IMD1101
Machine Learning

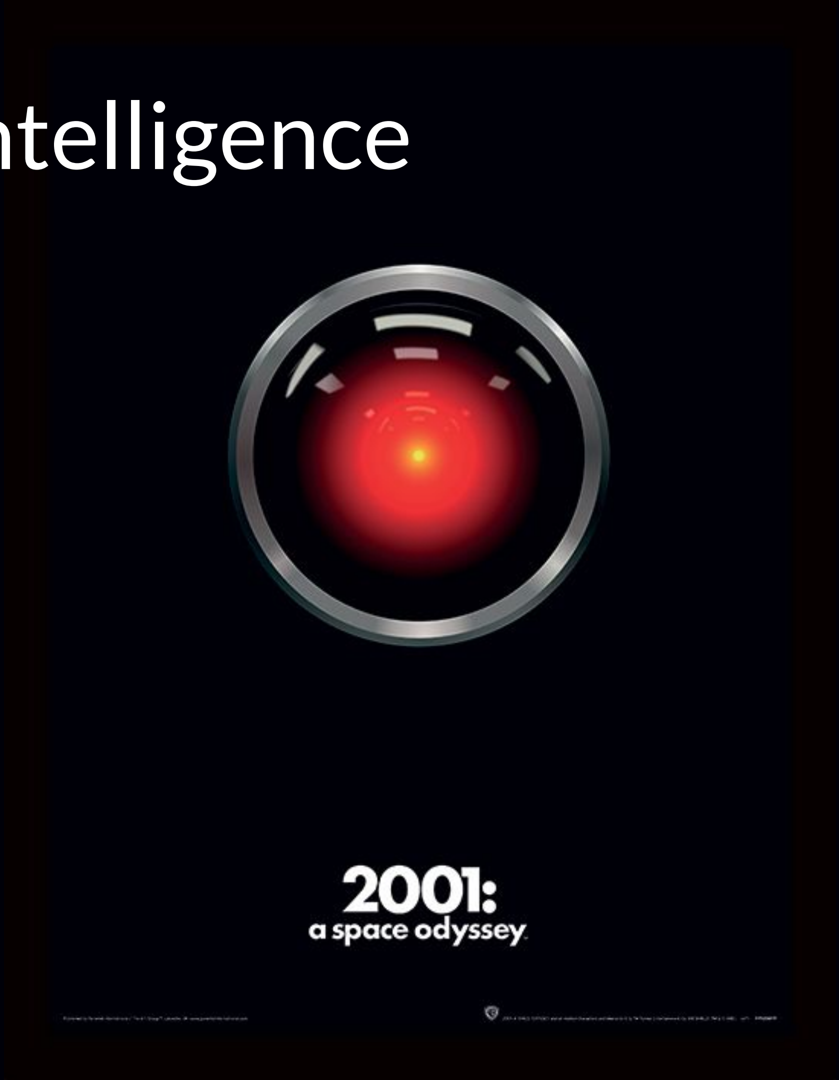
Lesson #03
Fundamentals of Machine Learning

notas

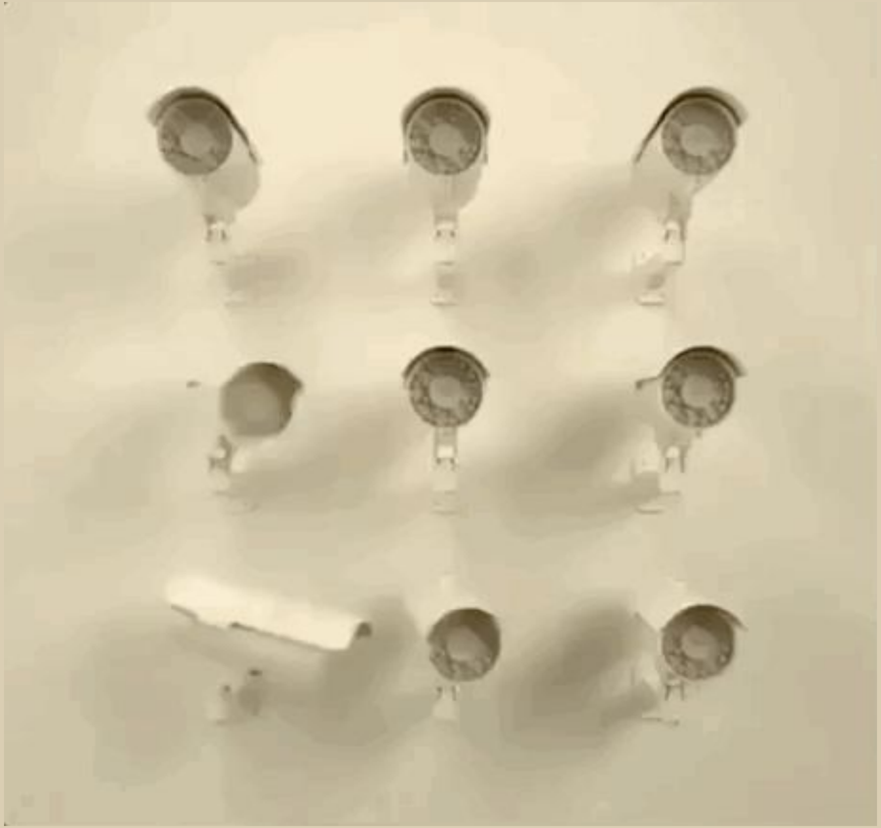
1. The Machine Learning Landscape
    a. What is ML?
    b. Types
    c. Main challenges
    d. Test & Validating

notas

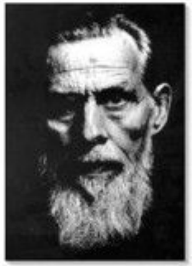
Artificial Intelligence




https://en.wikipedia.org/wiki/List_of_artificial_intelligence_films



World War II (1939 - 1945)




1943

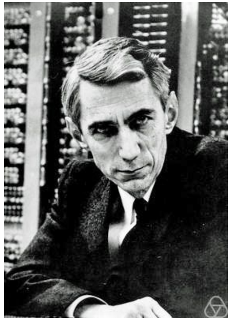
A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY

WARREN S. McCULLOCH and WALTER H. PITTS

Because of the "all-or-none" character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.

INTRODUCTION

THEORETICAL neurophysiology rests on certain cardinal assumptions. The nervous system is a net of neurons, each having a



1950

Philosophical Magazine, Ser.7, Vol. 41, No. 314 - March 1950.

XXII. Programming a Computer for Playing Chess[1]

By CLAUDE E. SHANNON

Bell Telephone Laboratories, Inc., Murray Hill, N.J.[2]

[Received November 8, 1949]

1. INTRODUCTION

This paper is concerned with the problem of constructing a computing routine or "program" for a modern general purpose computer which will enable it to play chess. Although perhaps of no practical importance, the question is of theoretical interest, and it is hoped that a satisfactory solution of this problem will act as a wedge in attacking other problems of a similar nature and of greater significance. Some possibilities in this direction are: -

(1)Machines for designing filters, equalizers, etc.
(2)Machines for designing relay and switching circuits.
(3)Machines which will handle routing of telephone calls based on the individual circumstances rather than by fixed patterns.
(4)Machines for performing symbolic (non-numerical) mathematical operations.
(5)Machines capable of translating from one language to another.
(6)Machines for making strategic decisions in simplified military operations.
(7)Machines capable of orchestrating a melody.
(8)Machines capable of logical deduction.

1950

A. M. Turing (1950) Computing Machinery and Intelligence. *Mind 49:* 433-460.

COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

1. The Imitation Game

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous, If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

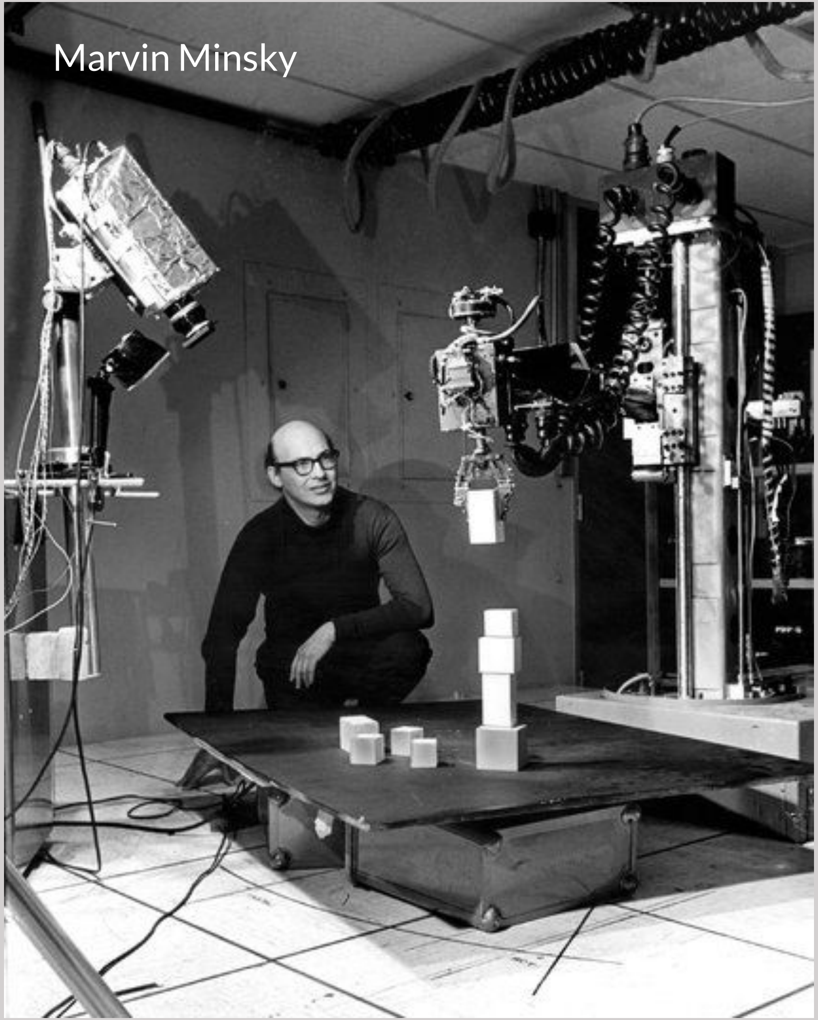
The new form of the problem can be described in terms of a game which we call the 'imitation game." It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart front the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A." The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

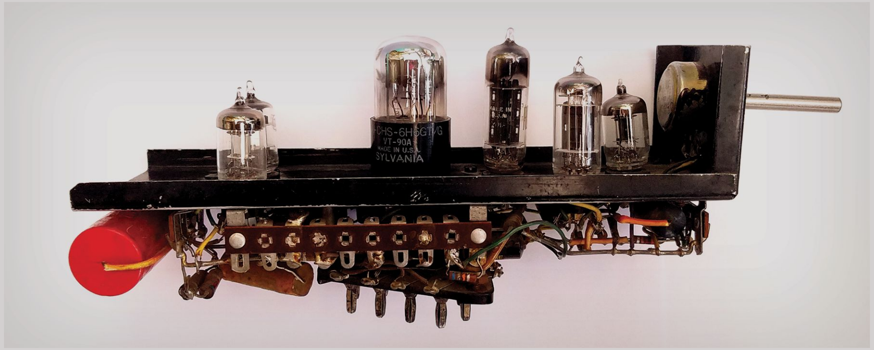
Now suppose X is actually A, then A must answer. It is A's object in the game to try and

Turing test

During the Turing test, the human questioner asks a series of questions to both respondents. After the specified time, the questioner tries to decide which terminal is operated by the human respondent and which terminal is operated by the computer.



Marvin Minsky

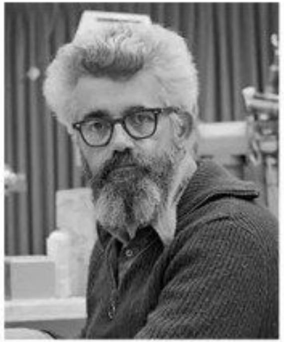


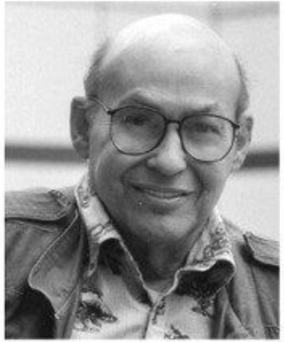
Stochastic Neural Analog Reinforcement Calculator (SNARC) 1951

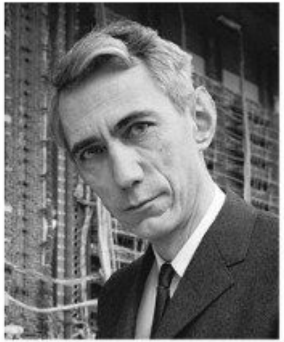
Arthur Samuel (1952)

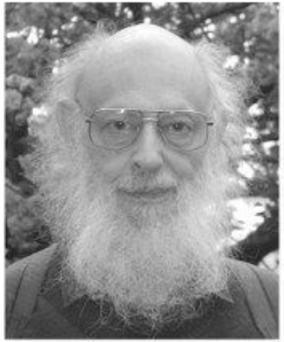
The Samuel Checkers playing Program was among the world's first successful self-learning programs

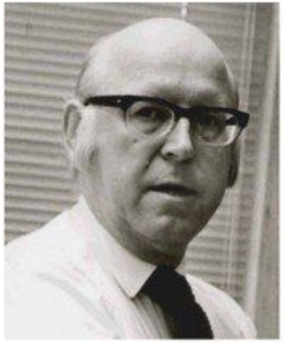
1956 Dartmouth Conference: The Founding Fathers of AI


John MacCarthy


Marvin Minsky

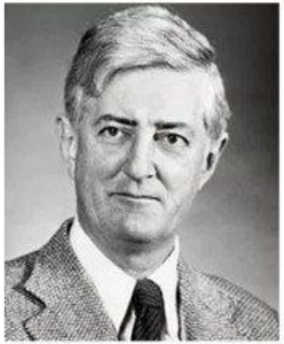

Claude Shannon


Ray Solomonoff


Alan Newell


Herbert Simon


Arthur Samuel

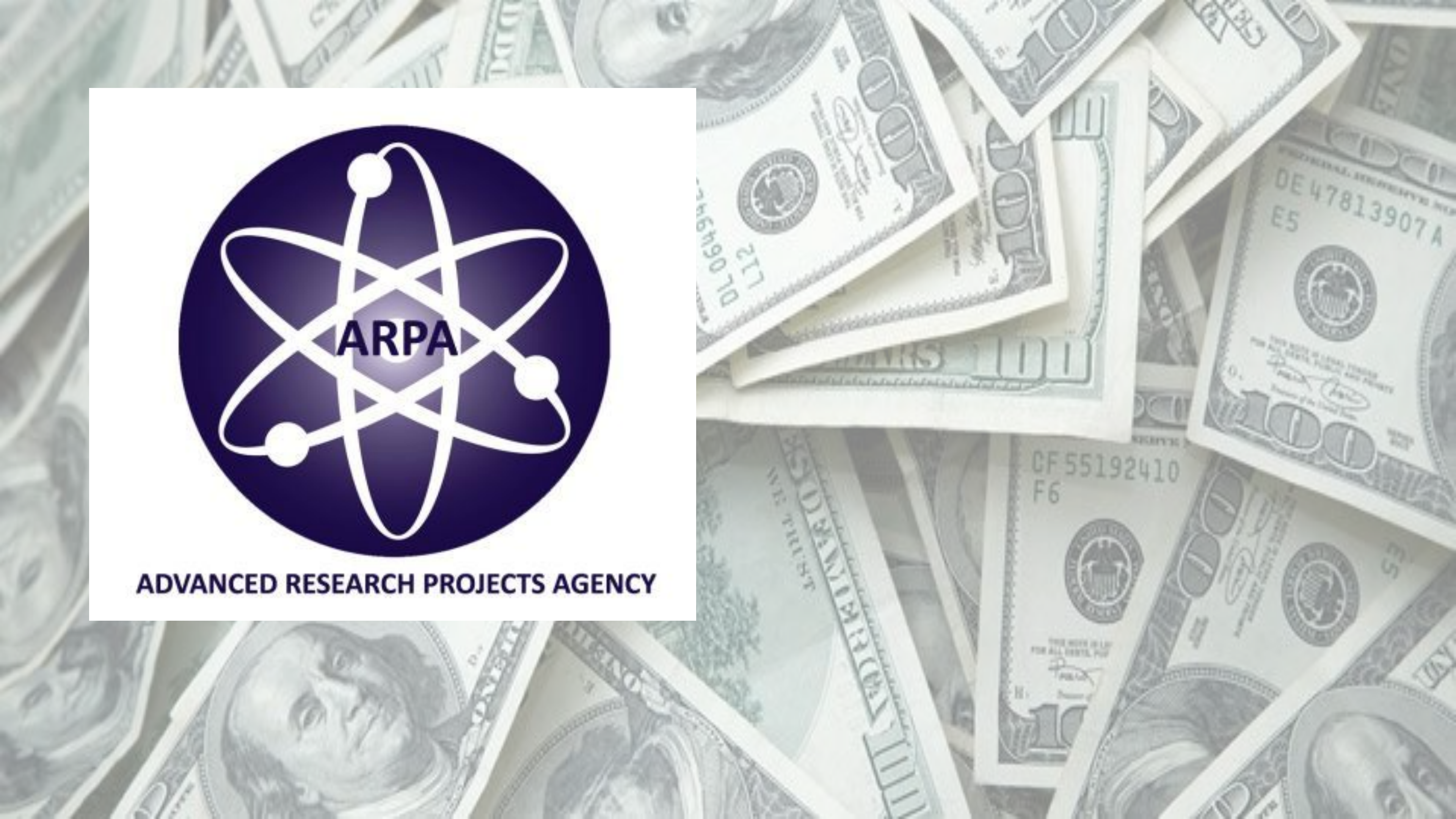

Oliver Selfridge


Nathaniel Rochester


Trenchard More

(2006) From left: Trenchard More, John McCarthy, Marvin Minsky, Oliver Selfridge, and Ray Solomonoff.

Dartmouth Conference, 1956


ARPA
ADVANCED RESEARCH PROJECTS AGENCY

1958

Psychological Review
Vol. 65, No. 6, 1958

THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN[1]

F. ROSENBLATT

Cornell Aeronautical Laboratory

If we are eventually to understand the capability of higher organisms for perceptual recognition, generalization, recall, and thinking, we must first have answers to three fundamental questions:

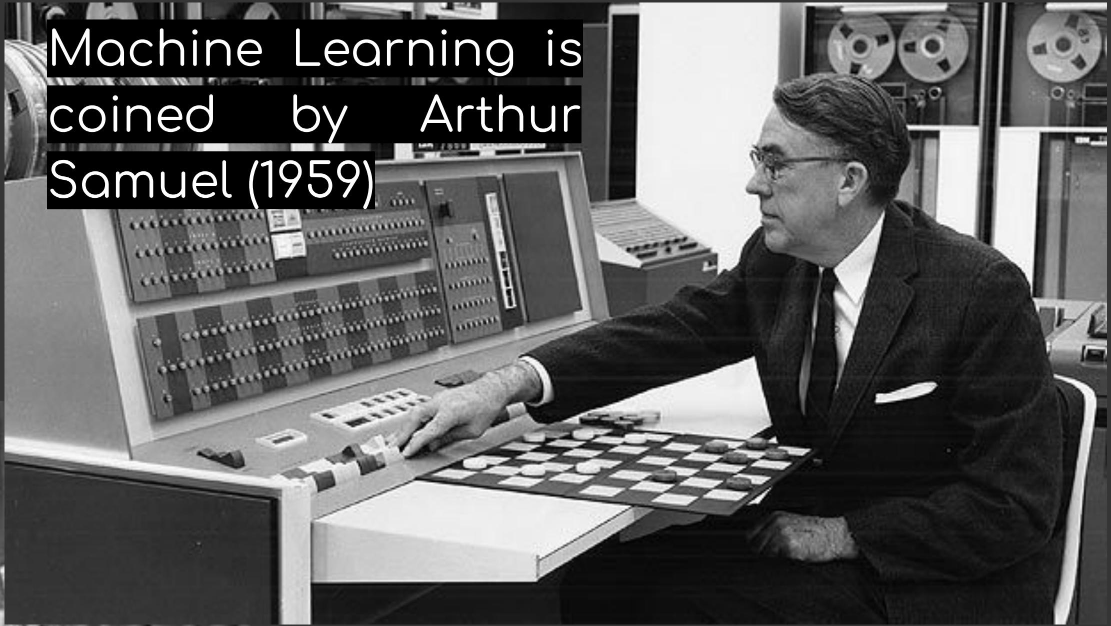
1. How is information about the physical world sensed, or detected, by the biological system?
2. In what form is information stored, or remembered?
3. How does information contained in storage, or in memory, influence recognition and behavior?

The first of these questions is in the province of sensory physiology, and is the only one for which appreciable understanding has been achieved. This article will be concerned primarily with the second and third questions, which are still subject to a vast amount of speculation, and where the few relevant facts currently supplied by neurophysiology have not yet been integrated into an acceptable theory.

With regard to the second question, two alternative positions have been maintained. The first suggests that

and the stored pattern. According to this hypothesis, if one understood the code or "wiring diagram" of the nervous system, one should, in principle, be able to discover exactly what an organism remembers by reconstructing the original sensory patterns from the "memory traces" which they have left, much as we might develop a photographic negative, or translate the pattern of electrical charges in the "memory" of a digital computer. This hypothesis is appealing in its simplicity and ready intelligibility, and a large family of theoretical brain models has been developed around the idea of a coded, representational memory (**2**, **3**, **9**, **14**). The alternative approach, which stems from the tradition of British empiricism, hazards the guess that the images of stimuli may never really be recorded at all, and that the central nervous system simply acts as an intricate switching network, where retention takes the form of new connections, or pathways, between centers of activity. In many of the more recent developments of this position (Hebb's "cell assembly," and Hull's "cortical anticipatory goal response," for example) the "re-

LISP: The language that made AI possible



"Machine Learning: field of study that gives computers the ability to learn without being explicitly programmed"

- 1960
  - one Defense computer mistakenly identified the moon as an incoming missile which understandably caused great consternation
- 1961
  - IBM declared separate departments for software and hardware, meaning pure programmers officially would have a declared place to develop programs and environments
  - In the academic sphere, universities began granting the first degrees in Computer Science.
- 1962
  - during the Cuban Missile crisis, when communications were blocked for several days.
- 1963
  - DARPA funding in 1963, MIT created a new research group Project MAC (Mathematics and Computation) - 2M U$

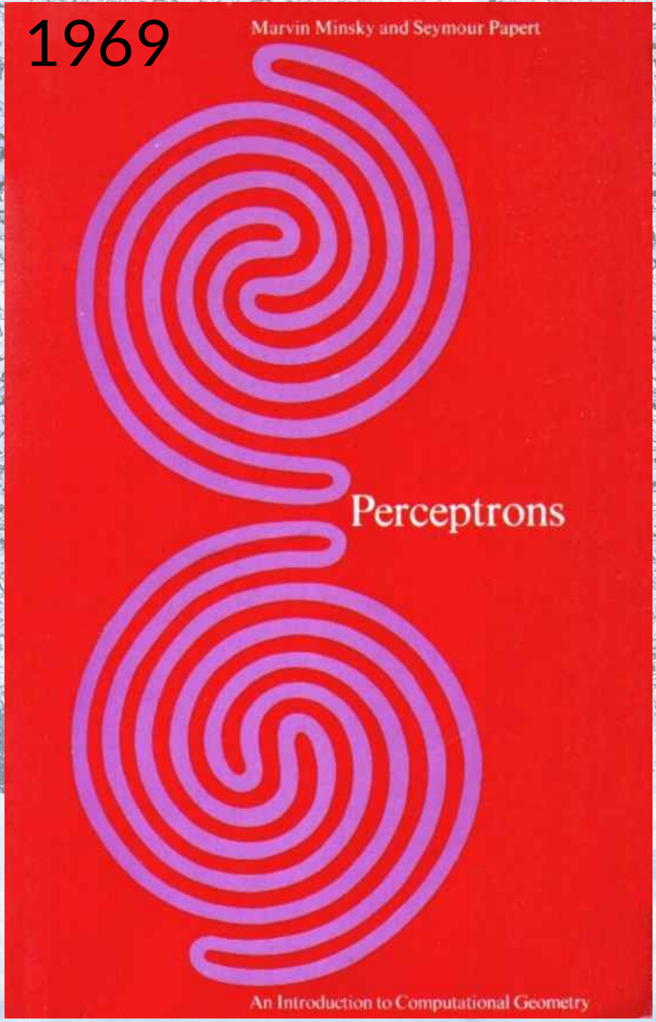
```
  Welcome to
                    EEEEEE  LL      IIII  ZZZZZZZ   AAAAA
                    EE      LL       II       ZZ   AA   AA
                    EEEEE   LL       II     ZZZ    AAAAAAA
                    EE      LL       II    ZZ      AA   AA
                    EEEEEE  LLLLLL  IIII  ZZZZZZZ  AA   AA

  Eliza is a mock Rogerian psychotherapist.
  The original program was described by Joseph Weizenbaum in 1966.
  This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

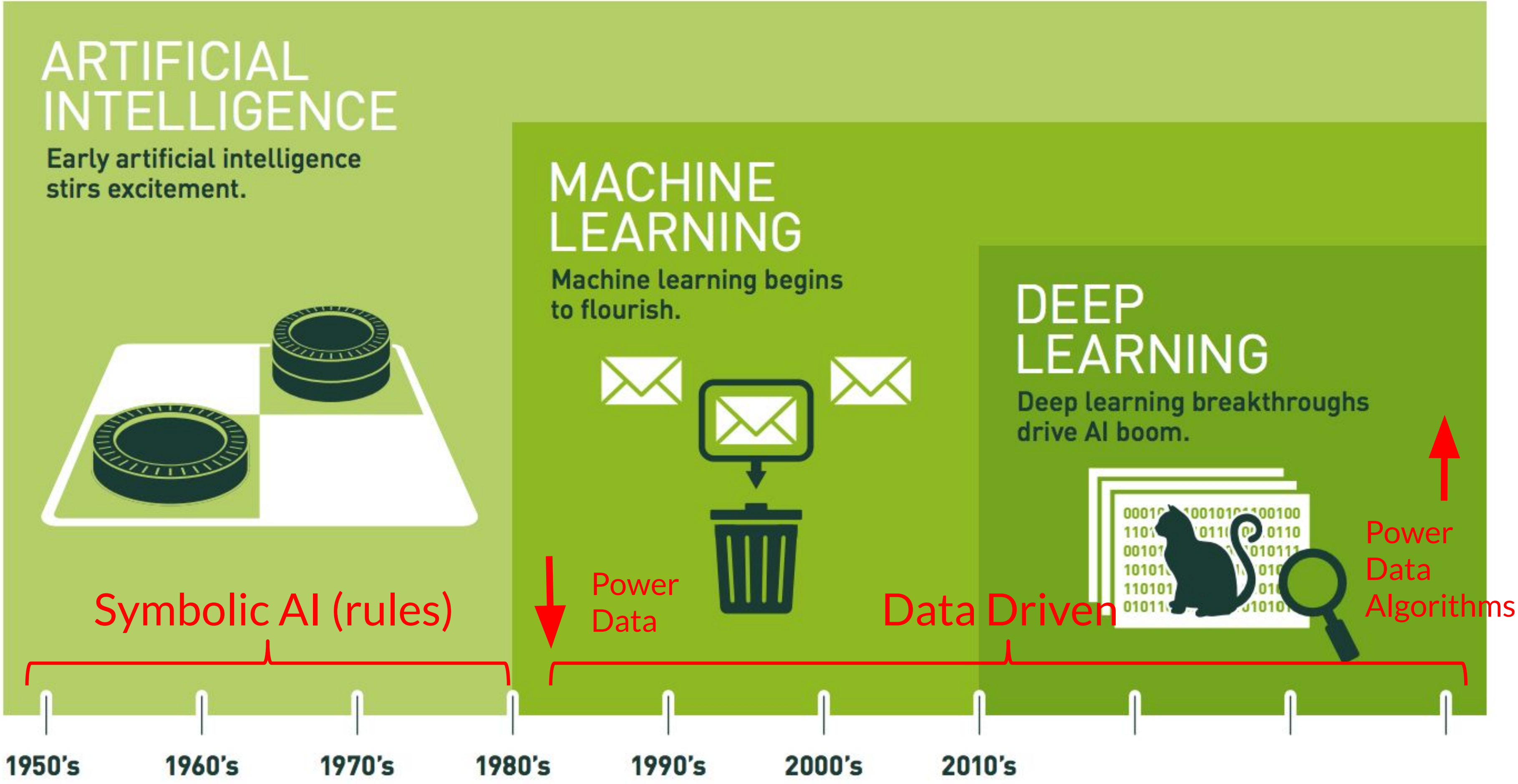
First chabot - Eliza (1964, Joseph Weizenbaum, MIT AI Lab)

1969


Marvin Minsky and Seymour Papert
Perceptrons
An Introduction to Computational Geometry

AI Winter
(1970-1980)

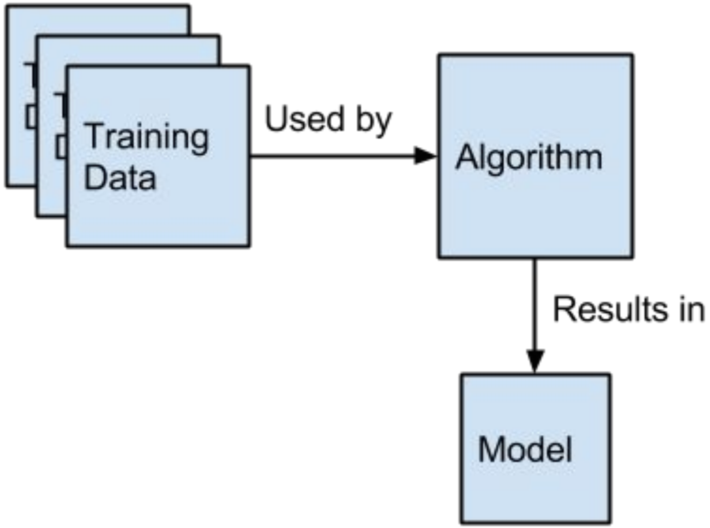
How do they related to each other?



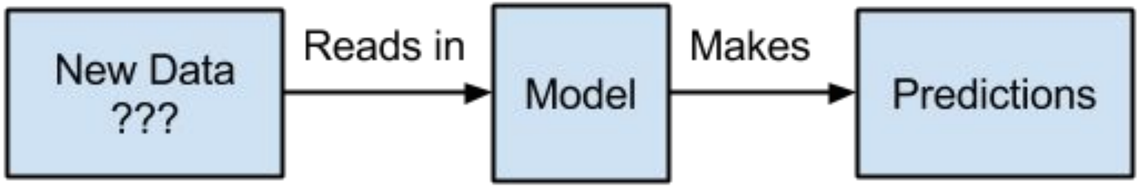

NVIDIA
280.00
260.00
240.00
220.00
200.00
174.82
160.00
140.00
120.00
100.00
80.00
60.00
40.00
20.00
0.00
−20.00
2006
2008
2010
2012
2014
2016
2018
2020

MACHINE LEARNING

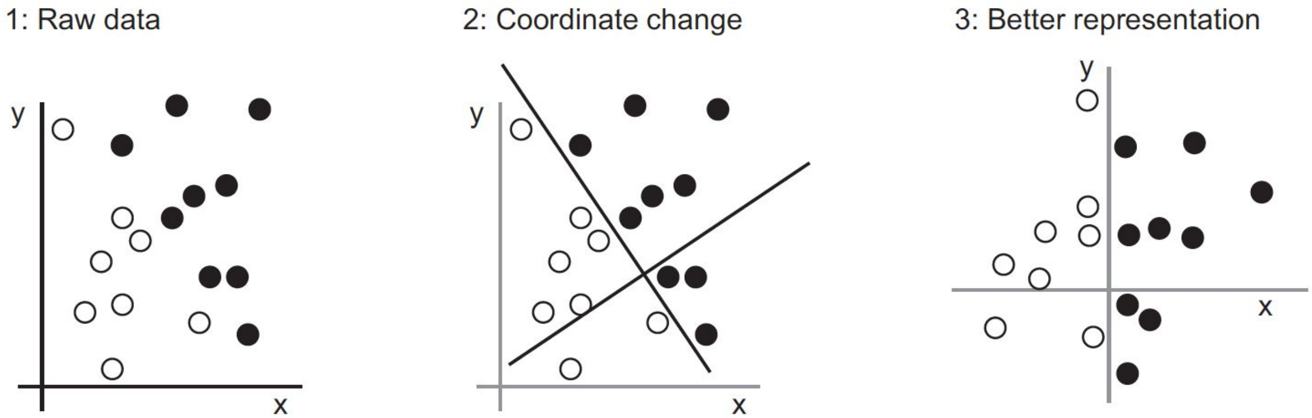
A NEW
PROGRAMMING
PARADIGM



The ability to acquire their own knowledge, by extracting patterns from raw data. This capability is known as **machine learning.**

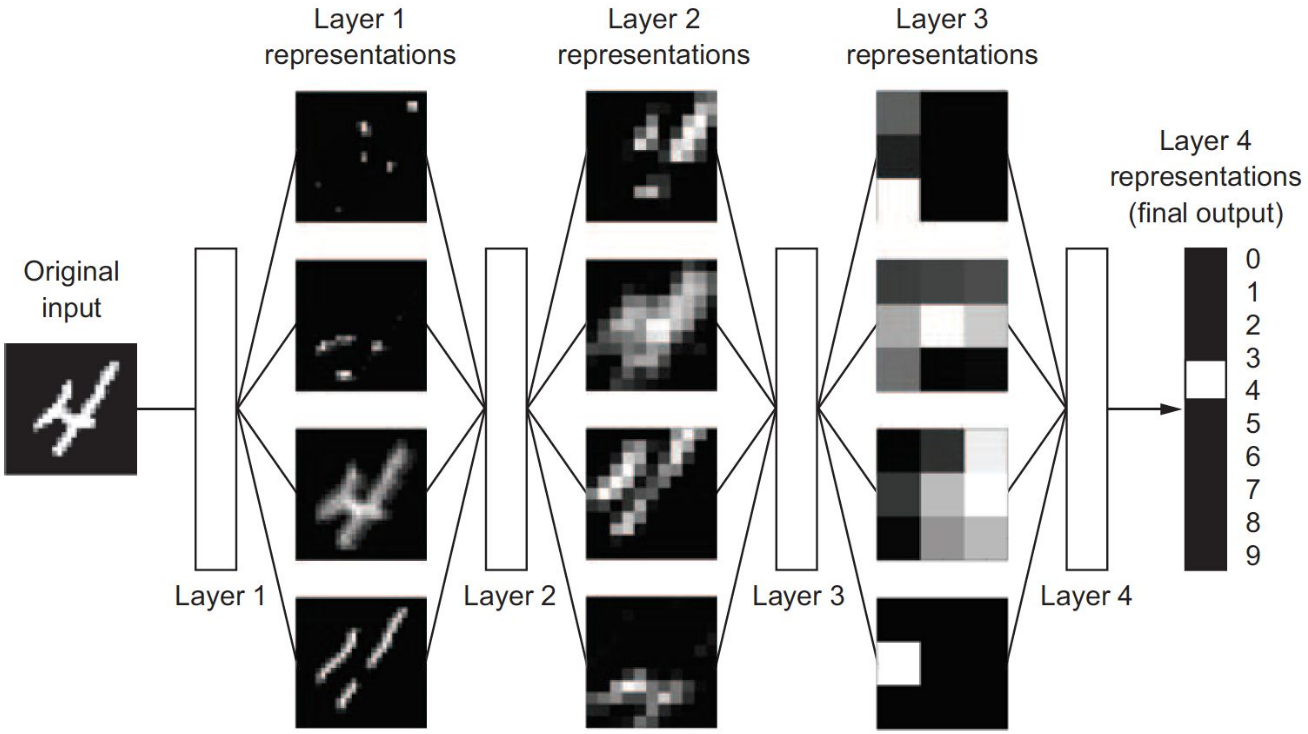


Machine Learning Definition



An automatic search process for better data representations

What is Deep Learning?

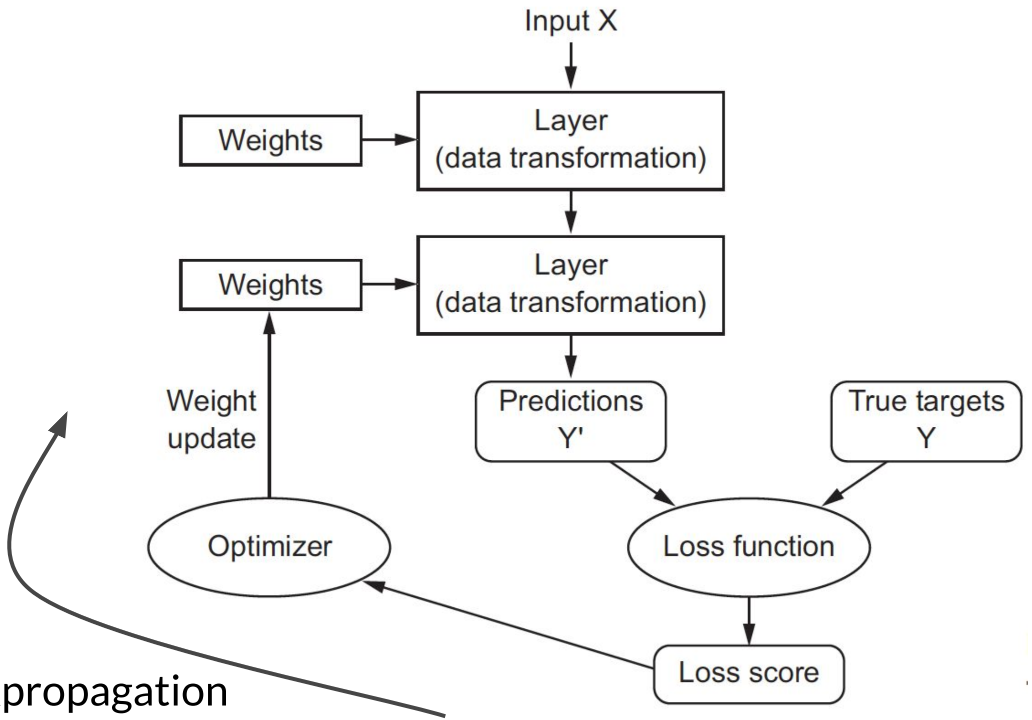


"Deep learning methods are representation-learning methods with multiple levels of representation, obtained by composing simple but nonlinear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level. [. . .] The key aspect of deep learning is that these layers are not designed by human engineers: they are learned from data using a general-purpose learning procedure"

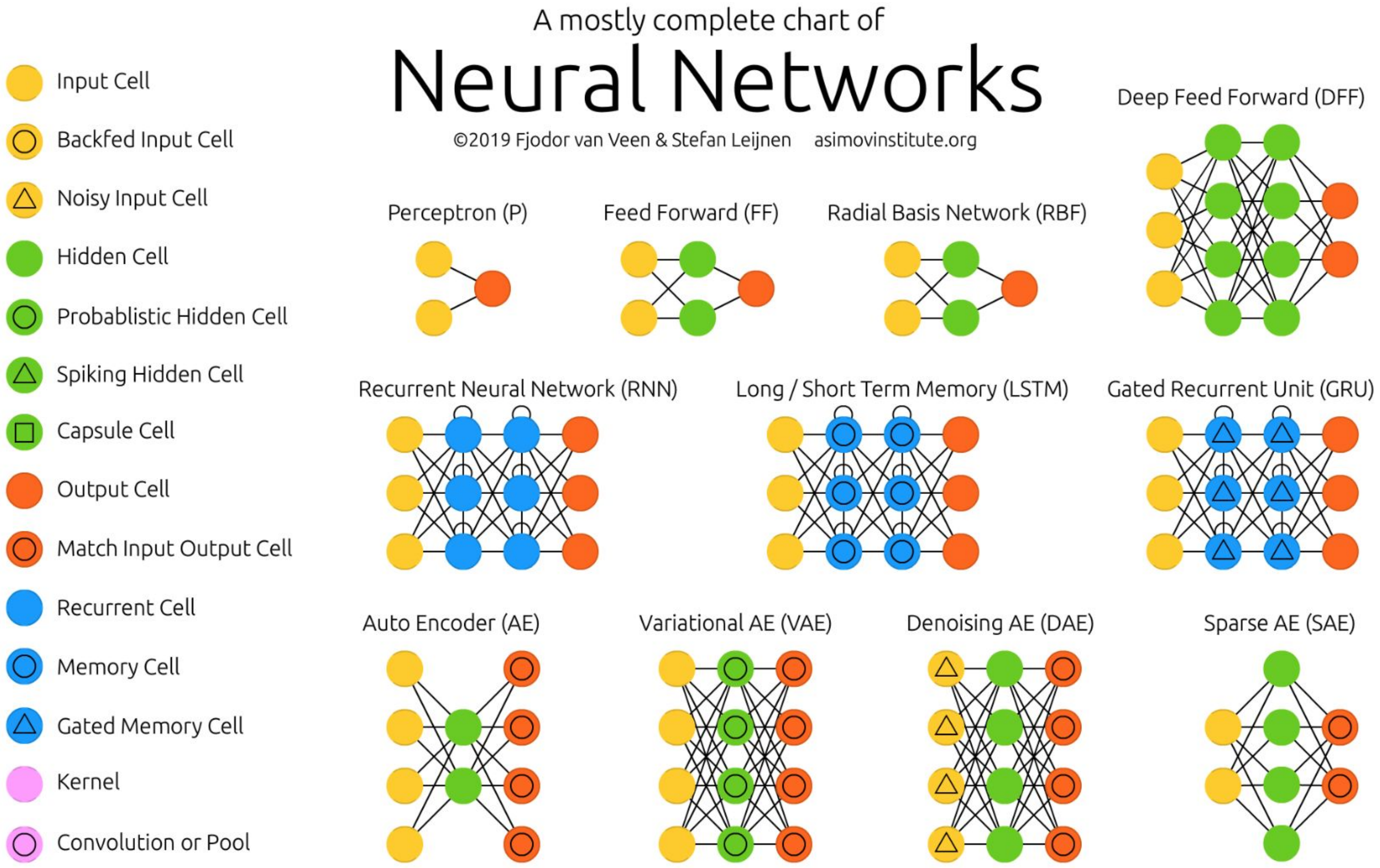
[Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, Nature 2015]

Understanding how DL works

Finding the right values of weights which minimize the error

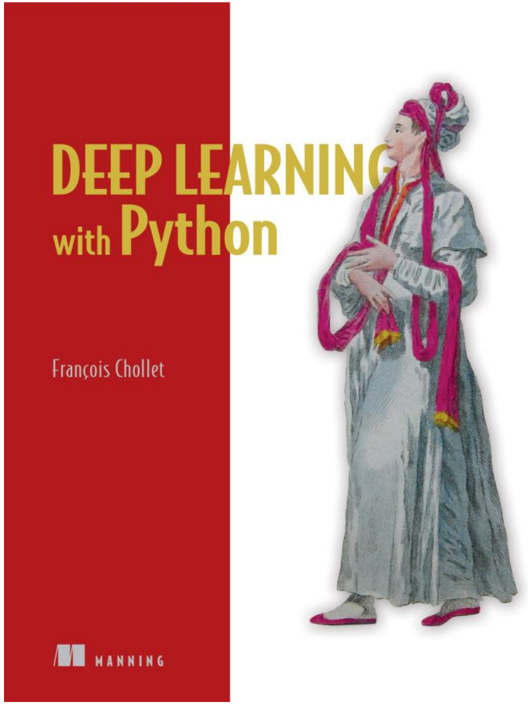


Backpropagation



http://www.asimovinstitute.org/neural-network-zoo/

A brief history of Machine Learning

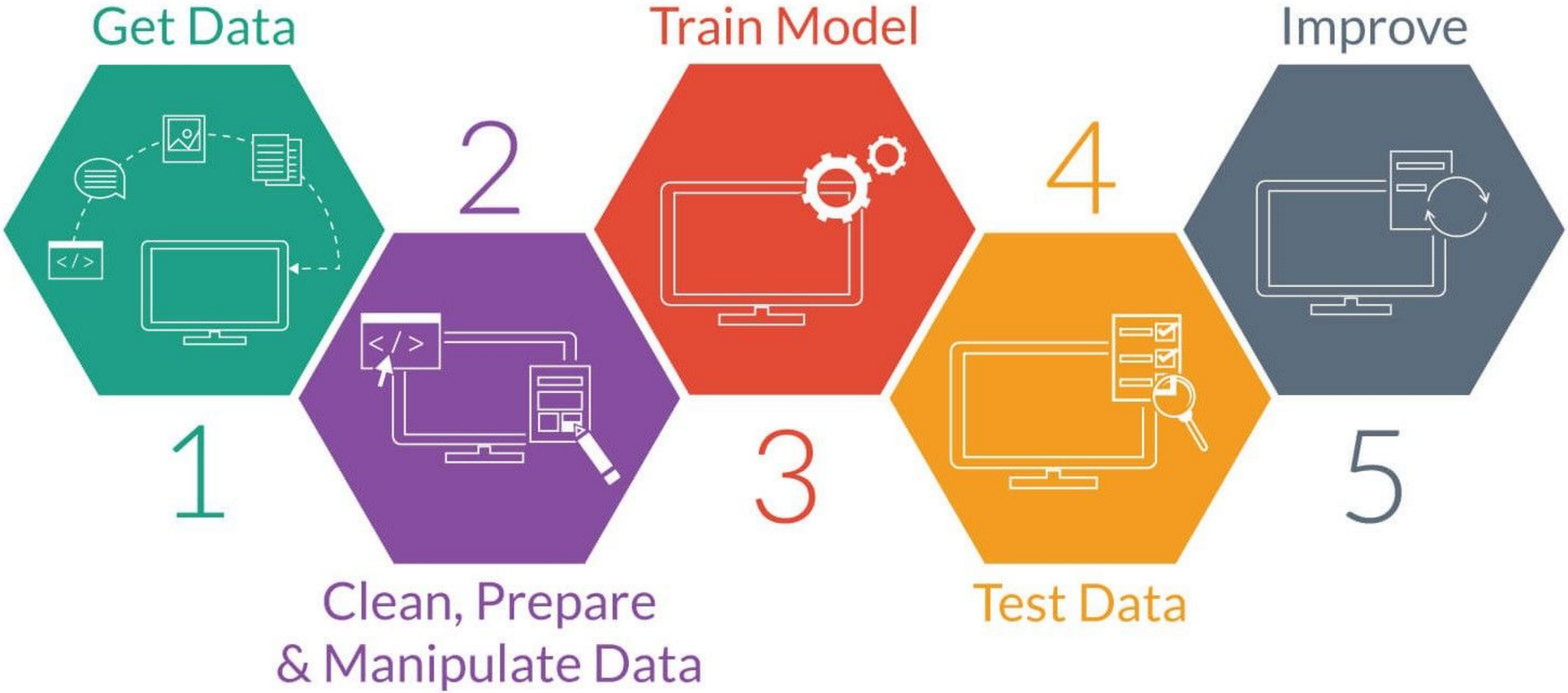


https://www.manning.com/books/deep-learning-with-python

Chapter 1, 2 and 3 are available for free!!!
Read the Chapter. 1, Section 1.2 !!!!!


It's time to
move on...

A general ML workflow



Types of Machine Learning Systems

- Whether or not they are trained with human supervision
  - Supervised
  - Unsupervised
  - Semi-supervised
  - Reinforcement learning
- Whether or not they can learn incrementally on the fly
  - Online
  - Batch learning
- Whether they work by simply comparing new data points to known data points, or instead detect patterns in the training data and build a predictive model
  - Instance-based learning
  - Model-based learning

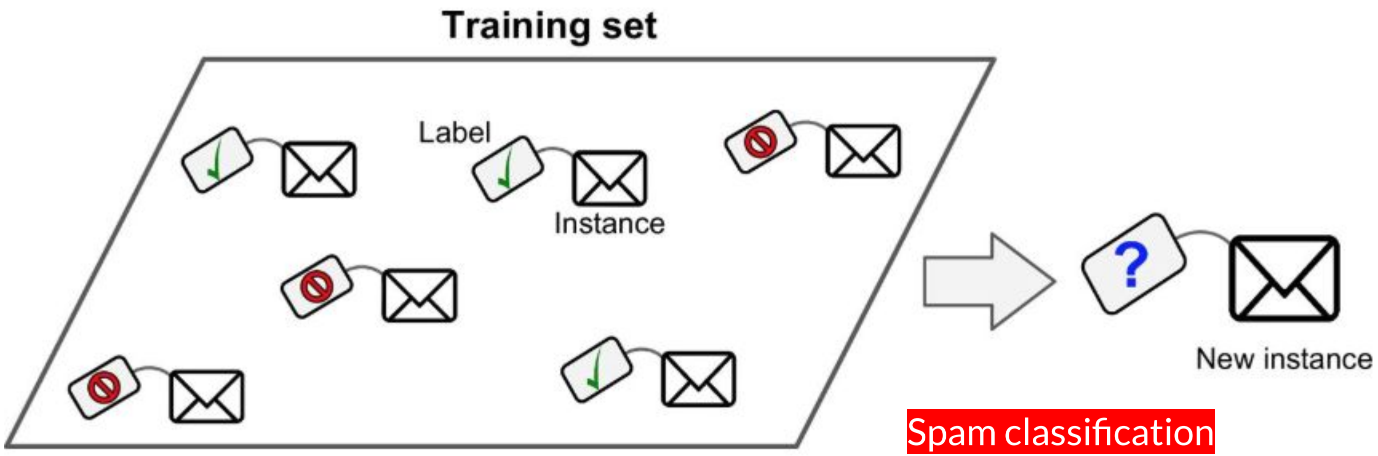
Types of Machine Learning Systems

- Whether or not they are trained with human supervision
  - Supervised
  - Unsupervised
  - Semi-supervised
  - Reinforcement learning
- Whether or not they can learn incrementally on the fly
  - Online
  - Batch learning
- Whether they work by simply comparing new data points to known data points, or instead detect patterns in the training data and build a predictive model
  - Instance-based learning
  - Model-based learning

SUPERVISED LEARNING

Supervised Learning

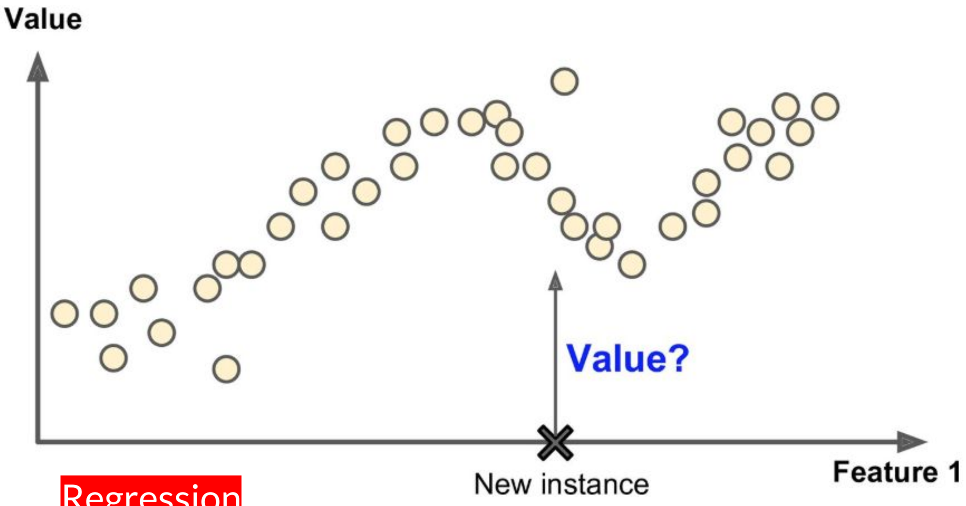
In supervised learning, the training data you feed to the algorithm includes the desired solutions, called labels.



Spam classification

Supervised Learning

Another typical task is to predict a target numeric value, such as the price of a car, given a set of features (mileage, age, brand, etc) called predictors.

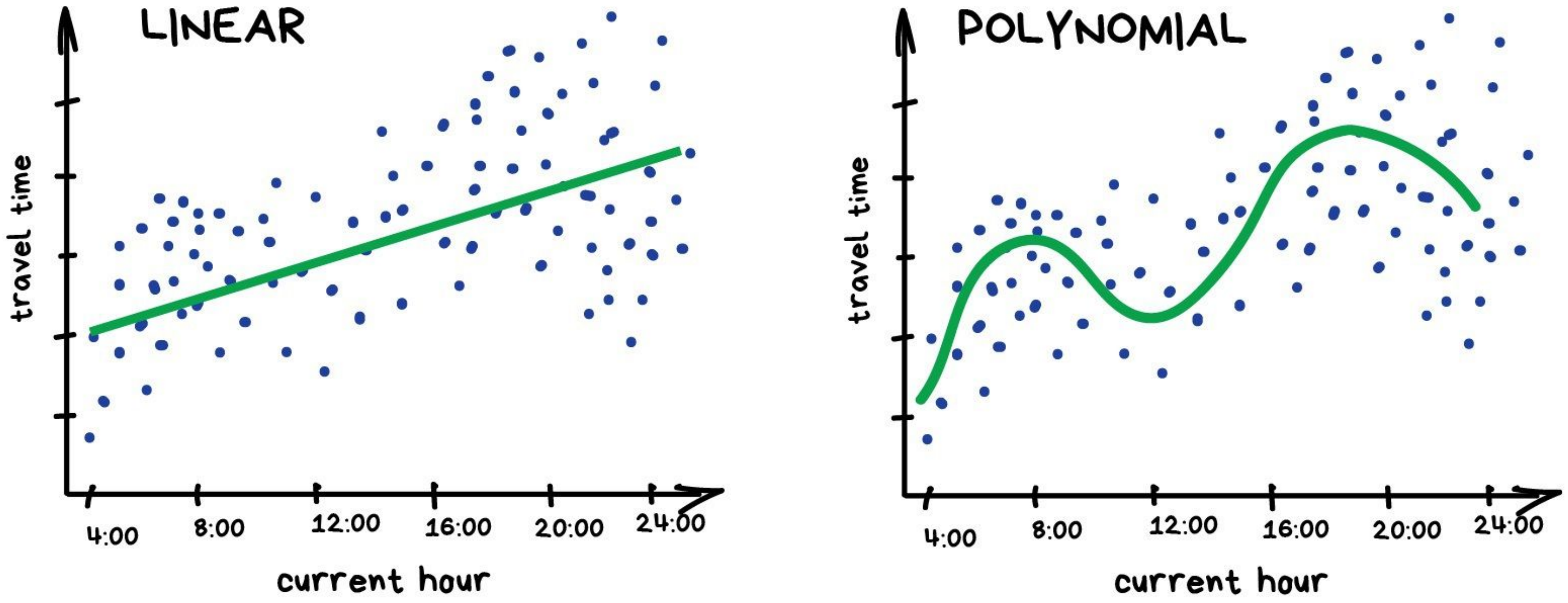


Regression

Supervised Learning

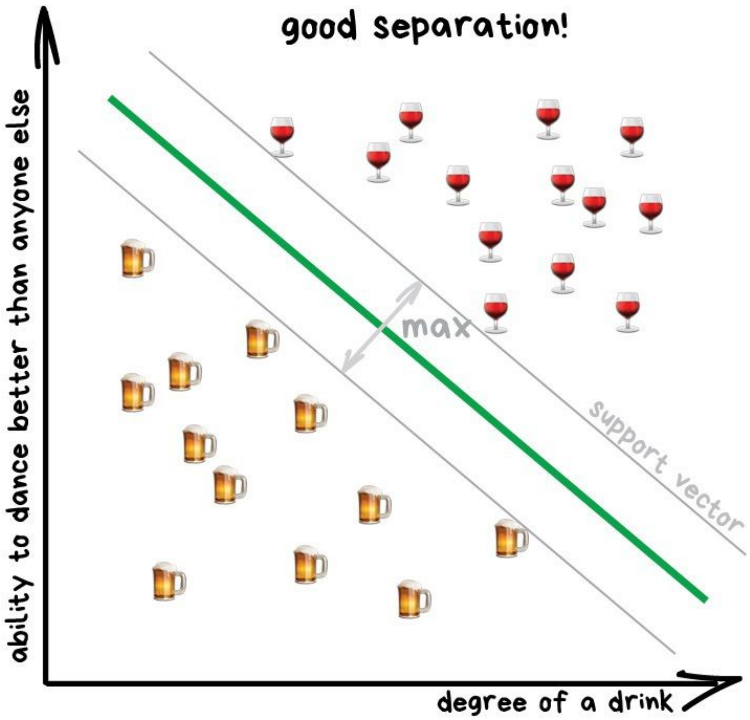
- K-Nearest Neighbors (KNN)
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVM)
- Decision Trees and Random Forests
- Neural Networks
- XGBoost
- Deep Learning
- **Ensemble

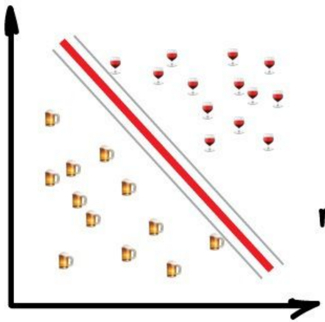
PREDICT TRAFFIC JAMS

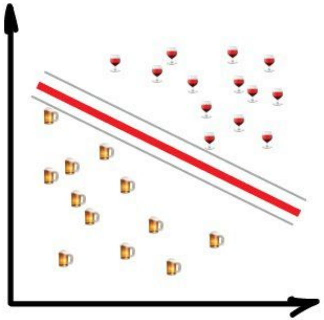


REGRESSION

SEPARATE TYPES OF ALCOHOL

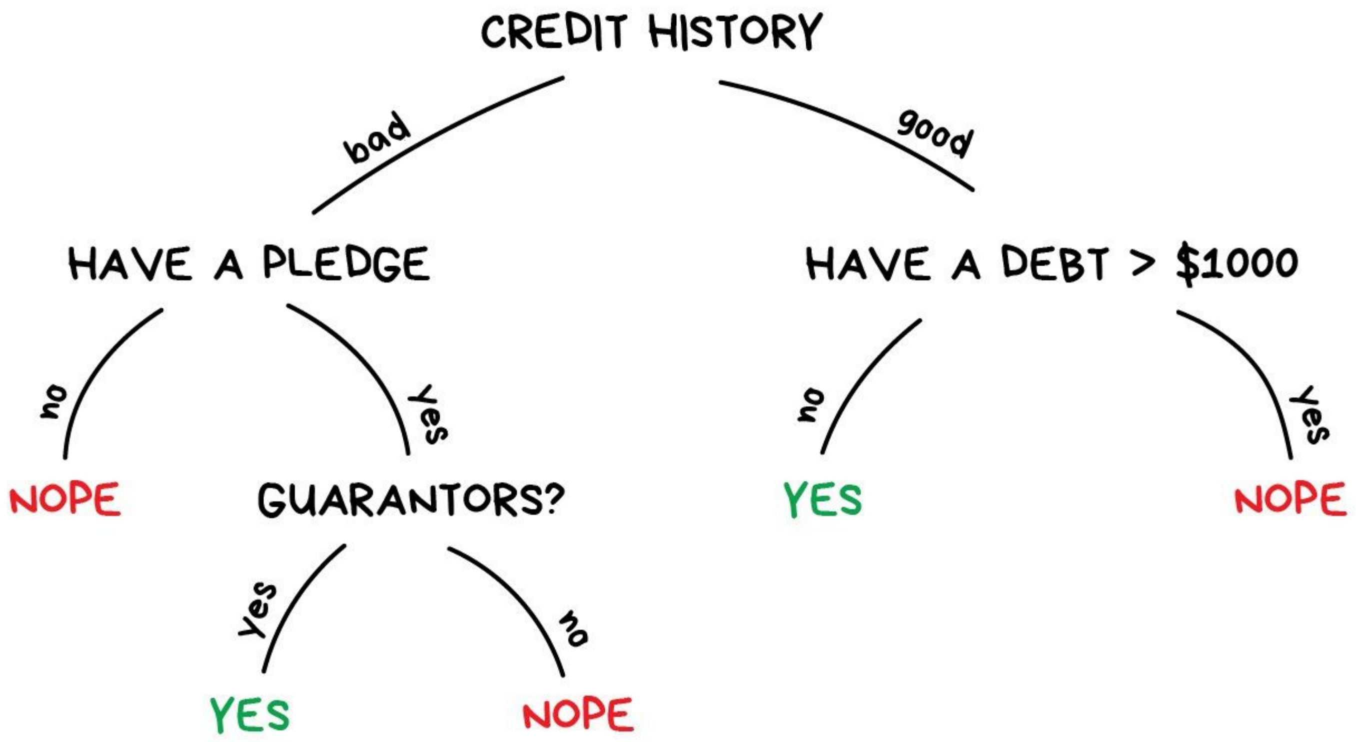




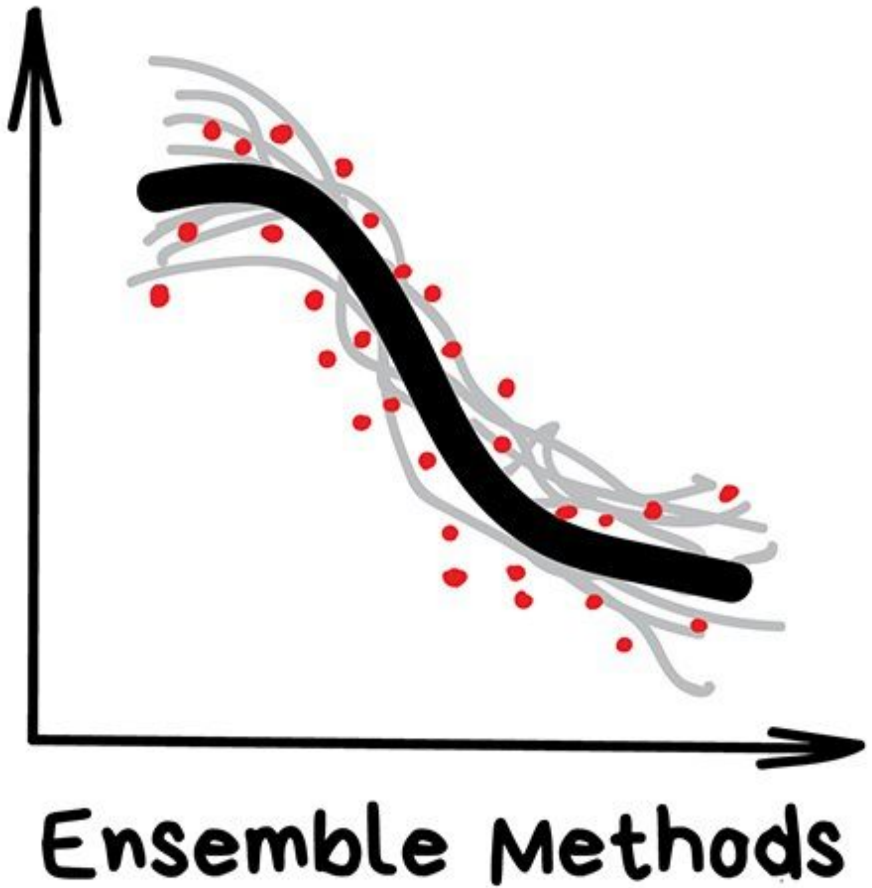


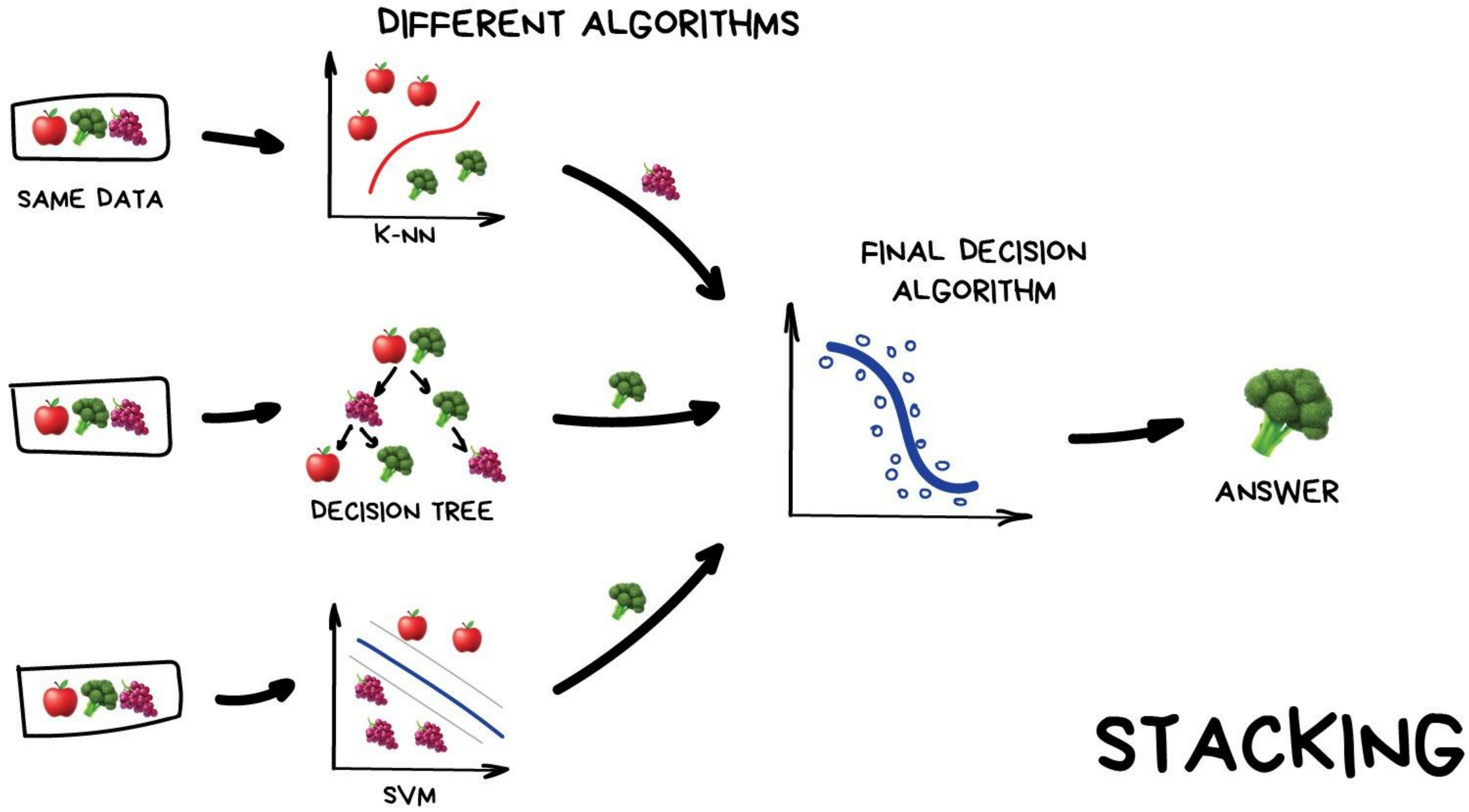
SUPPORT VECTOR MACHINE

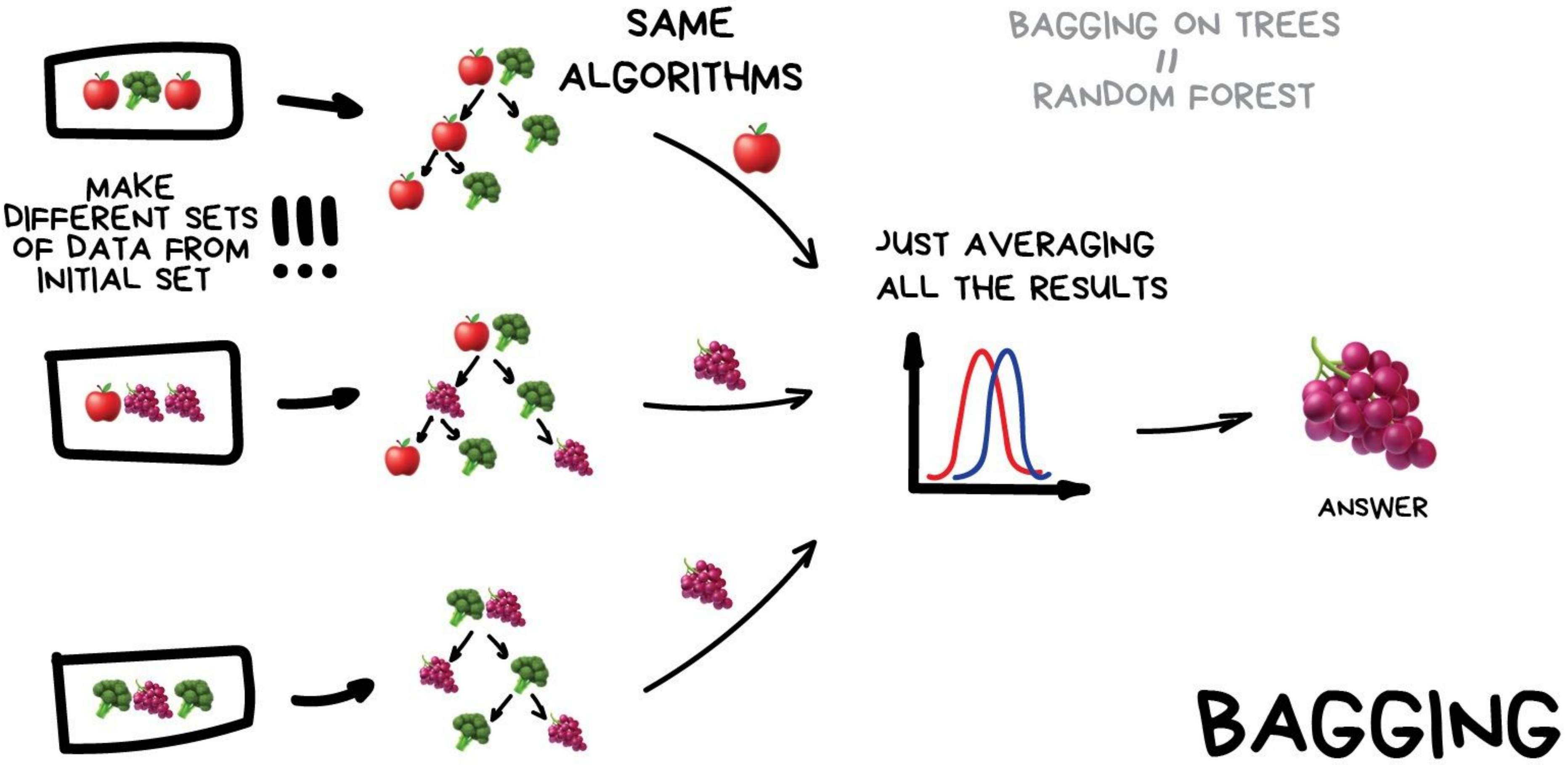
GIVE A LOAN?

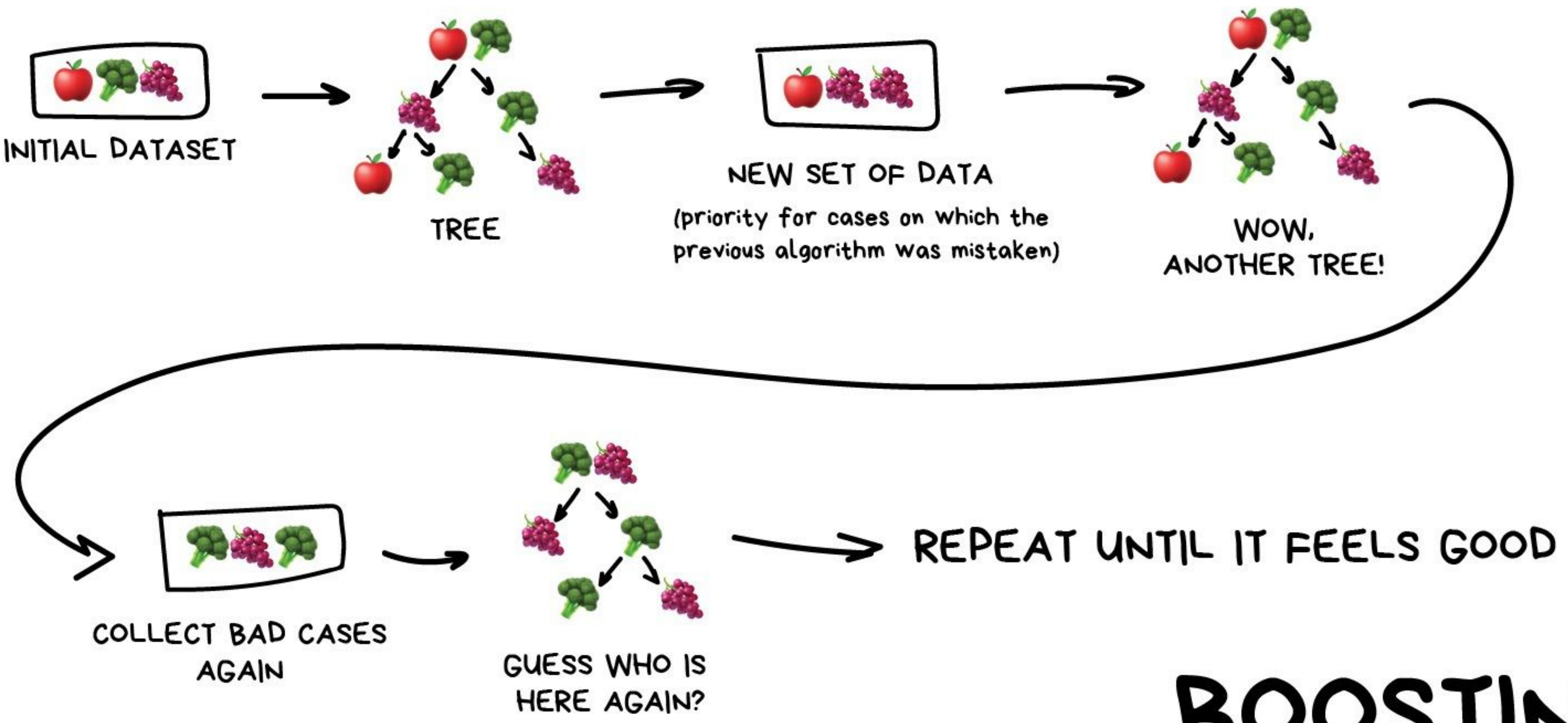


DECISION TREE


Ensemble Methods


DIFFERENT ALGORITHMS
SAME DATA
K-NN
DECISION TREE
SVM
FINAL DECISION ALGORITHM
ANSWER
STACKING


SAME ALGORITHMS
BAGGING ON TREES
=
RANDOM FOREST
MAKE DIFFERENT SETS OF DATA FROM INITIAL SET !!! ...
JUST AVERAGING ALL THE RESULTS
ANSWER
BAGGING

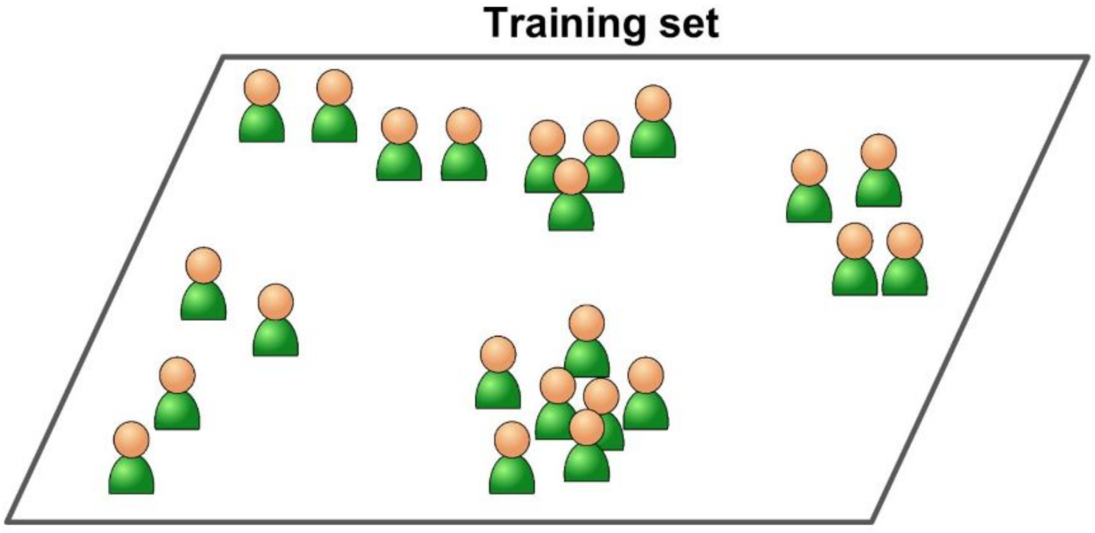

INITIAL DATASET
TREE
NEW SET OF DATA
(priority for cases on which the previous algorithm was mistaken)
WOW, ANOTHER TREE!
COLLECT BAD CASES AGAIN
GUESS WHO IS HERE AGAIN?
REPEAT UNTIL IT FEELS GOOD

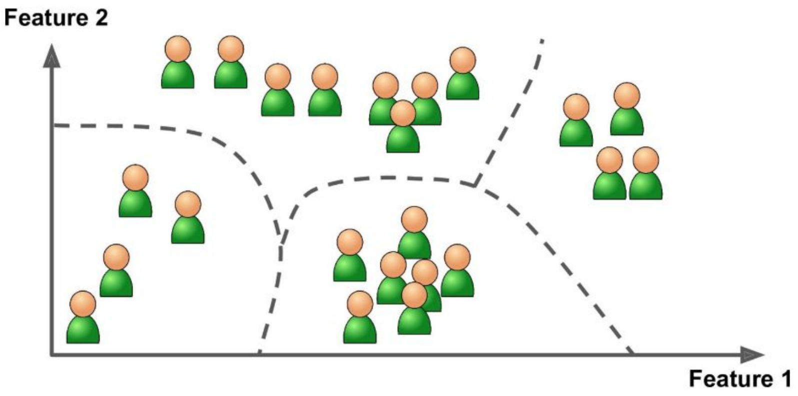
BOOSTING

UNSUPERVISED LEARNING

Unsupervised Learning

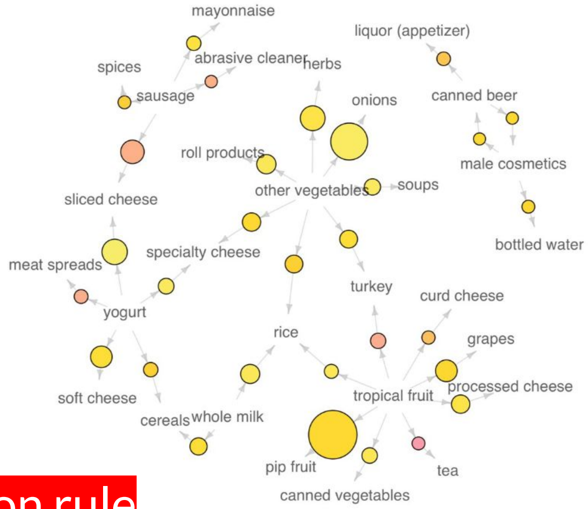
In unsupervised learning, as you might guess, the training data is unlabeled. The system tries to learn without a teacher.



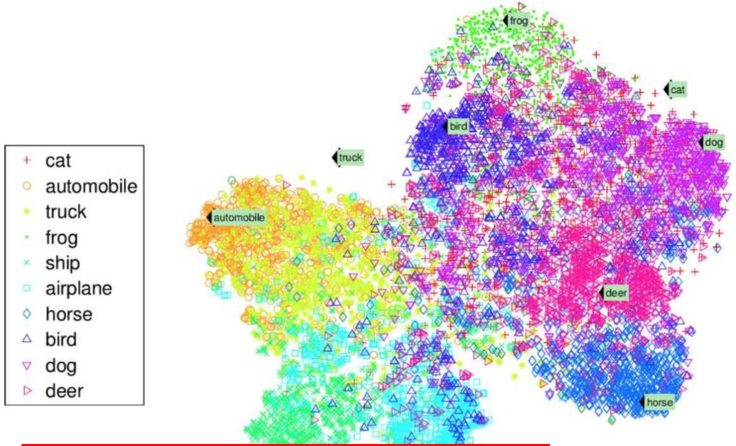


Clustering

Anomaly detection



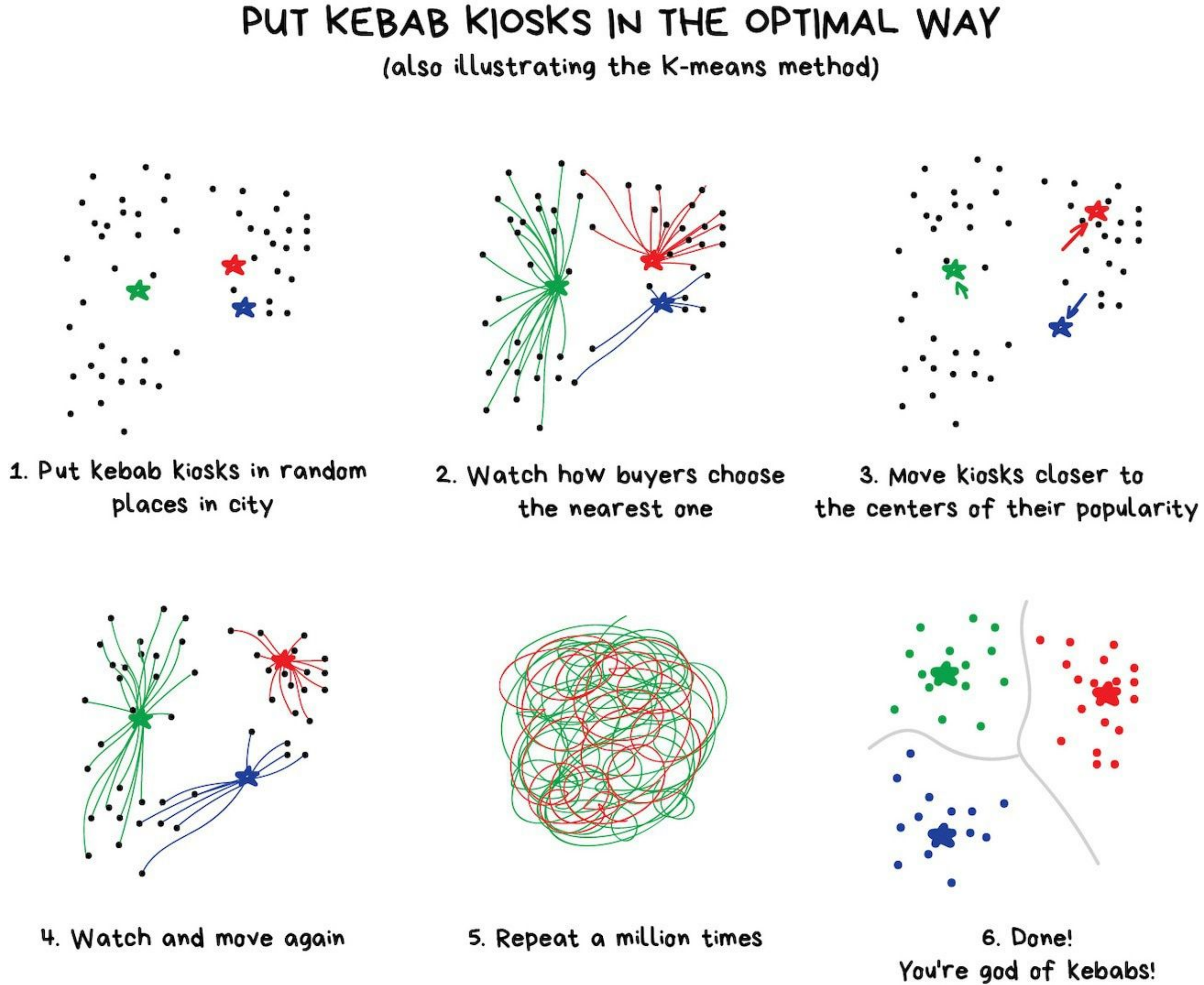
Association rule



Visualization highlighting

Unsupervised Learning

- Clustering
    - K-Means
    - Hierarchical Cluster Analysis (HCA)
    - Expectation Maximization
- Visualization and dimensionality reduction
    - Principal Component Analysis (PCA)
    - Kernel PCA
    - Locally-Linear Embedding (LLE)
    - T-distributed Stochastic Neighbor Embedding (t-SNE)
- Association rule learning
    - Apriori
    - Eclat


PUT KEBAB KIOSKS IN THE OPTIMAL WAY
(also illustrating the K-means method)
1. Put kebab kiosks in random places in city
2. Watch how buyers choose the nearest one
3. Move kiosks closer to the centers of their popularity
4. Watch and move again
5. Repeat a million times
6. Done! You're god of kebabs!

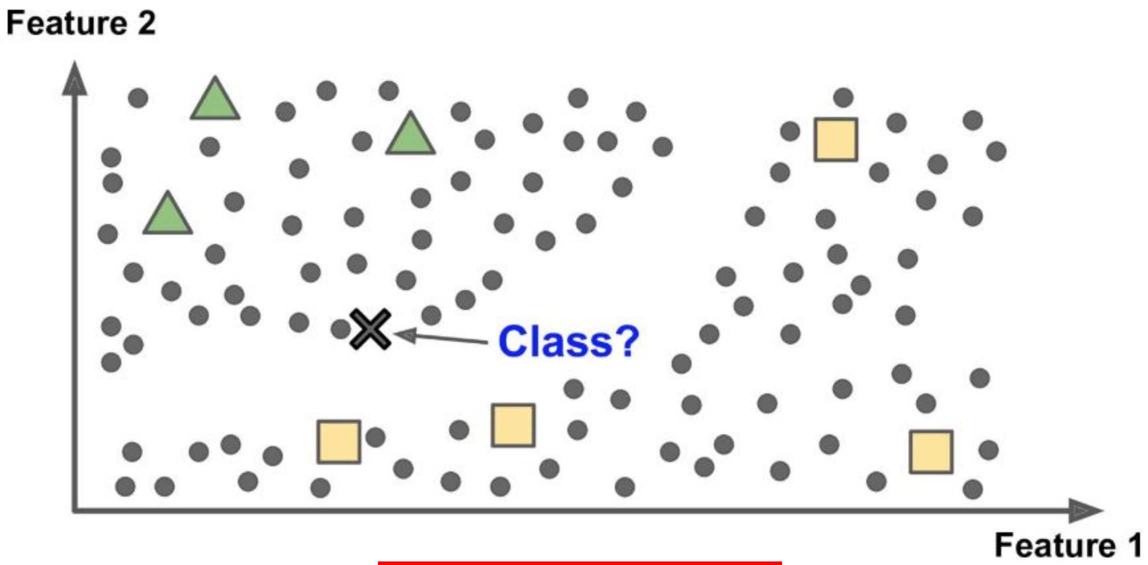
SEMI-SUPERVISED LEARNING

Semi-supervised Learning

Semi-supervised learning algorithms are trained on a combination of labeled and unlabeled data.

- The process of labeling massive amounts of data for supervised learning is often prohibitively time-consuming and expensive.
- What's more, too much labeling can impose human biases on the model.
- That means including lots of unlabeled data during the training process actually tends to improve the accuracy of the final model while reducing the time and cost spent building it

Semi-supervised Learning



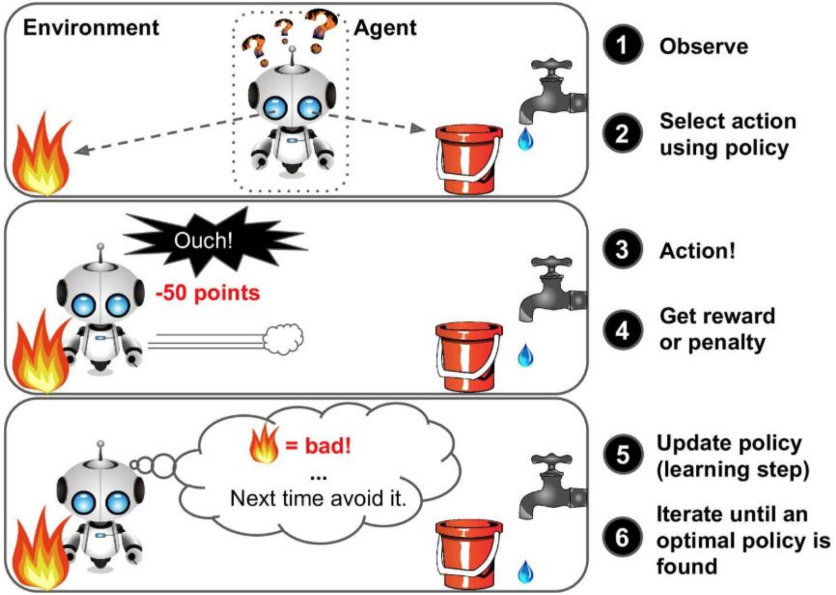
Web Page Classification
Google Photos

REINFORCEMENT LEARNING

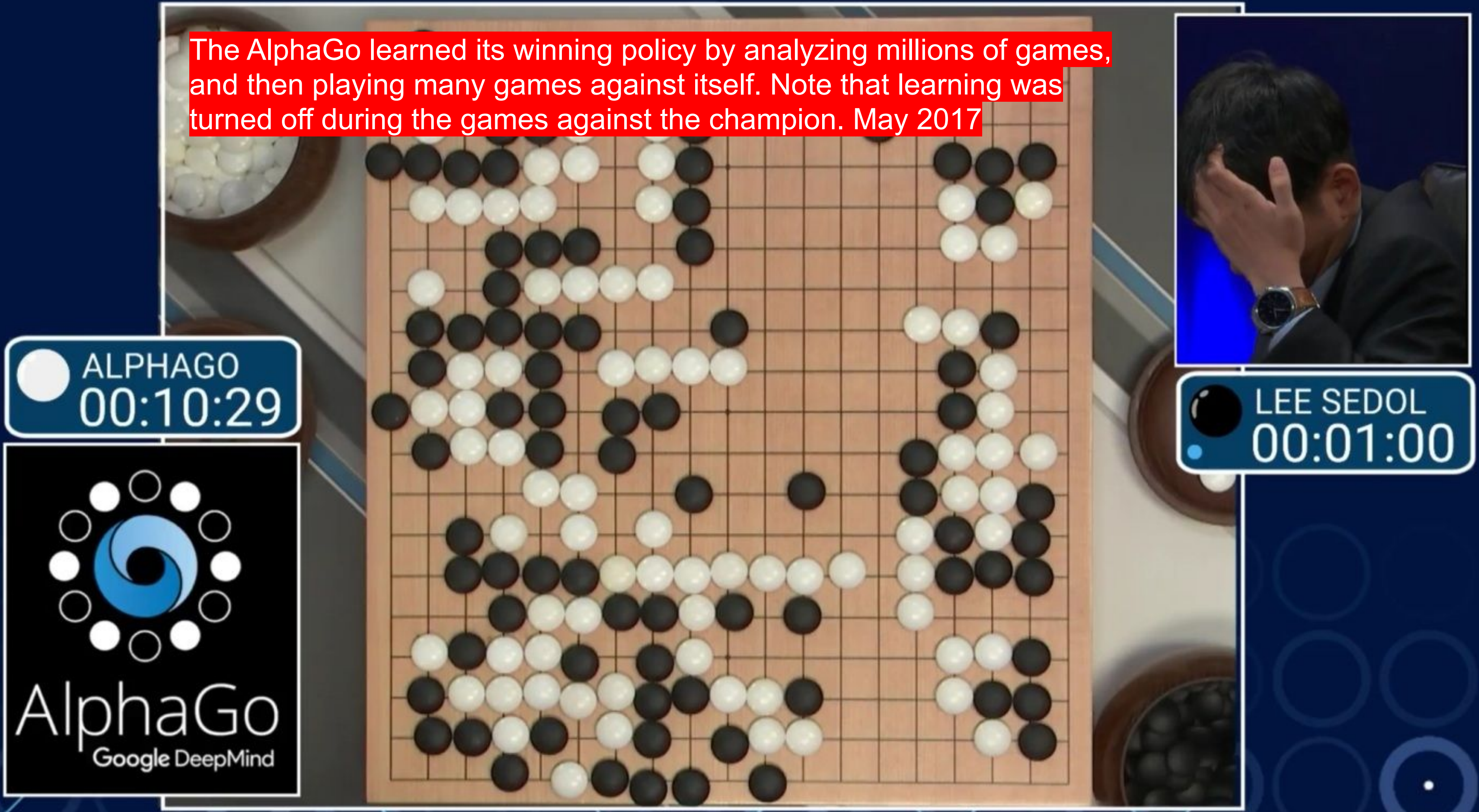
Reinforcement Learning

The learning system, called an agent in this context, can :

- Observe the environment
- Select and perform actions
- Get rewards in return (positive or negative)
- Learn by self what is the best strategy, called a policy

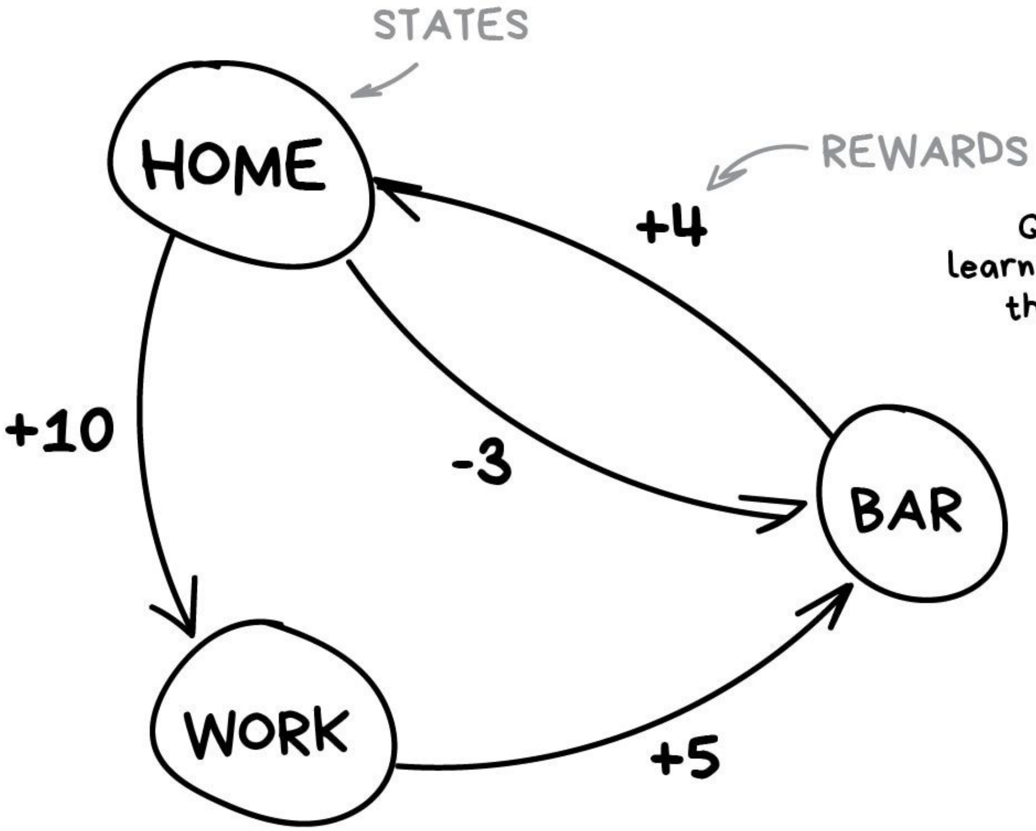


The AlphaGo learned its winning policy by analyzing millions of games, and then playing many games against itself. Note that learning was turned off during the games against the champion. May 2017

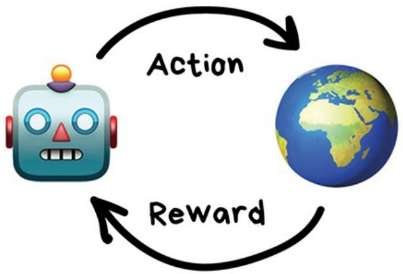


Reinforcement Learning

- Q-Learning
- State-Action-Reward-State-Action (SARSA)
- Deep Q Network (DQN)
- Deep Deterministic Policy Gradient (DDPG)
- A Brief Survey of Deep Reinforcement Learning (Dec, 2017)
  - https://arxiv.org/pdf/1708.05866.pdf

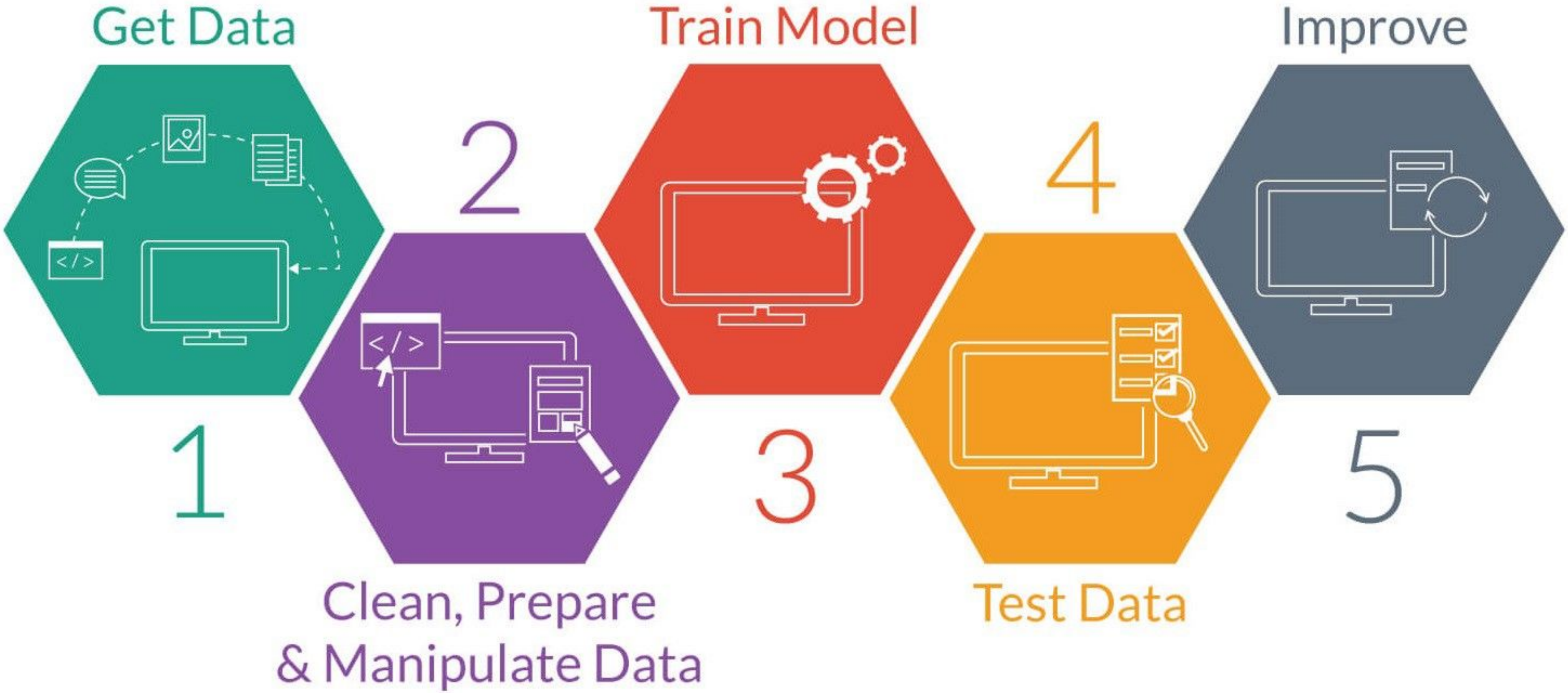


Q-learning
learns to maximize
the rewards



ROUTINE MARKOV PROCESS

A general ML workflow

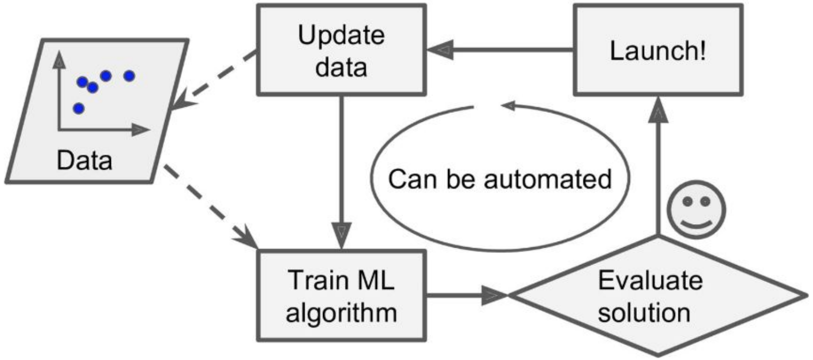


Types of Machine Learning Systems

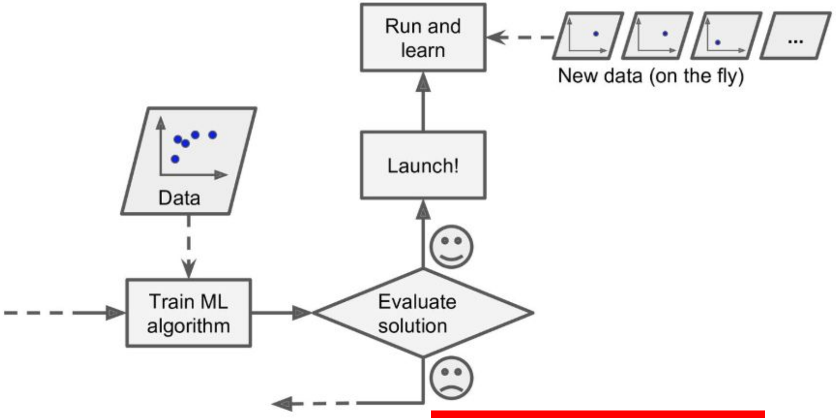
- Whether or not they are trained with human supervision
  - Supervised
  - Unsupervised
  - Semi-supervised
  - Reinforcement learning
- Whether or not they can learn incrementally on the fly
  - Online
  - Batch learning
- Whether they work by simply comparing new data points to known data points, or instead detect patterns in the training data and build a predictive model
  - Instance-based learning
  - Model-based learning

Batch and Online Learning

Another criterion used to classify Machine Learning systems is whether or not the system can learn incrementally from a stream of incoming data.



Batch Learning



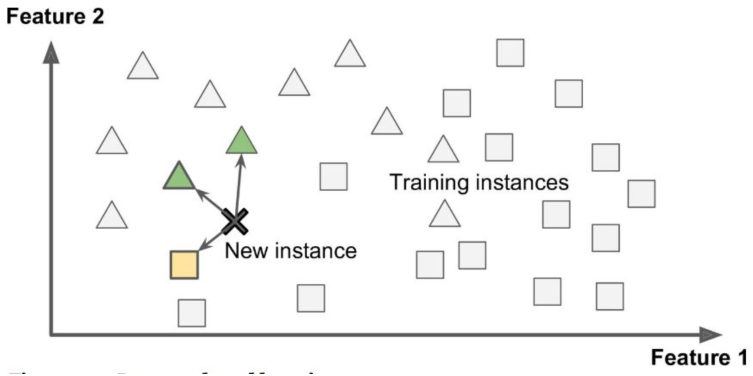
Online Learning

Types of Machine Learning Systems

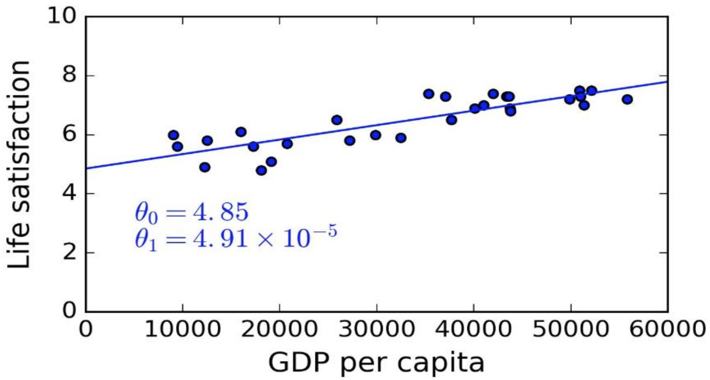
- Whether or not they are trained with human supervision
  - Supervised
  - Unsupervised
  - Semi-supervised
  - Reinforcement learning
- Whether or not they can learn incrementally on the fly
  - Online
  - Batch learning
- Whether they work by simply comparing new data points to known data points, or instead detect patterns in the training data and build a predictive model
  - Instance-based learning
  - Model-based learning

Instance-Based or Model-Based Learning

One more way to categorize ML is by how they generalize (classify or predict to examples it has never seen before).



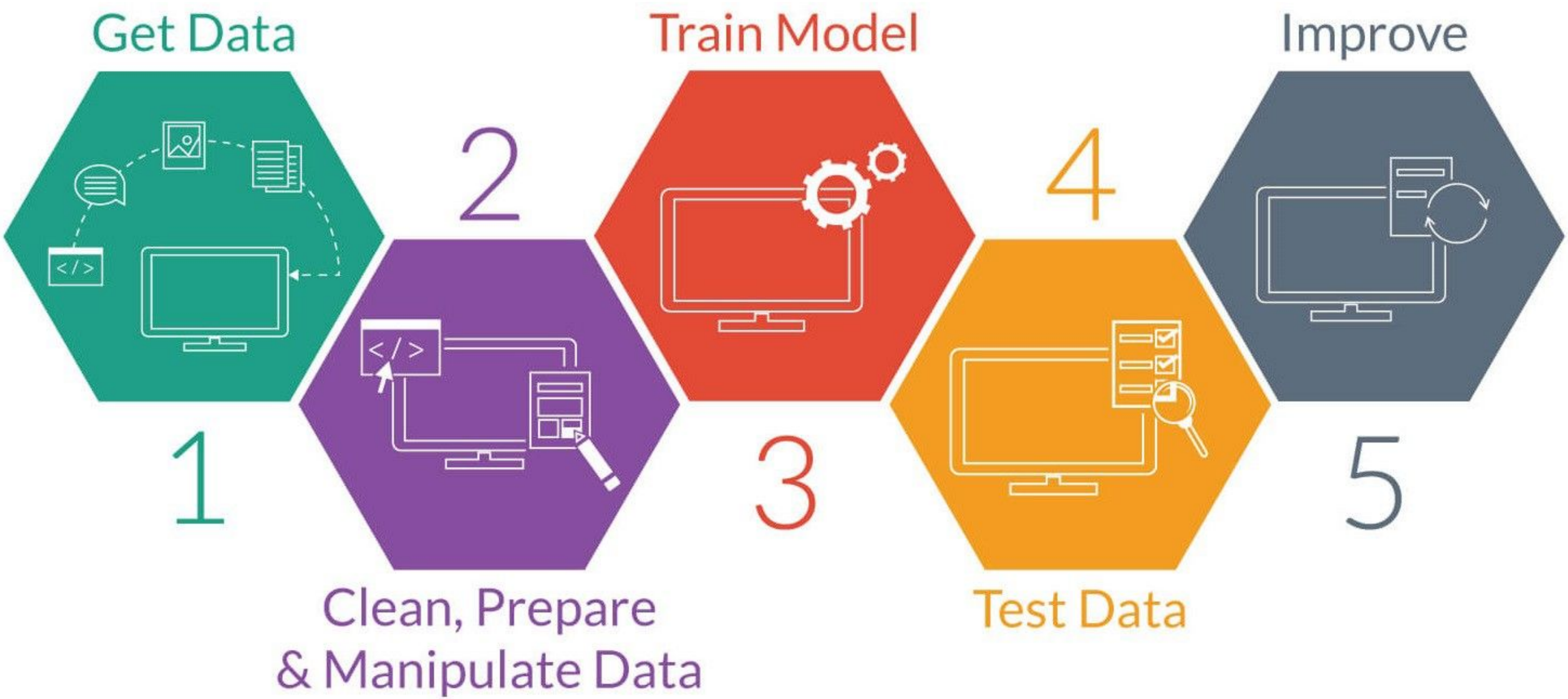
Instance-Based



Model-Based

MAIN CHALLENGES OF MACHINE LEARNING

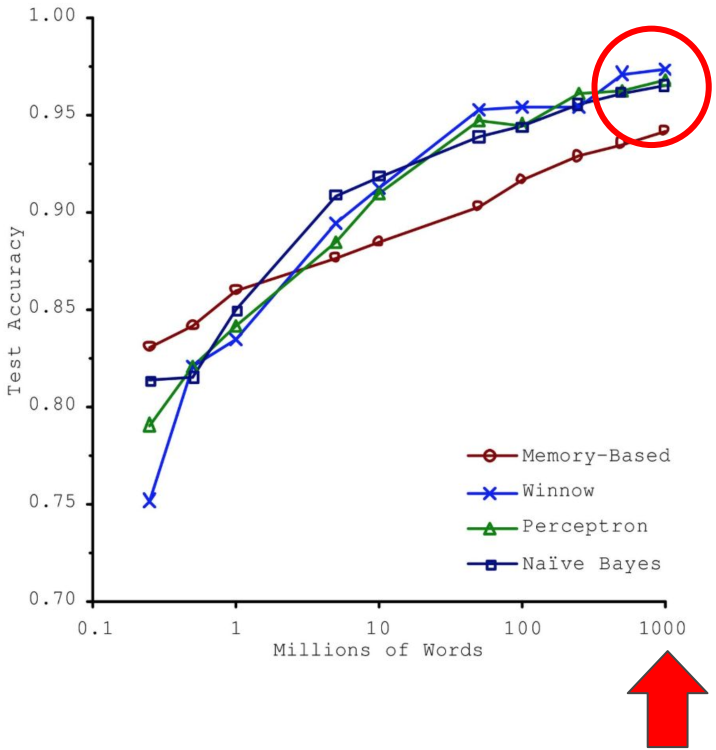
A general ML workflow



Main Challenges of Machine Learning

- Insufficient quantity of training data
- Nonrepresentative training data
- Poor quality-data
- Irrelevant features
- Overfitting the training data
- Underfitting the training data

Insufficient quantity of training data



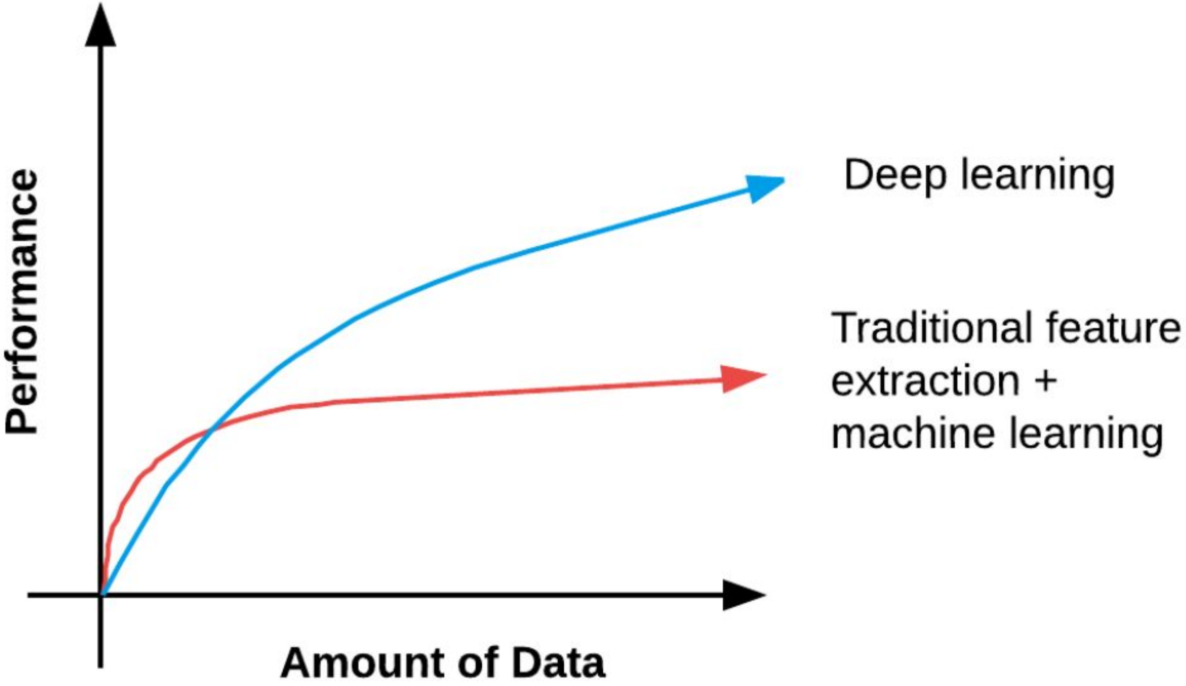
In a famous paper published in 2001, Microsoft researchers Michele Banko and Eric Brill showed that very differente ML algorithms, including fairly simples ones, performed almost identically well on a complex problem of natural language disambiguation once they were given enough data

Insufficient quantity of training data

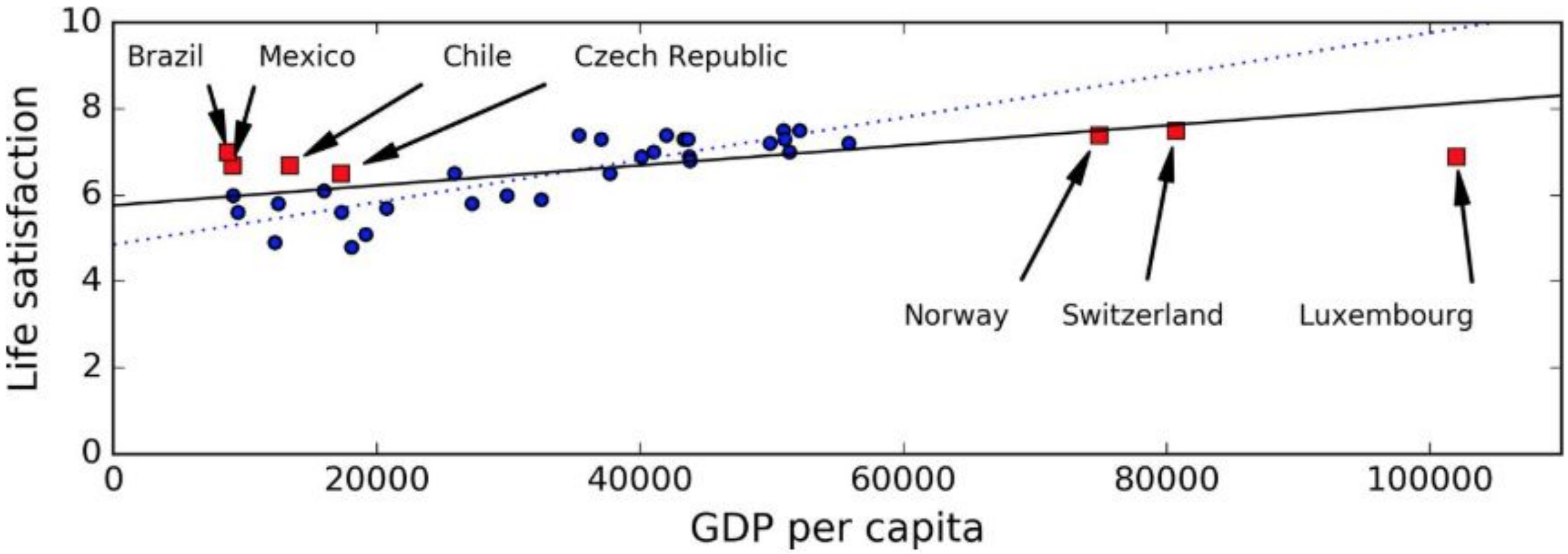
The idea that data matters more than algorithms for complex problems was further popularized by Peter Norvig et al. in a paper titled "The Unreasonable Effectiveness of Data" published in 2009.

Researchers from Google and Carnegie Mellon (2017) took a step towards clearing the clouds of mystery surrounding the relationship between "enormous data" and visual deep learning. we find that the performance on vision tasks increases logarithmically based on volume of training data size (300M images were labeled with 18291 categories resulting in more than 1 billion of labels)

Insufficient quantity of training data



Nonrepresentative Training Data



By using a nonrepresentative training set, we trained a model that is unlikely to make accurate predictions (sampling bias), especially for very poor and very rich countries

Poor-Quality Data

Obviously, if your training data is full of:

- Erros
- Outliers
- Noise
- Missing data

It will make it harder for the system to detect underlying patterns.

The truth is, most data scientist spend a significant part of their time doing cleaning up your training data.

Irrelevant Features

A critical part of the success of a ML project is coming up with a good set of features to train on. This process, called feature engineering involves:

- Feature selection
  - Selecting the most useful features to train on among existing features.
- Feature extraction
  - Combining existing features to produce a more useful one
- Creating new features by gathering new data.

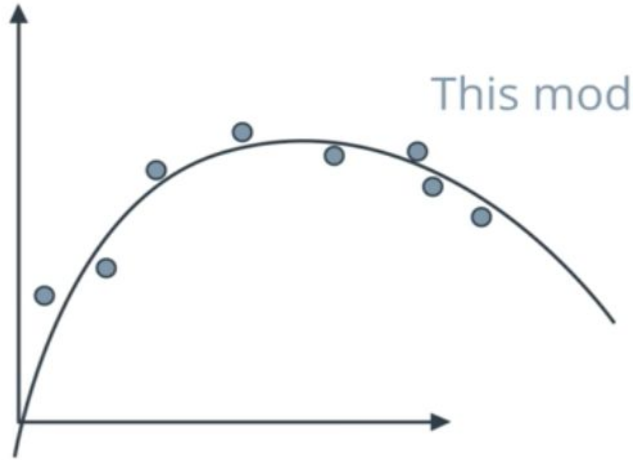
UNDERFITTING

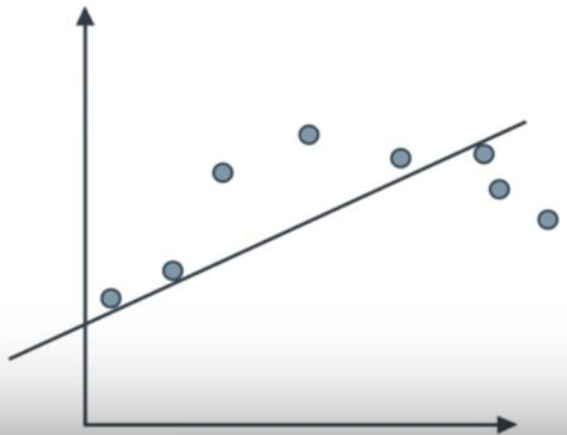
OVERFITTING

○ UNDERFITTING

Error due to bias

This model will not do well in the training set

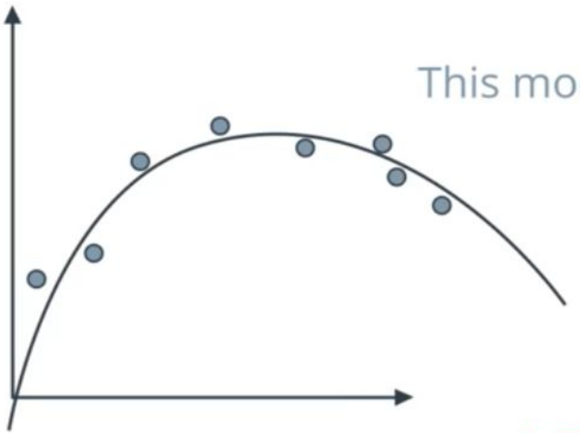


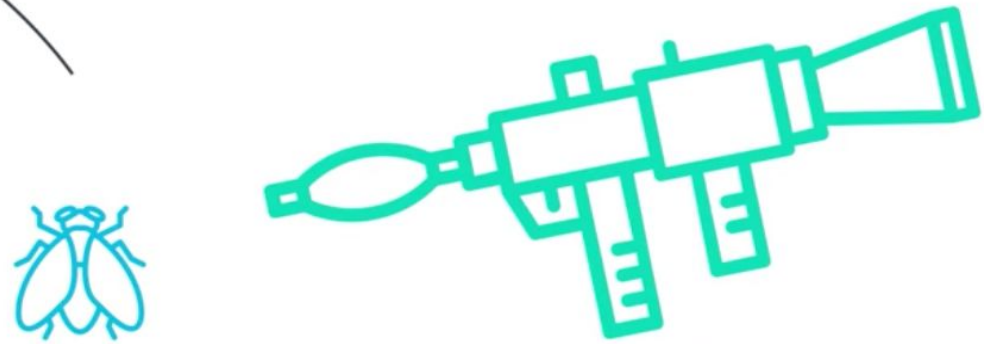


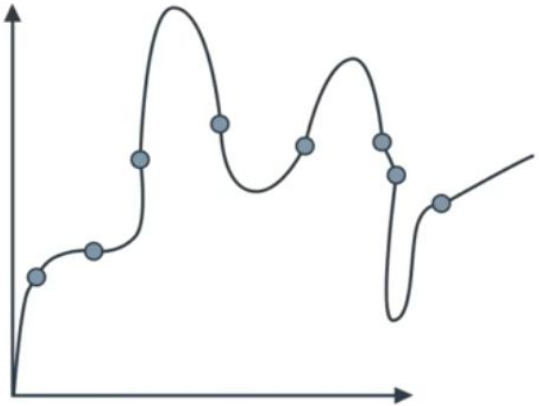
OVERFITTING

Error due to variance

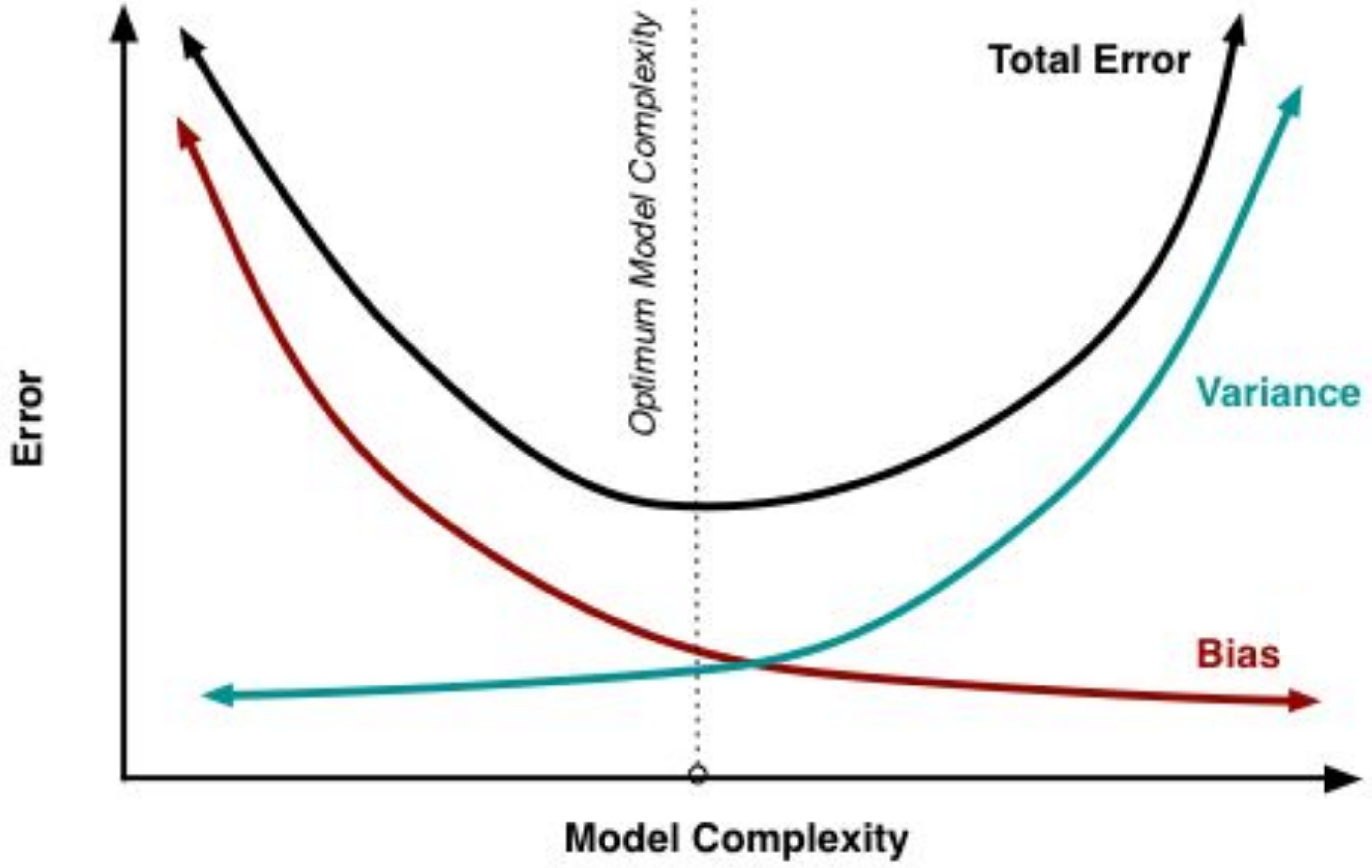
This model performs poorly in the testing set





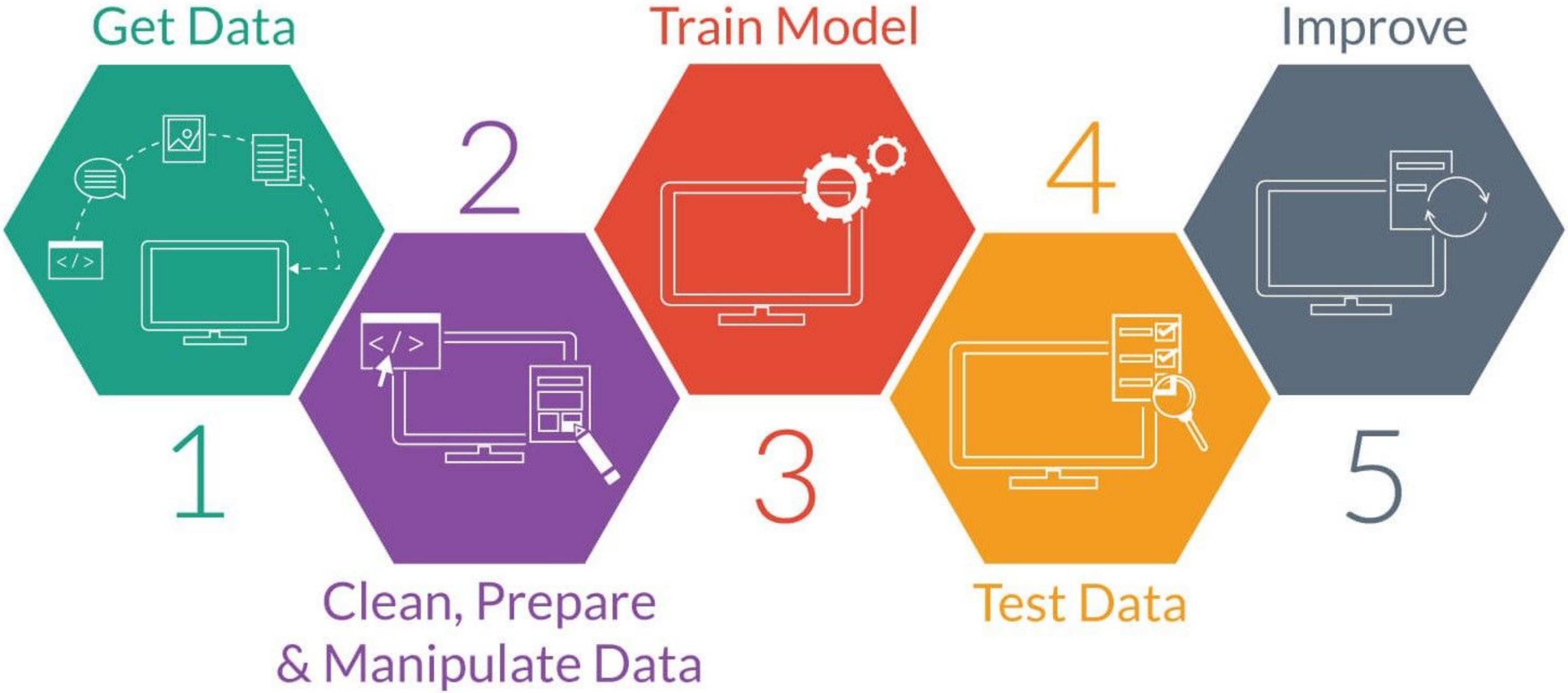


Underfitting Overfitting



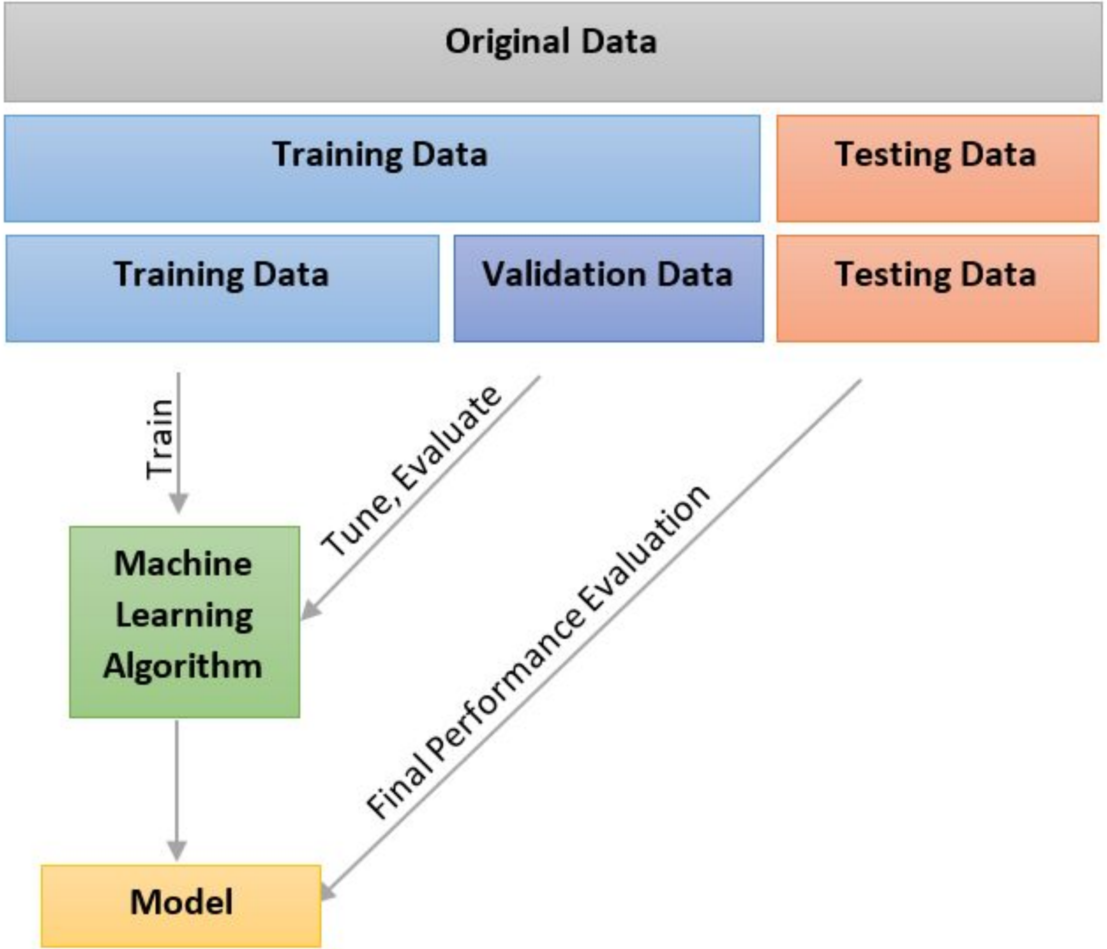
Testing
&
Validating

A general ML workflow



The only way to know how well a model will generalize to new cases is to actually try it out on new cases.

One way to do that is to put your model in production and monitor how well it performs!!!



Original Data
Training Data
Testing Data
Training Data
Validation Data
Testing Data
Train
Tune, Evaluate
Final Performance Evaluation
Machine Learning Algorithm
Model

THE END

End-to-end Machine Learning Project