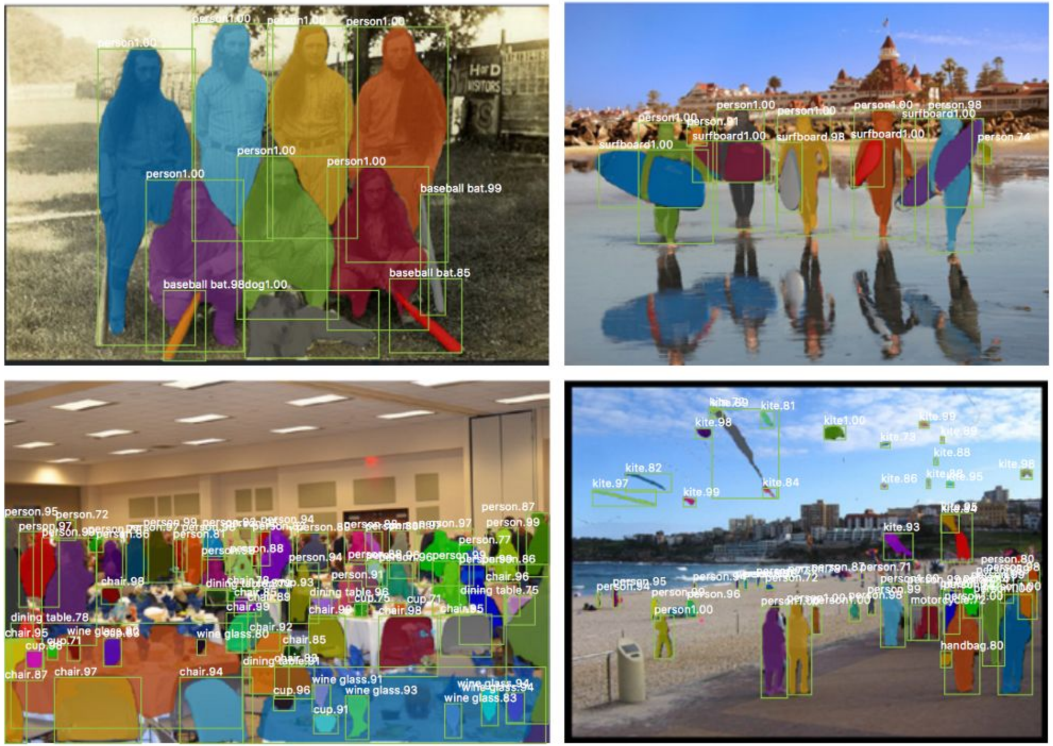
"A group of young people playing a game of frisbee"

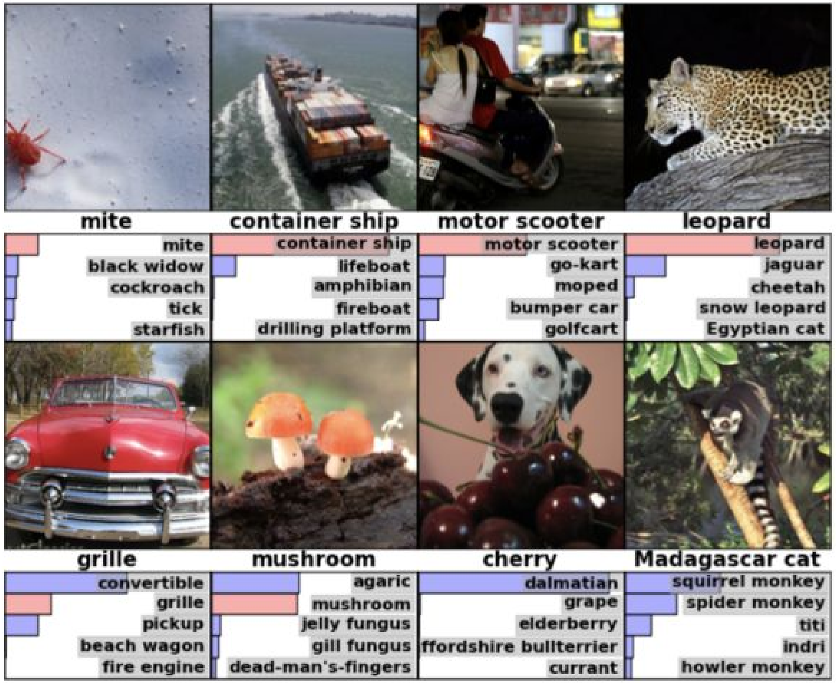
Types of learning

- Supervised learning
  - Given a set of data and labels, learn to predict the labels
- Reinforcement learning
  - Given a task and a reward function, learn to perform the task
- Unsupervised learning
  - Given a set of data *without* labels, learn underlying features of the data

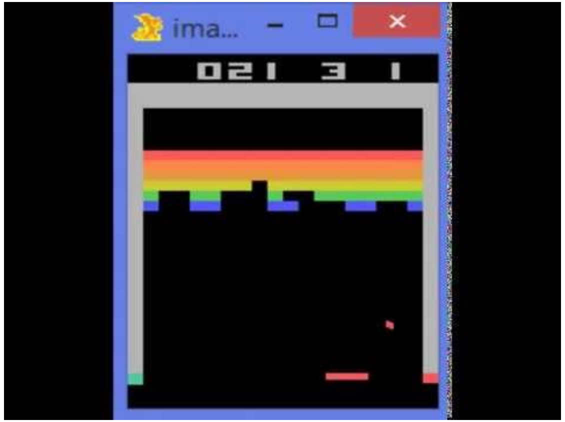
Supervised learning

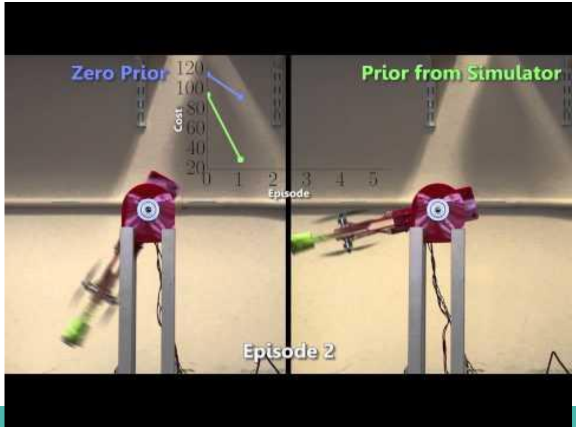
-1 -1 +1 +1



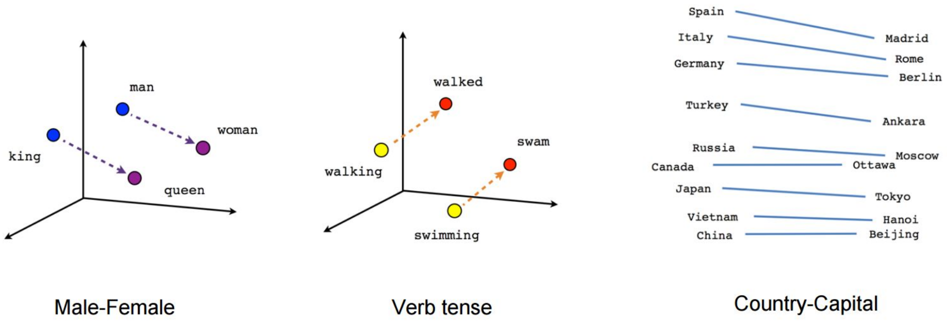


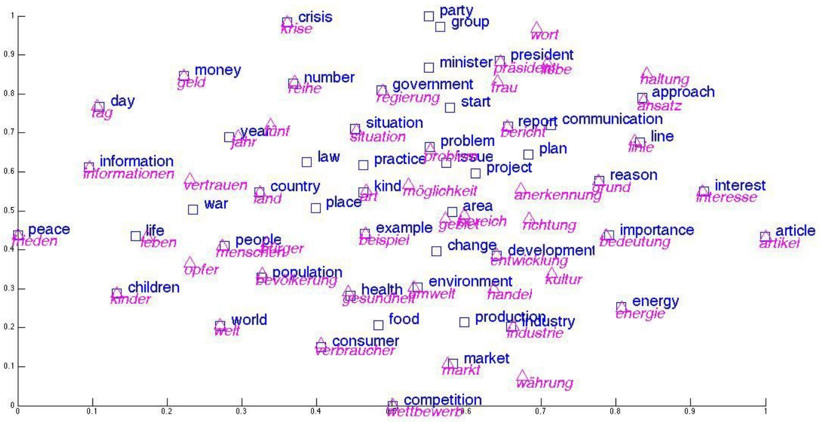
Reinforcement learning

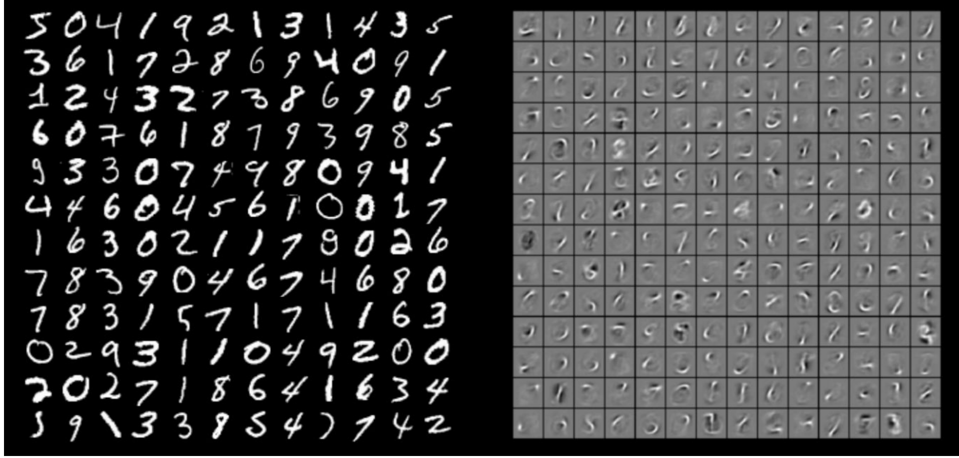


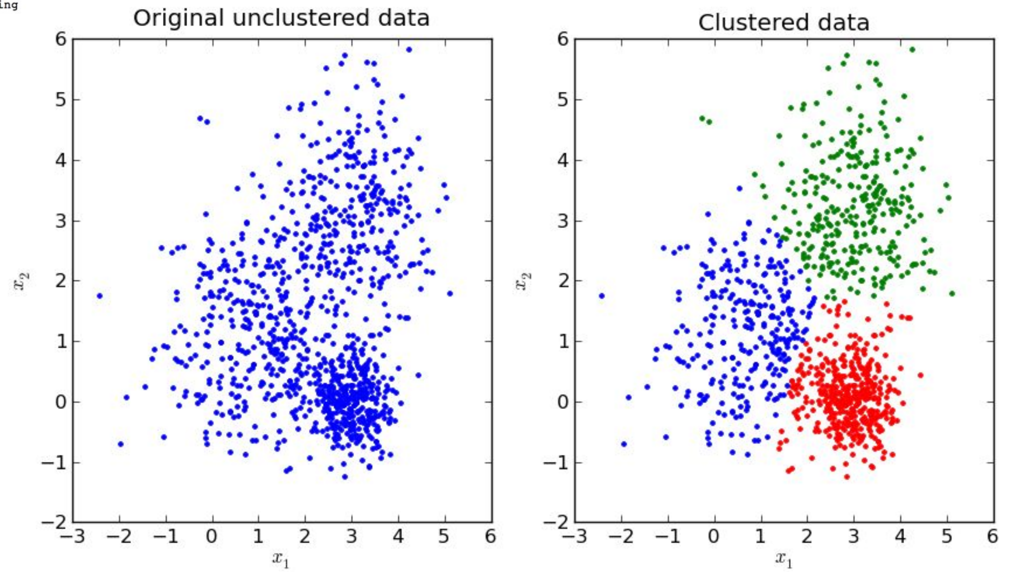


Unsupervised learning



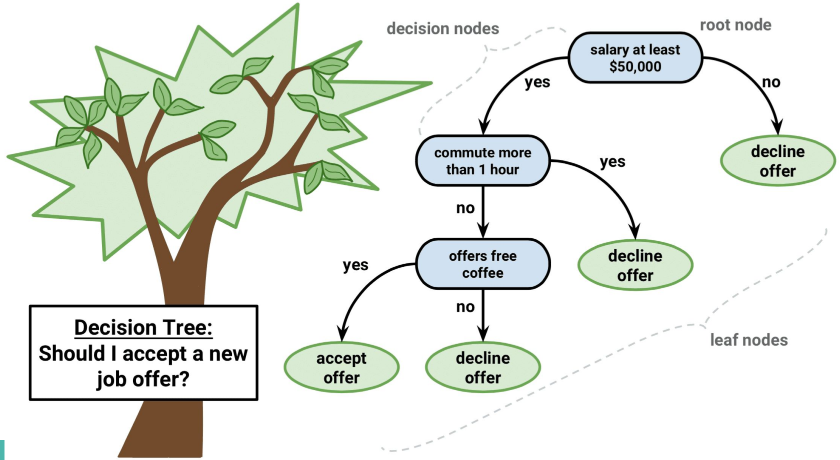




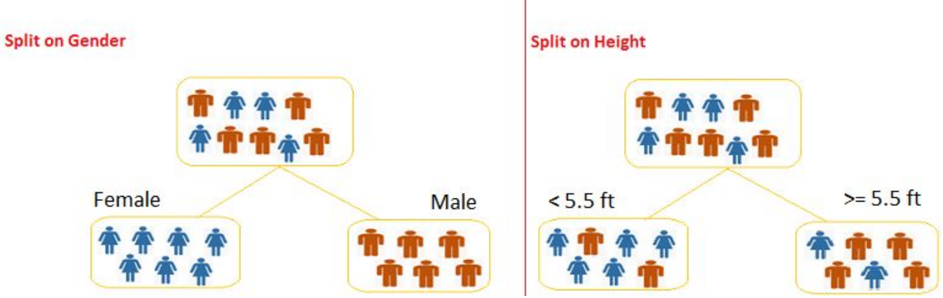


Decision trees

- Simple supervised classification algorithm
- Builds a flowchart for making a decision
- Rather than manually constructing the rules, the rules are *learned*
- Rules are selected based on which features best split the data
- Multiple flowcharts are combined and averaged for better accuracy

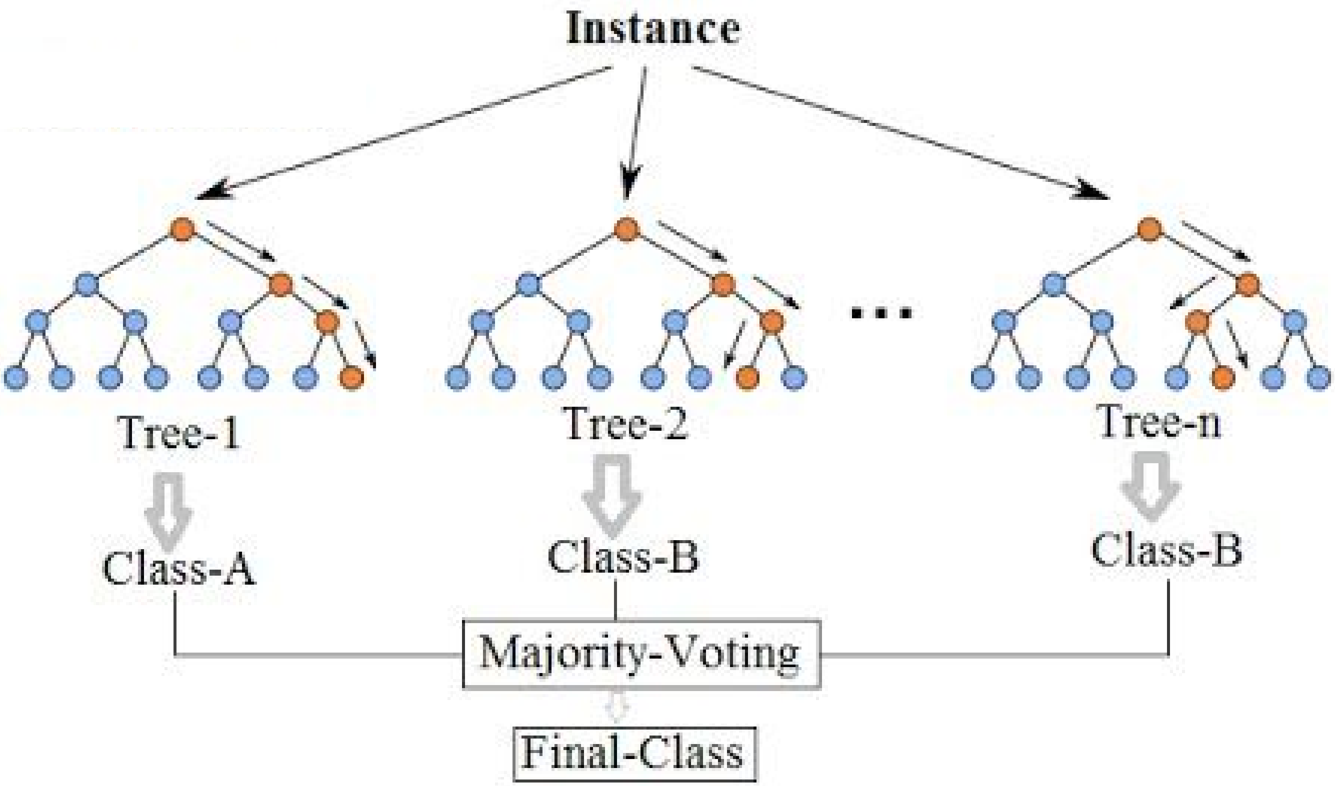


Decision tree algorithm



1. Select a subset of the data
2. Determine the information gain for each possible rule
   a. Intuition: which rule best splits the data?
   b. Math: which rule produces a split with the least entropy?
3. Select the rule with the greatest information gain
4. Insert a decision node with two child nodes (one for yes, one for no)
5. Repeat steps 2-4 for each child node
   a. Stop when a child node perfectly splits the remaining data
6. Build additional decision trees with different subsets of the data
   a. The final prediction is the average prediction of the trees

Decision tree example



Machine learning applications

- Machine learning is everywhere!
  - On the web
  - On your phone
  - In your house
  - In the hospital
  - Etc.
- Two main areas of machine learning:
  - Natural language processing
  - Computer vision

Fraud/spam detection

Dear Dr Pape,

My client is looking for a Java developer.
Are you ready for the next challenge?
Call me: +49(0)40XXX-XXX-XXX-XX

Yours faithfully,
XYZ

SPAM

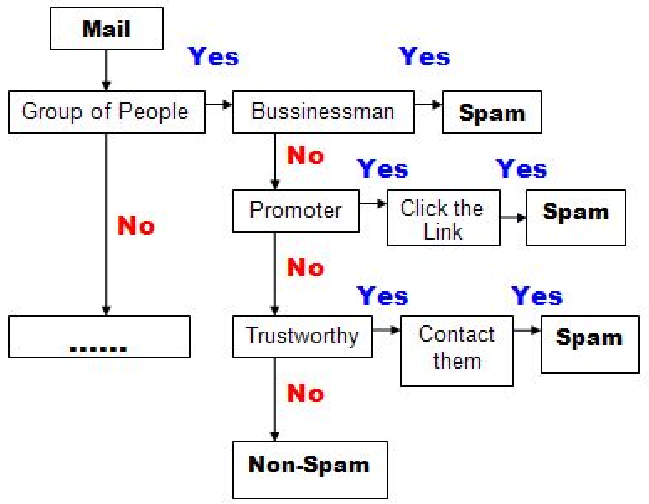
vs.

Hey Daniel,

Thanks again for the talk at yesterdays meetup. I think I've found an answer to the question we've been discussing and wanted to share....

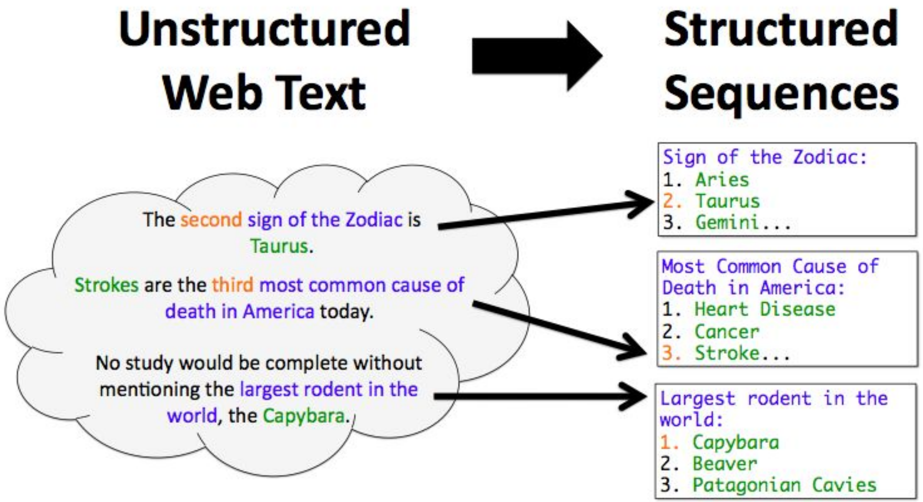
Yours,
XYZ

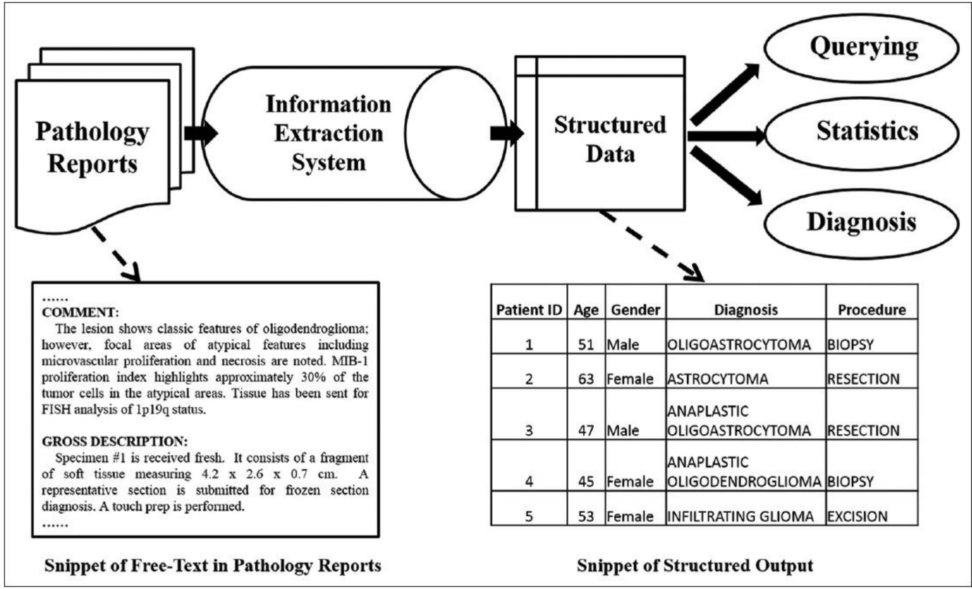
HAM



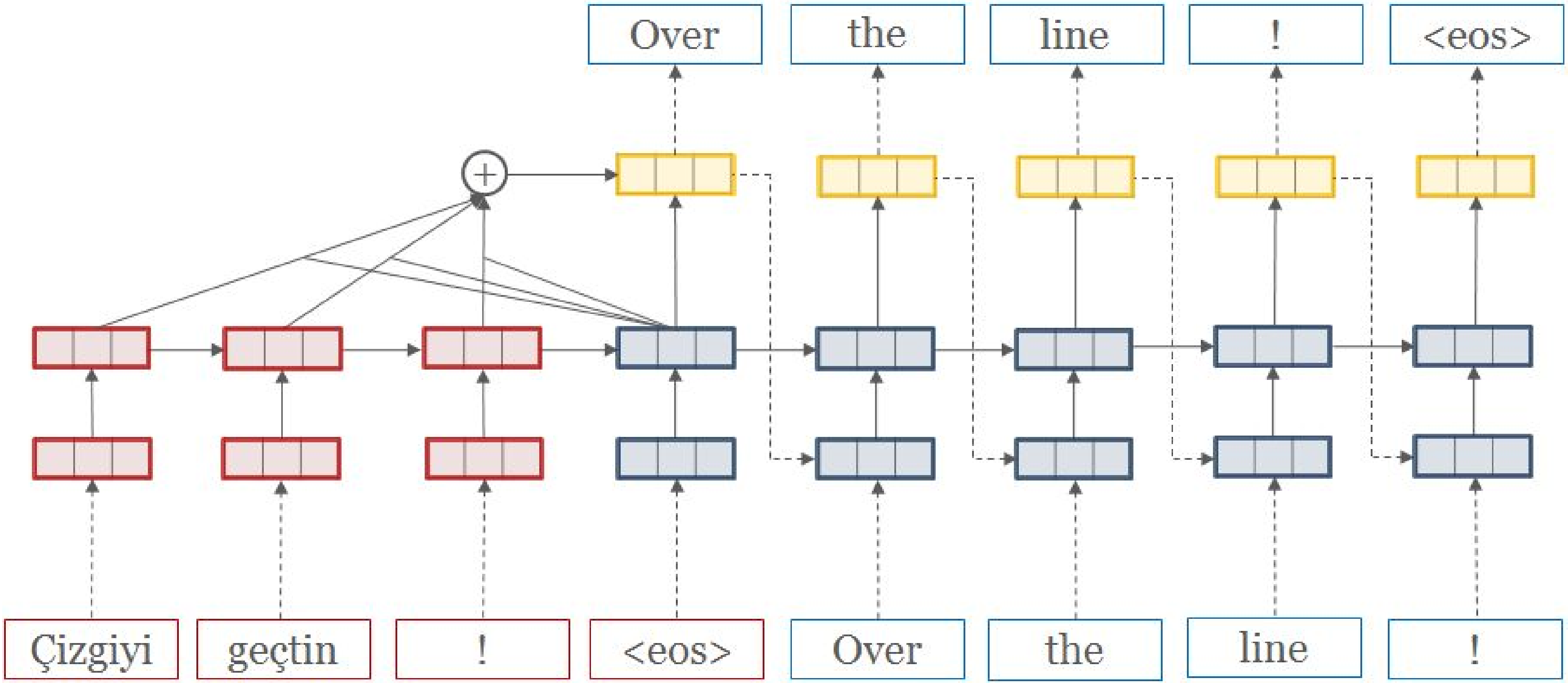
The figure of cascaded model designed

Information extraction

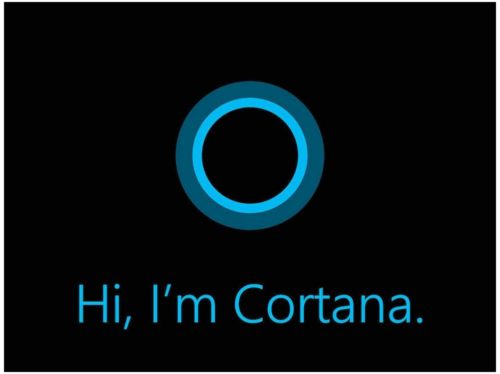


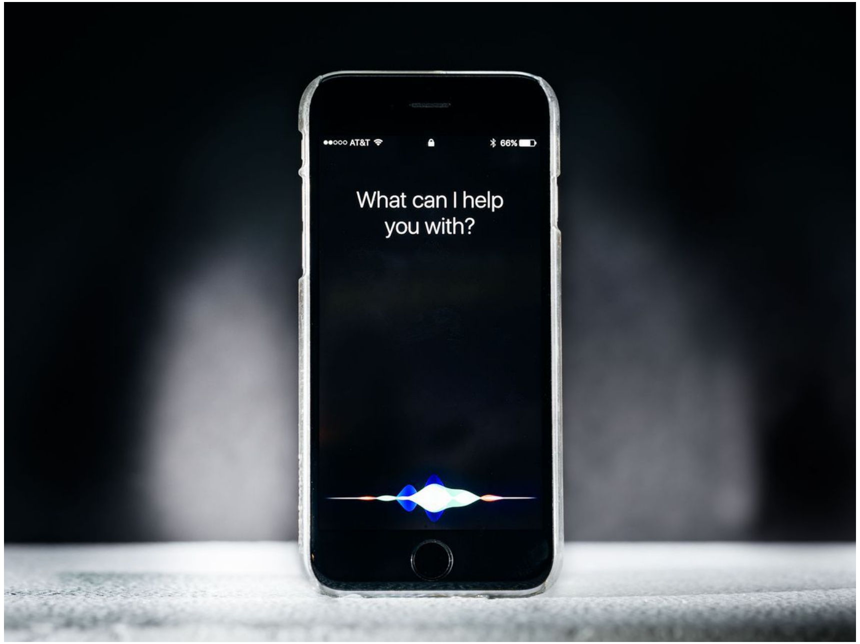


Machine translation



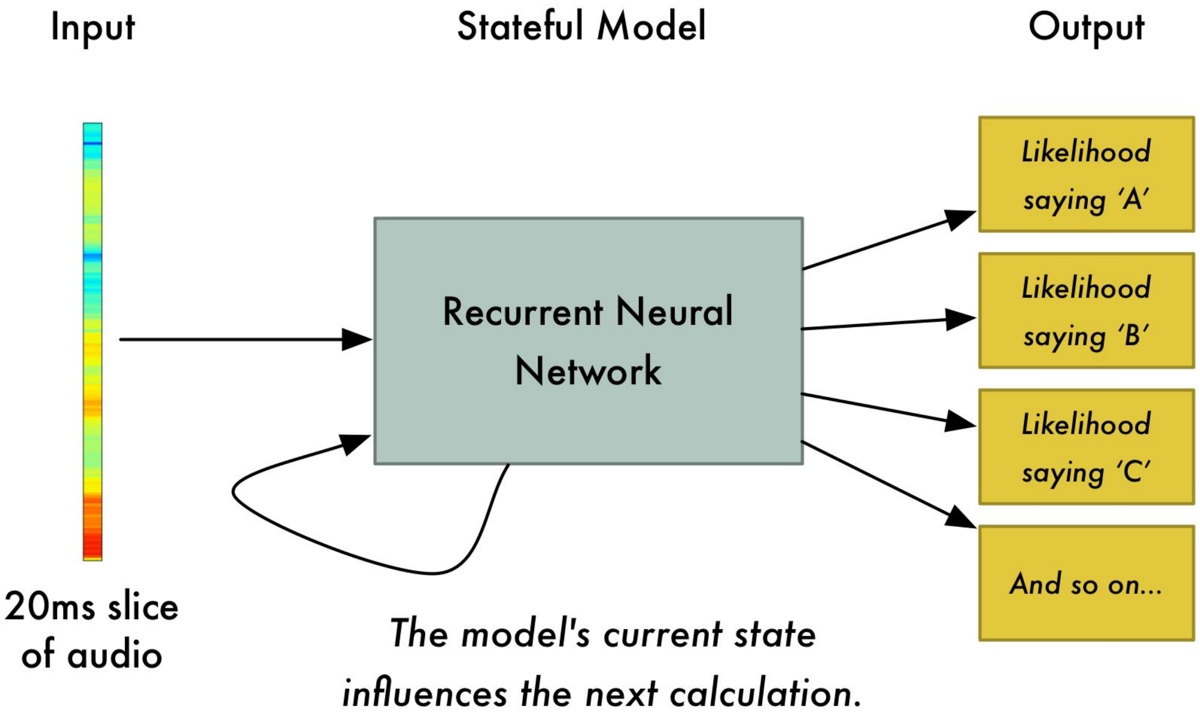
Chatbots

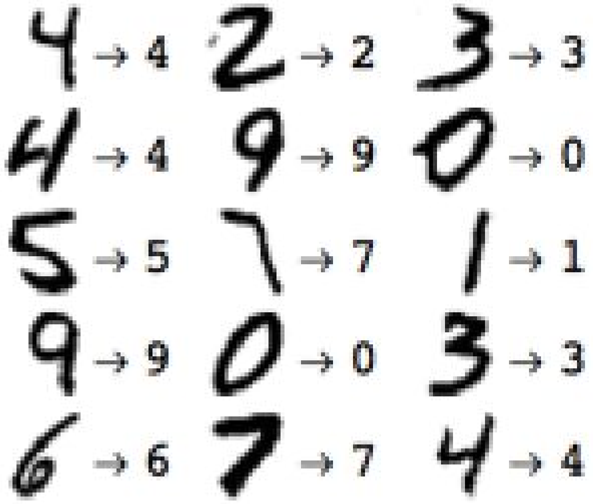




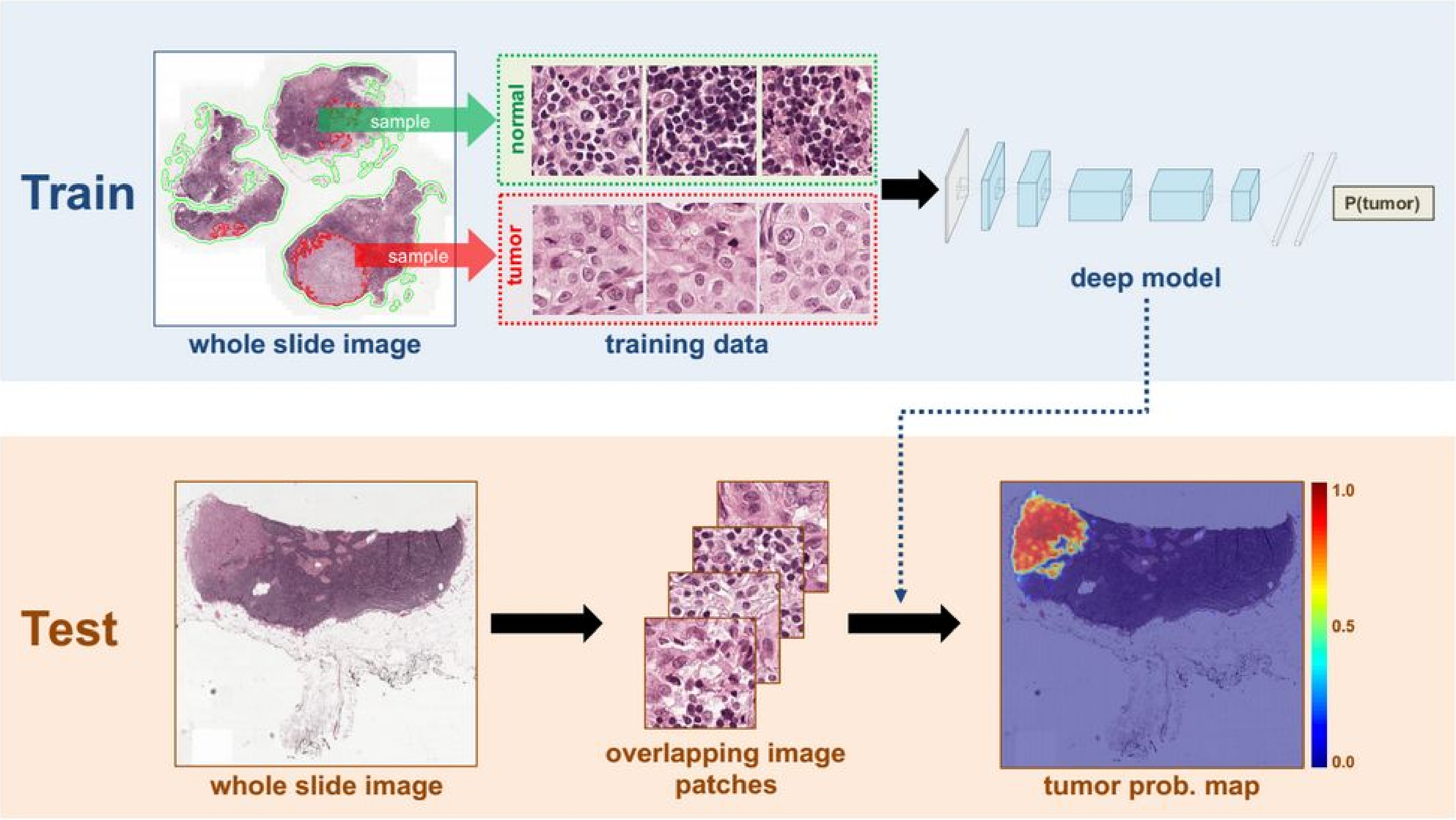
Text suggestion

Speech/handwriting recognition

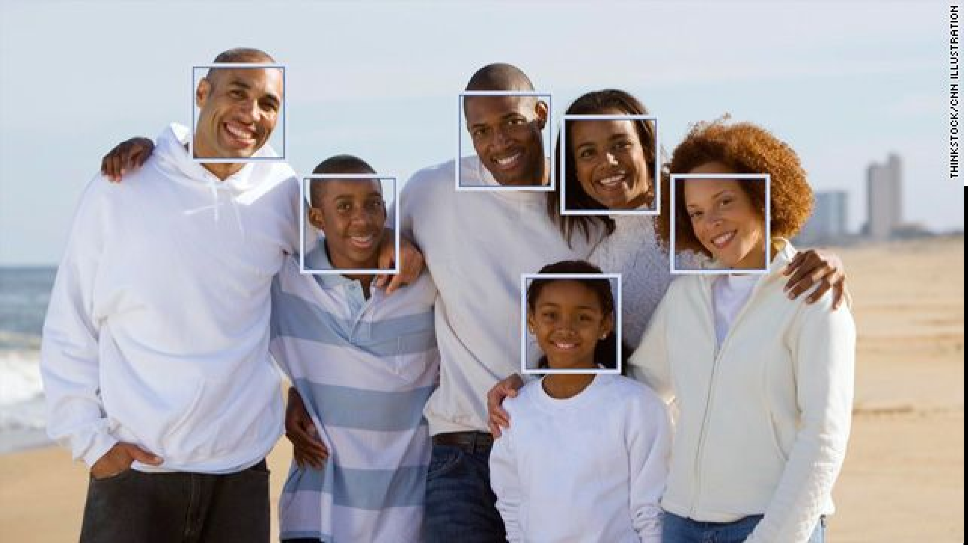


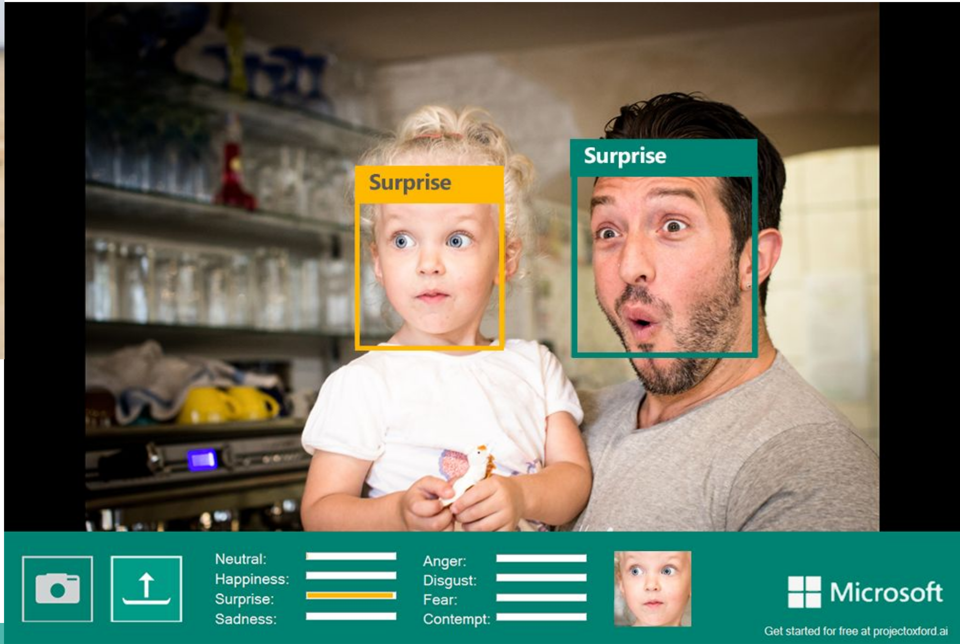


Medical diagnosis

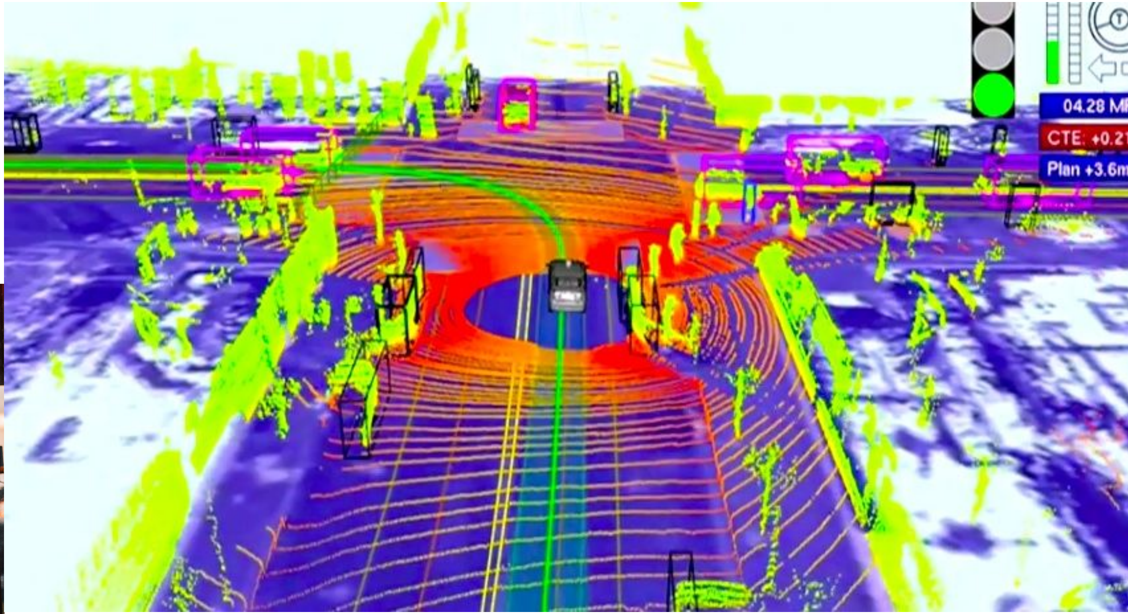


Facial detection

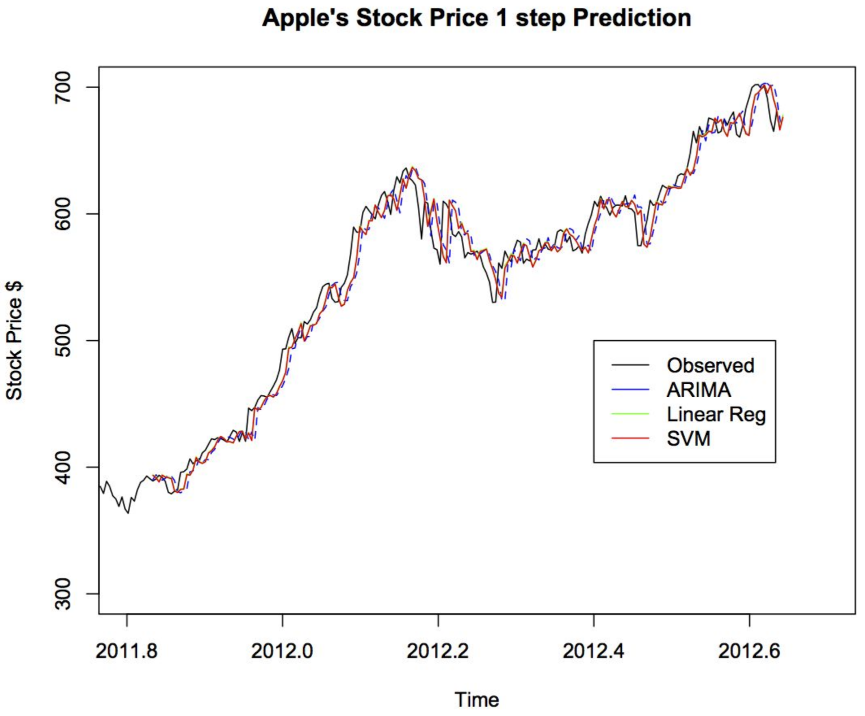




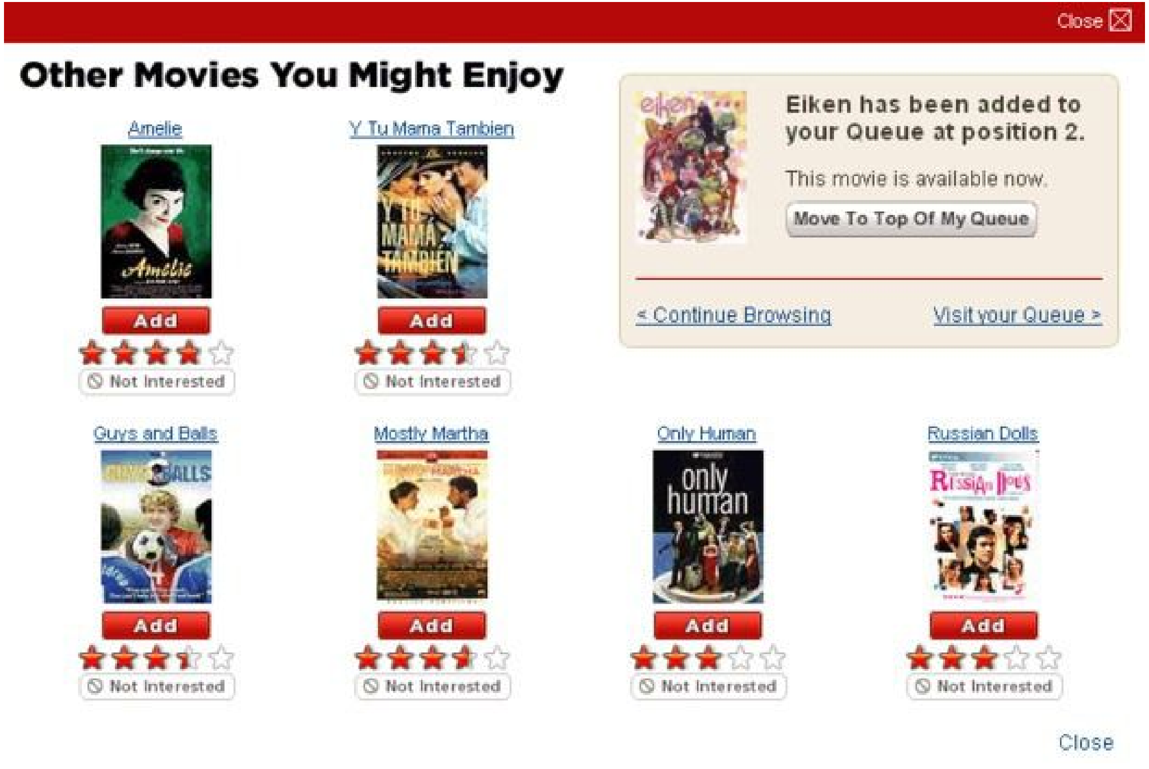
Self driving cars



Stock market prediction

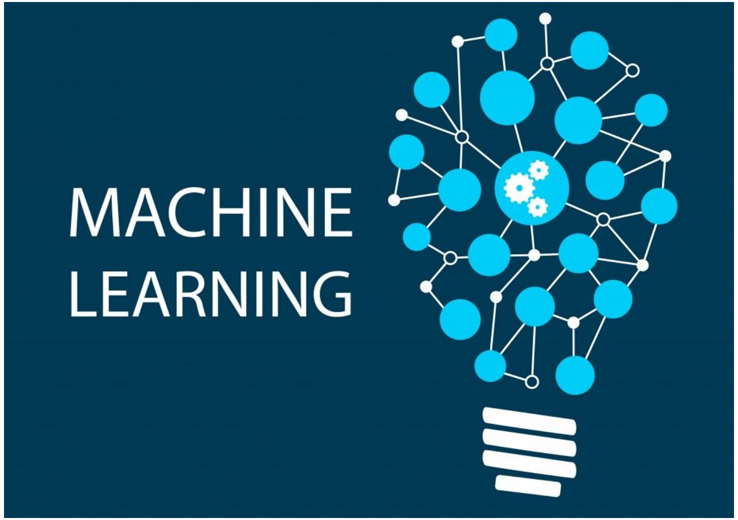


Content recommendation



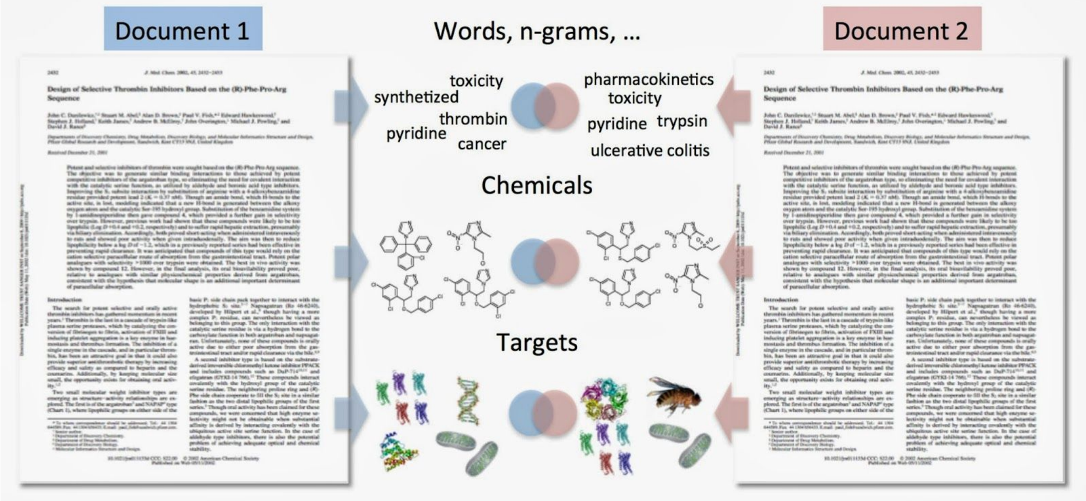
How to apply machine learning

1. Define the problem
2. Collect and process the data
3. Determine a baseline
4. Improve results
5. Implement the model



1 - Define the problem

- What processes are manual, slow, and inefficient?
- Problems typically involve processing large amounts of data
- How could automation fit into the workflow?
- Example: Detecting similar medical records



2 - Collect and process the data

- Large amounts of training data are necessary
- Data needs to be labelled!
  - Internal workers
  - Outsource to Mechanical Turk
- Good data storage and formatting are essential
  - Half of the challenge is data processing

Bad data formatting

- Free text
- Loosely organized data
- Data in multiple locations
- Bad data management practices
  - Hard to access
  - Rarely updated
  - Insecure

BeamReaders®

7117 W. Hood Place #110
Kennewick, WA 99336
P 916.771.3505
F 916.646.3742
https://www.beamreaders.com

Patient: Patient Name
DOB: 04/13/1900
Report Date: 11/27/2012
Study Date: 10/25/2012

Ref. Doctor: Dr. Doctor
Scan Source: *Your Imaging Center*

Study Purpose: TMD Orofacial Pain
Dr. Notes: Please evaluate TMJ's & possible DJD especially within the left TMJ

OBSERVATIONS

DENTAL FINDINGS: A mixed dentition consistent with the patient's age is present.
Left mandible – The ramus and the body are shorter, the mandibular plane is steeper and the occlusal plane in elevated relative to the right side.

TMJs: *Left TMJ* – The condyle is small and exhibits marked reduction in the vertical dimension. Significant flattening is present on the superior aspect. Flattening and sclerosis is noted along the posterior slope of the articular eminence and the glenoid fossa is shallow. The condyle is minimally anterior to the center of its glenoid fossa; however no significant reduction of the joint spaces is noted.
Right TMJ – The condyle is normal in size and shape with smooth, rounded contours. The cortical outline is diffuse and indistinguishable from the underlying trabecular bone; this is considered normal for the patient's age. There is evidence of minimal flattening and focal areas of subchondral sclerosis on the superior aspect of the condylar head and along the posterior slope of the articular eminence. The condyle is superior to the center of its glenoid fossa and there is resultant narrowing of the superior joint space.

SINUSES: Soft tissue collection is noted along the walls of the maxillary sinuses (R>L), in some ethmoid air cells, and in the sphenoid sinuses (R>L). There is evidence of air bubbles in the right sphenoid sinus. The ostiomeatal complex is narrowed but patent on the right and blocked on the left.

NOSE: Soft tissue collection is noted along the walls and in the meati.

AIRWAY: The dimensions of the airway, posterior to the soft palate and tongue base, are within normal limits. Enlargement of the adenoids and the palatine tonsils is noted; airway patency in the nasopharynx and the oropharynx is however not compromised. Tonsillar enlargement is a common finding in children and they tend to gradually regress after age 12.

OTHER FINDINGS: Cerumen (ear wax) is noted in both external auditory canals.

IMPRESSIONS

- The osseous findings in the **left** TMJ are consistent with advanced degenerative joint disease (DJD) that is radiographically stable at this time. DJD involves the destruction of the articular tissues and occurs when the remodeling capacity of those tissues has been exceeded by the functional demands. The presence of the DJD increases the probability of a displaced disc, has reduced the size of the condyle, may be associated with a change in occlusion and mandibular posture and may predispose the joint to dysfunction. The mandibular changes are secondary to the unilateral DJD of the left TMJ.
  Findings in the **right** TMJ are consistent with osseous remodeling, most likely of functional origin. These changes are typically adaptive and not progressive. The presence of the narrowed superior joint space increases the probability of a displaced disc.
- Soft tissue collection in the nose and paranasal sinuses is consistent with mild to moderate chronic panrhinosinusitis. An acute exacerbation is suspected in the right sphenoid sinus.
- Radiographic findings in the remainder of the CBCT scan are within normal limits; soft tissue evaluation is limited by the CBCT modality.

Sincerely,

Dr. OMR
Dip., American Board of Oral & Maxillofacial Radiology

Good data formatting

- Structured data format
  - Ex. CSV, JSON, XML, SQL table
- Data in one location
- Good data management practices
  - Easy to access
  - Frequently updated
  - Secure

Employees Table

IdNum	LName	FName	JobCode	Salary	Phone
1876	CHIN	JACK	TA1	42400	212/588-5634
1114	GREENWALD	JANICE	ME3	38000	212/588-1092
1556	PENNINGTON	MICHAEL	ME1	29860	718/383-5681
1354	PARKER	MARY	FA3	65800	914/455-2337
1130	WOOD	DEBORAH	PT2	36514	212/587-0013

CSV

	A	B	C	D
1	ID	Gender	City	Monthly_I
2	ID000002C	Female	Delhi	20000
3	ID000004E	Male	Mumbai	35000
4	ID000007H	Male	Panchkula	22500
5	ID000008I	Male	Saharsa	35000
6	ID000009J	Male	Bengaluru	100000
7	ID000010K	Male	Bengaluru	45000
8	ID000011L	Female	Sindhudui	70000
9	ID000012M	Male	Bengaluru	20000
10	ID000013N	Male	Kochi	75000
11	ID000014C	Female	Mumbai	30000
12	ID000016C	Male	Mumbai	25000
13	ID000018S	Female	Surat	25000
14	ID000019T	Female	Pune	24000
15	ID000021V	Male	Bhubanes	27000
16	ID000022V	Female	Howrah	28000

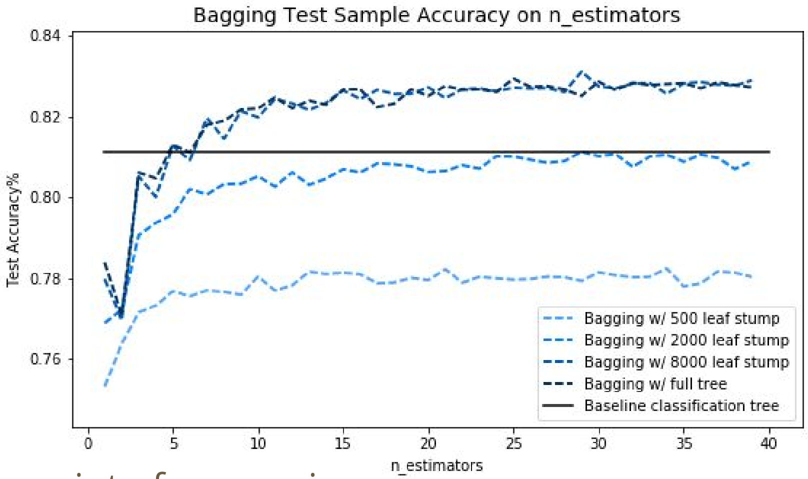
JSON

```
{
  "Employee": [
    {
      "id":"1",
      "Name": "Ankit",
      "Sal": "1000",
    },
    {
      "id":"2",
      "Name": "Faizy",
```

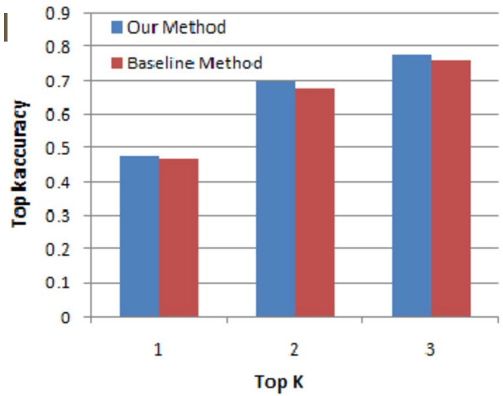
XML

```
<?xml version="1.0"?>
<contact-info>
<name>Ankit</name>
<company>Anlytics Vidhya</company>
<phone>+9187654321</phone>
</contact-info>
```

3 - Determine a baseline

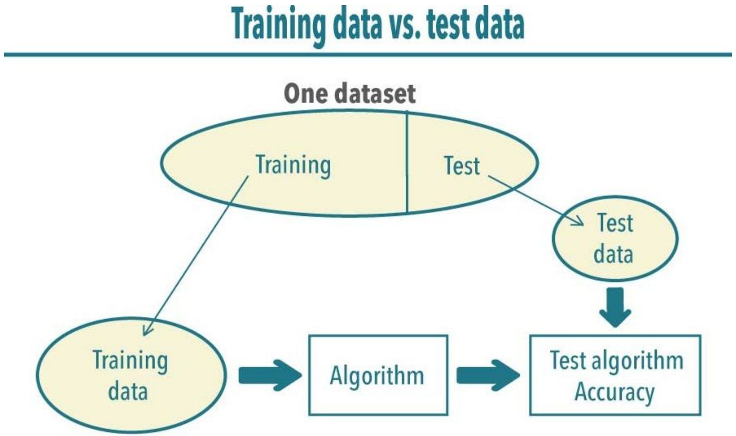


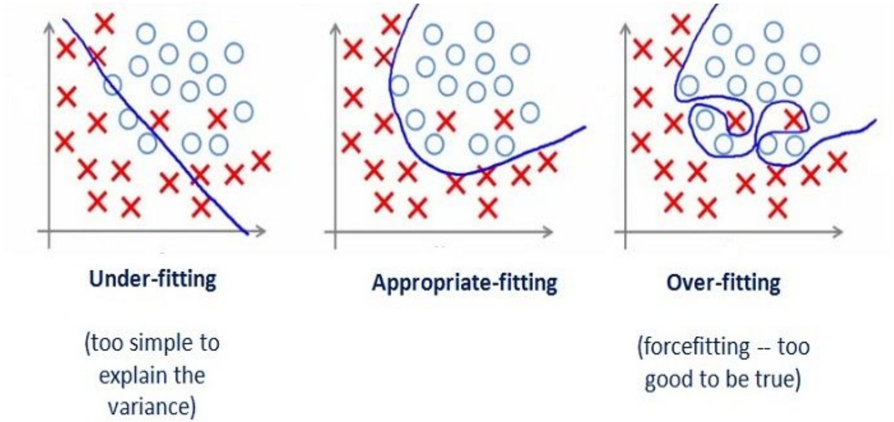
- Decide on a metric
  - Accuracy is most common
- Determine a baseline
  - Performance of models is meaningless without a point of comparison
    - Is 90% accuracy good?
  - Try the simplest possible model
    - Example: majority baseline - guess the most likely answer
      - If 80% of emails are junk mail, always guess junk mail
- An ML model is only useful if it beats the baseline



4 - Improve results

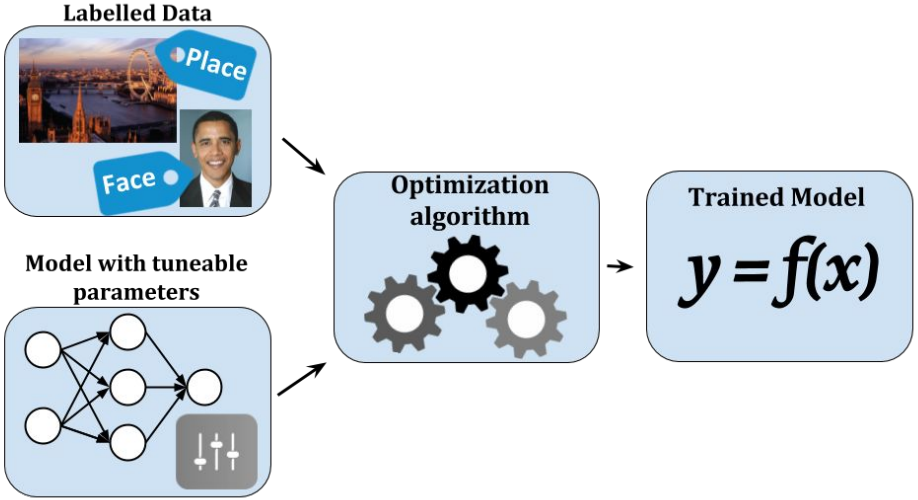
- Try more advanced models and compare to the baseline
- Select the model with the best **test** performance
  - Importance of training vs. testing
- Often a relatively simple model works well enough
  - Improvement with a more advanced model may not be worth the effort

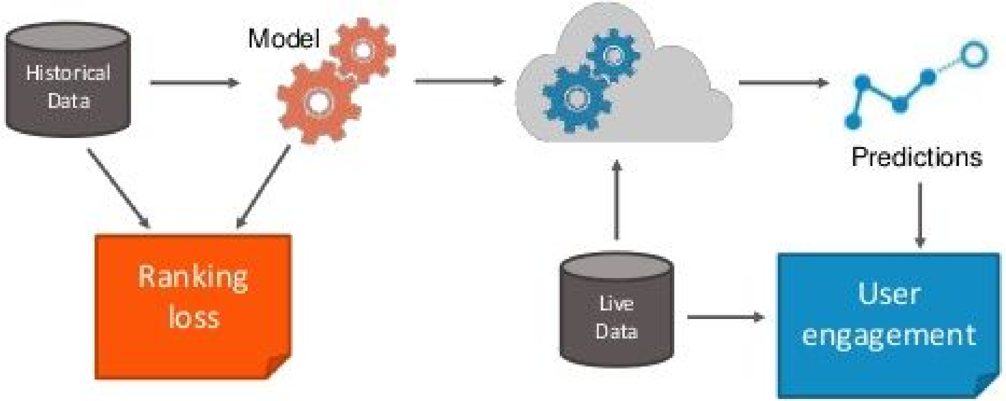




5 - Implement the model

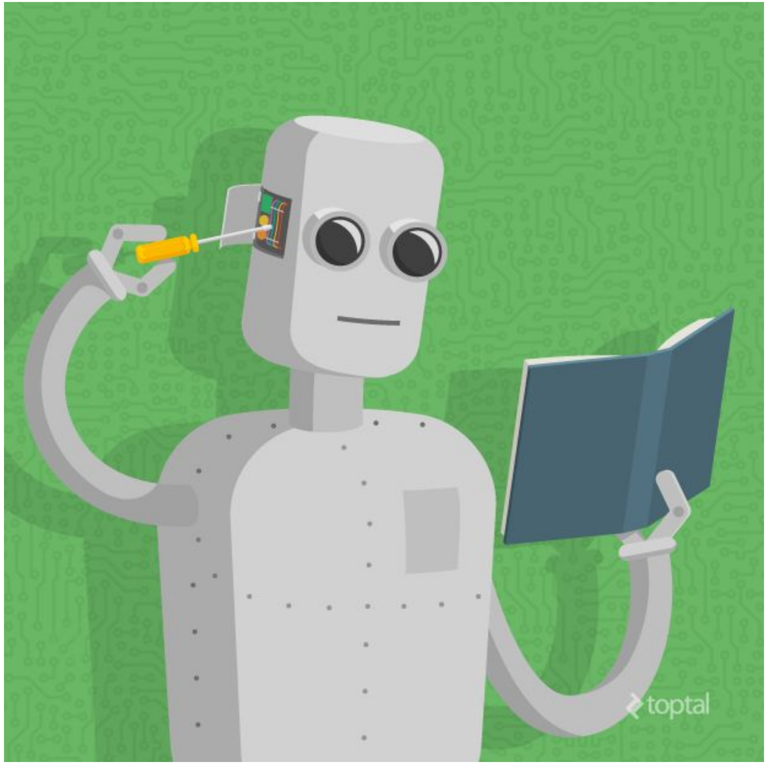
- Build the model
    - Select the best model architecture
    - Train the model on the data
    - Save the best parameters
- Insert the trained model into the workflow
- Maintain the model
    - Monitor model predictions
    - Periodically retrain with new data
- Enjoy the benefits!





Summary

- What is machine learning?
- History
- Types of machine learning problems
- Methods of learning
- Example: Decision trees
- Machine learning applications
- How to apply machine learning



Next Saturday - Deep Learning

- What is deep learning?
- How has deep learning revolutionized AI?
- Deep learning applications
- Case study: deep learning for medical image analysis