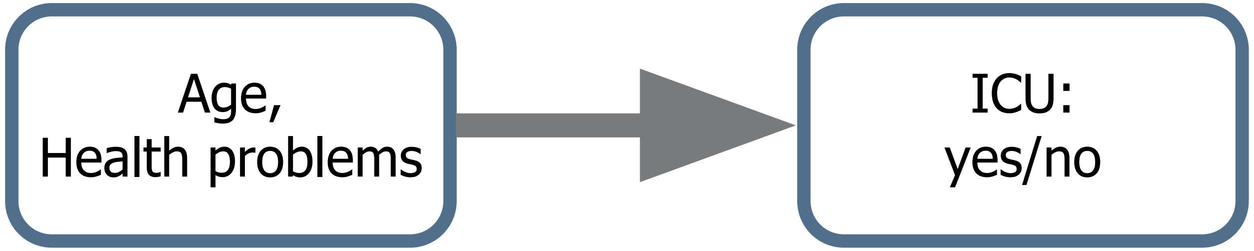
Introduction to Causal Inference

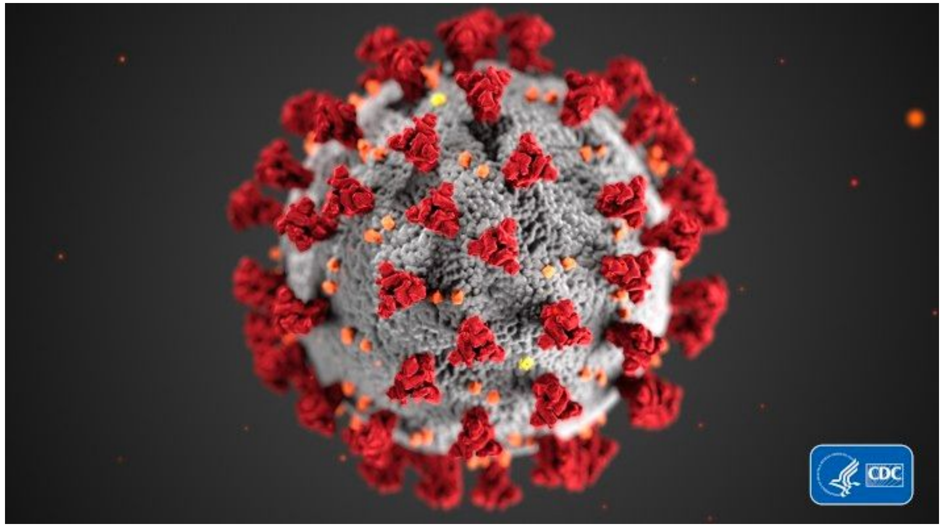
Sesiones de Estadística DAP-Cat

Aleix Ruiz de Villa (aleixrvr@gmail.com)

Main Objective of these 2 sessions

ICU ingress model due to covid





https://www.cancer.org/es/noticias-recientes/preguntas-comunes-acerca-del-brote-del-nuevo-coronavirus.html

STATISTICS / CAUSAL INFERENCE

Objective: ¿do age and healthcare affect ICU ingress?

Decision: design public health strategies

Model: finding the correct model

MACHINE LEARNING

Objective: Predict the risk

Decision: patient ranking

Model: finding the most accurate

Machine Learning main assumption

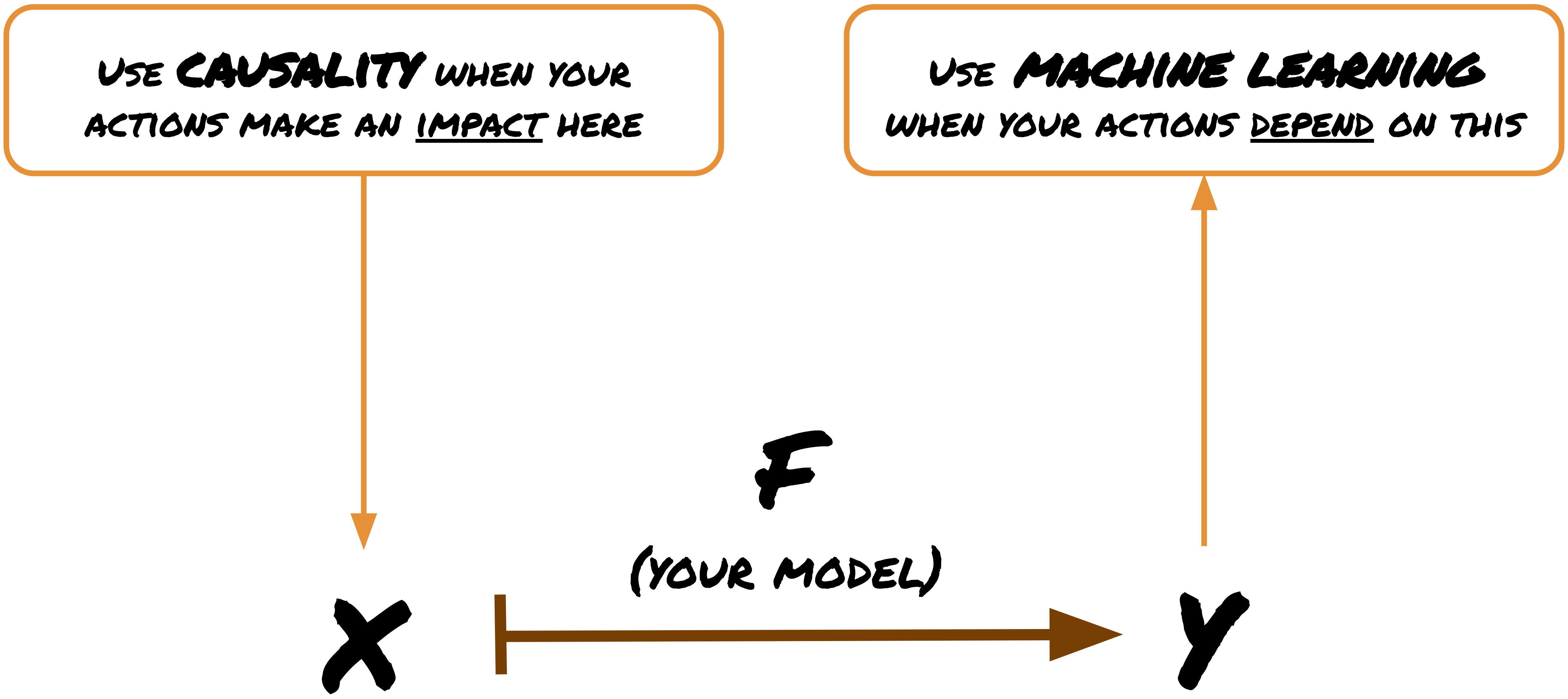
Past and Future behave the same

https://www.cancer.org/es/noticias-recientes/preguntas-comunes-acerca-del-brote-del-nuevo-coronavirus.html

Machine Learning limitation

What if we behave in a different manner?

https://www.alpinerecoverylodge.com/intervention-assistance/

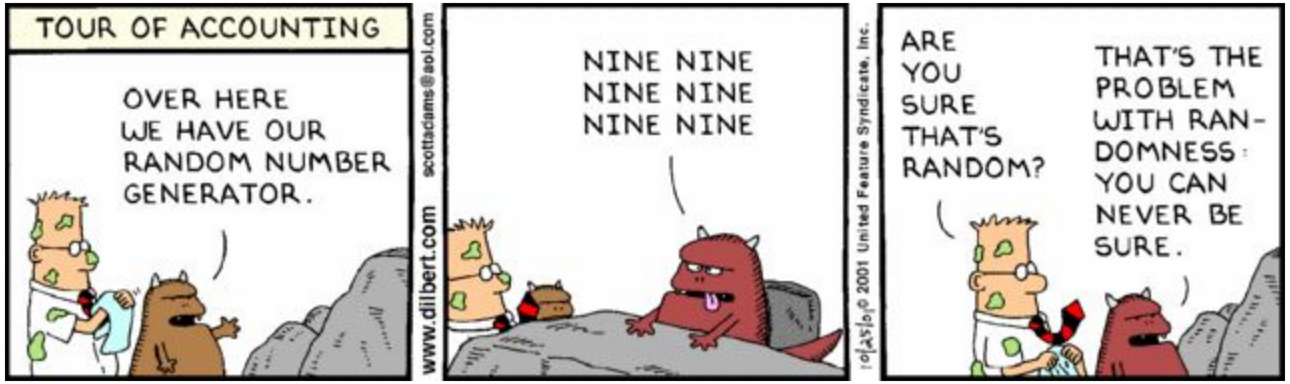


Causality = RCTs + Causal Inference

The birth of new fields

STATISTICS

Initially: What is random and what it is not?



The birth of causal inference

WELL KNOWN IN STATISTICS

- Correlation is not Causation
- Intervention is not observation

FOUNDATIONS OF

CAUSAL INFERENCE

The birth of machine learning

WELL KNOWN IN STATISTICS

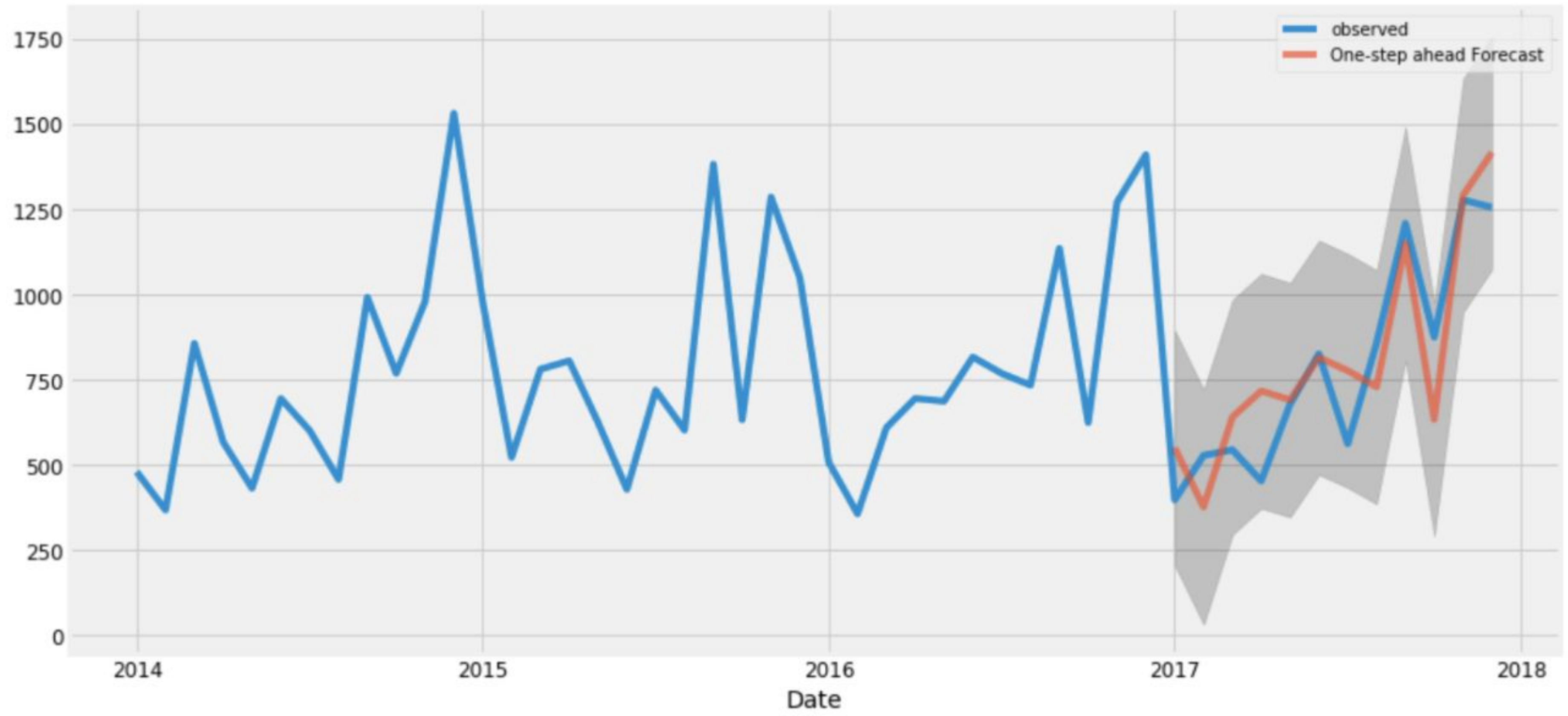
- Parsimonia Principle
- “All models are wrong, but some useful” (Box)

FOUNDATIONS OF

MACHINE LEARNING

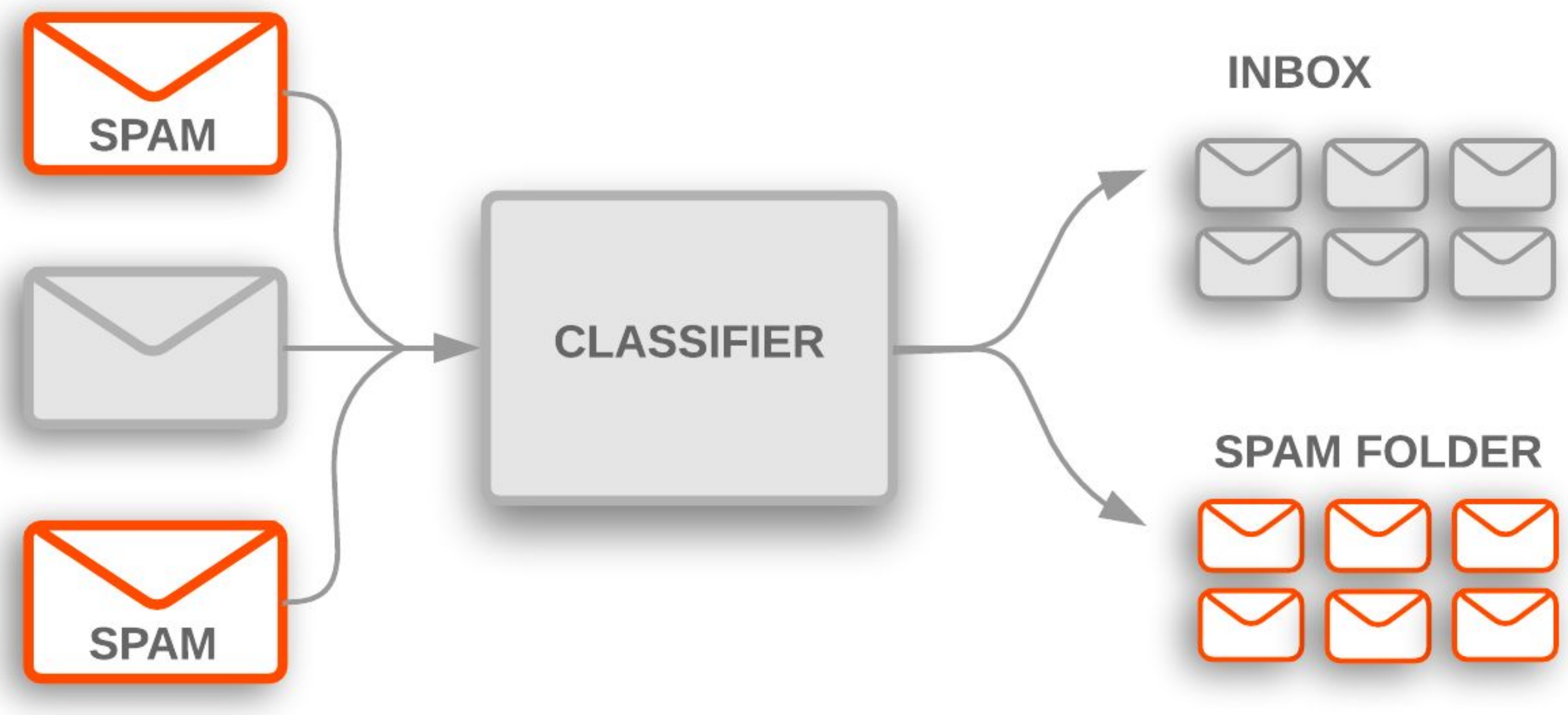
Introduction to Machine Learning

What are we talking about when we talk about machine learning?
Predictive Modelling



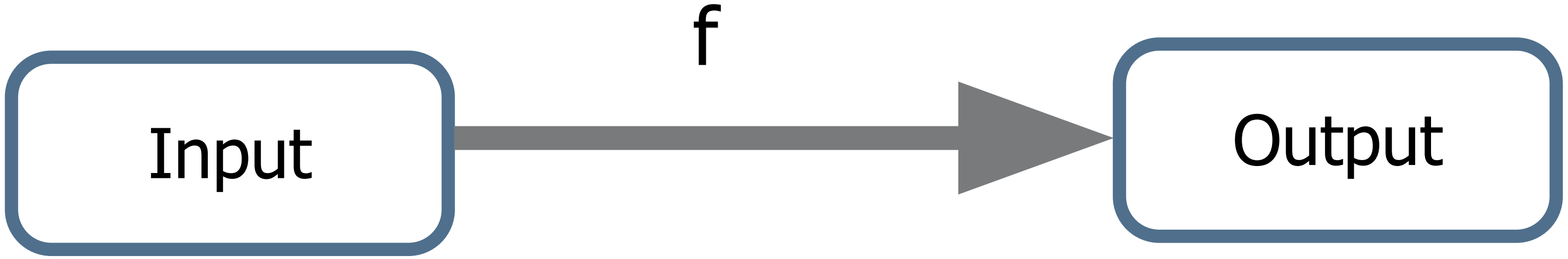
Source: https://becominghuman.ai/time-series-forecasting-7ac3344a8588

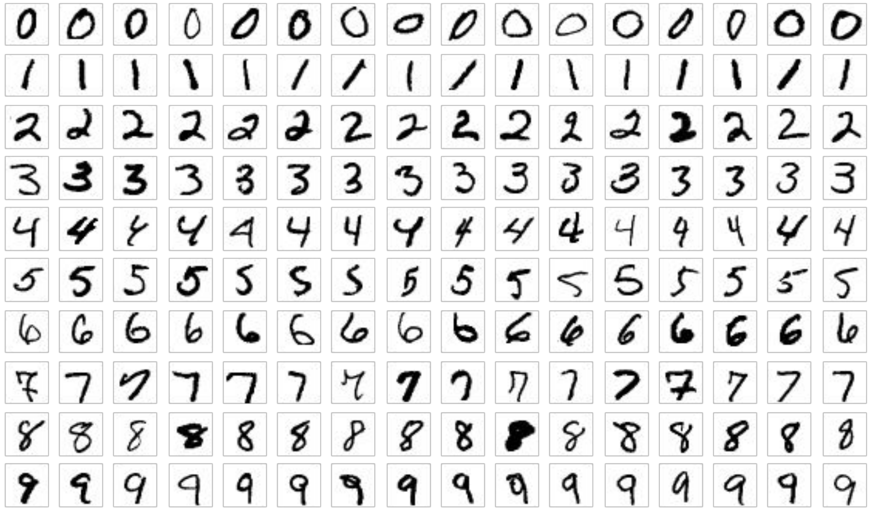
What are we talking about when we talk about machine learning?
Task Automation



Source:
https://medium.com/@naveeen.kumar.k/naive-bayes-spam-detection-7d087cc96d9d

Learning through example





0,0,0, ...
1,1,1, ...
2,2,2, ...
3,3,3, ...
4,4,4, ...
5,5,5, ...
6,6,6, ...
7,7,7, ...
8,8,8, ...
9,9,9, ...

Medical Diagnosis: detecting illness, chances of complications

Input: patients description

Output: Is ill?

Management: resource allocation

Input: hospital resources distribution and current demand

Output: Tomorrow’s level of congestion

Medical Diagnosis through image: radiography

Input: image

Output: Is ill?

Medical Diagnosis through sensors:

Input: patients current health data

Output: Is ill?

Personalized Medicine: using genomics data

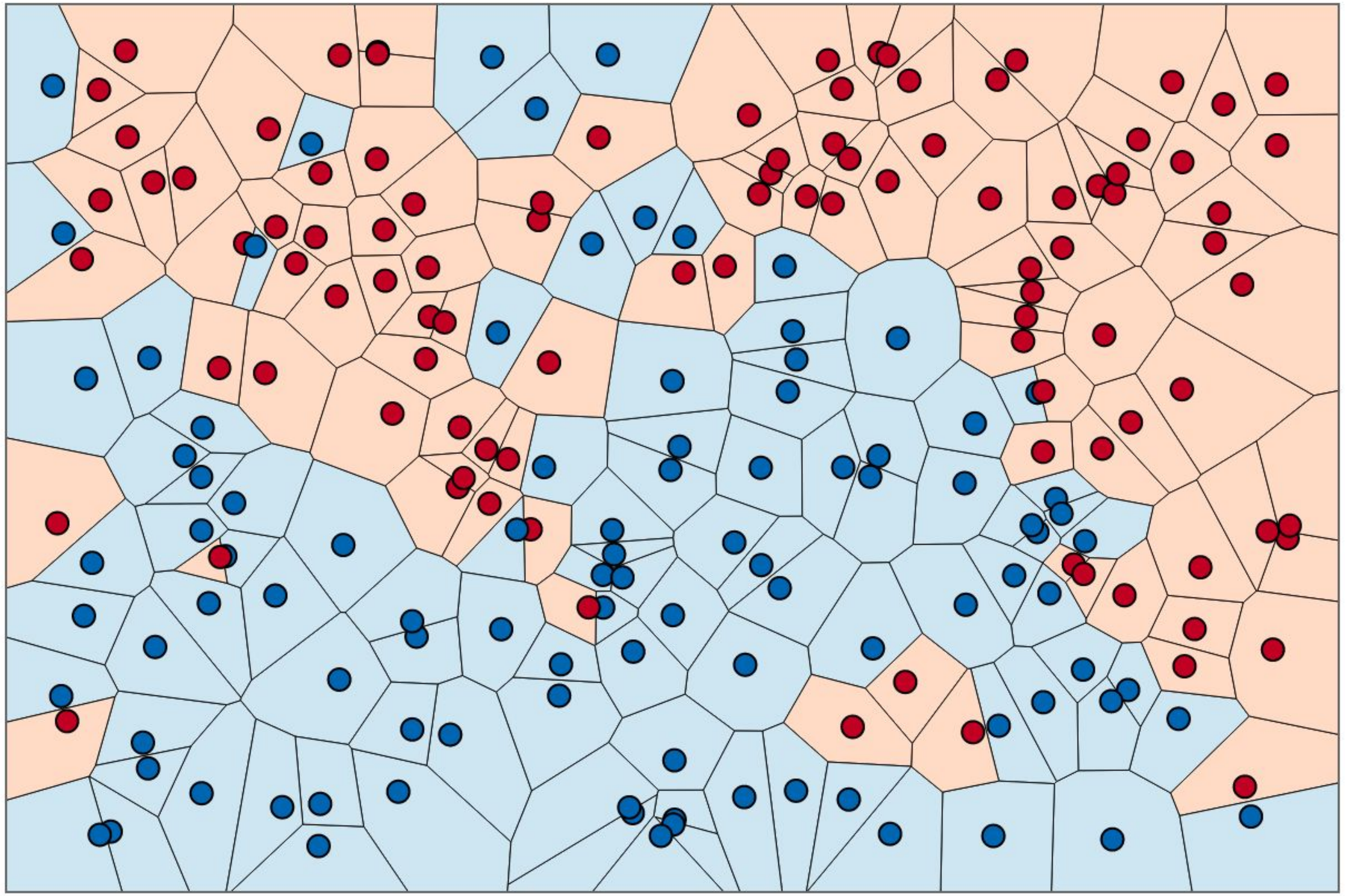
Input: genome + treatment

Output: Is ill?

Note: this can be tricky, combines causality!

The history of Machine Learning

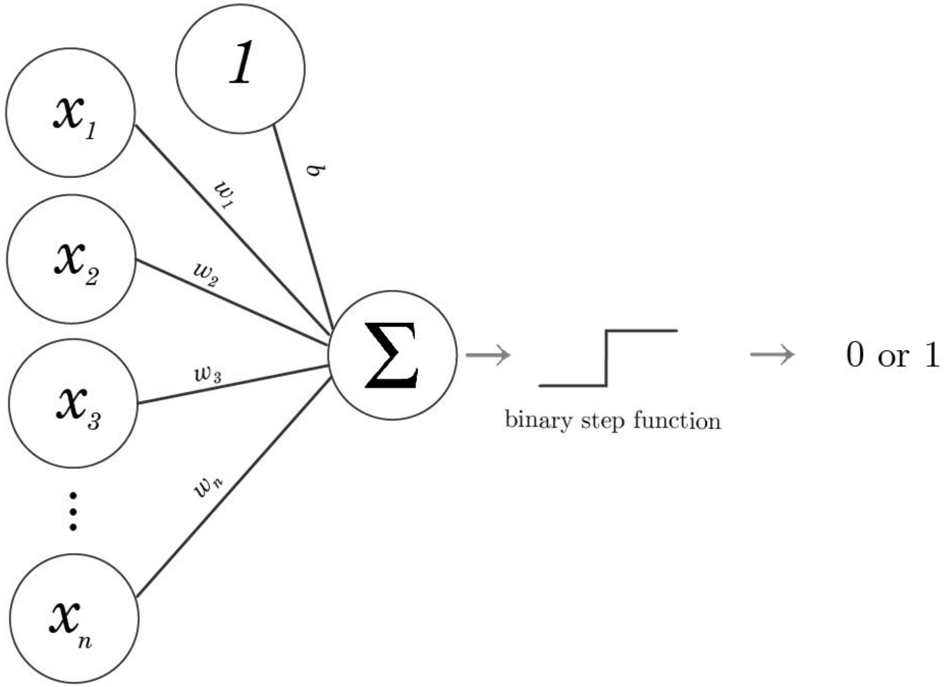
k Nearest Neighbors - Fix, Hodges (1951)

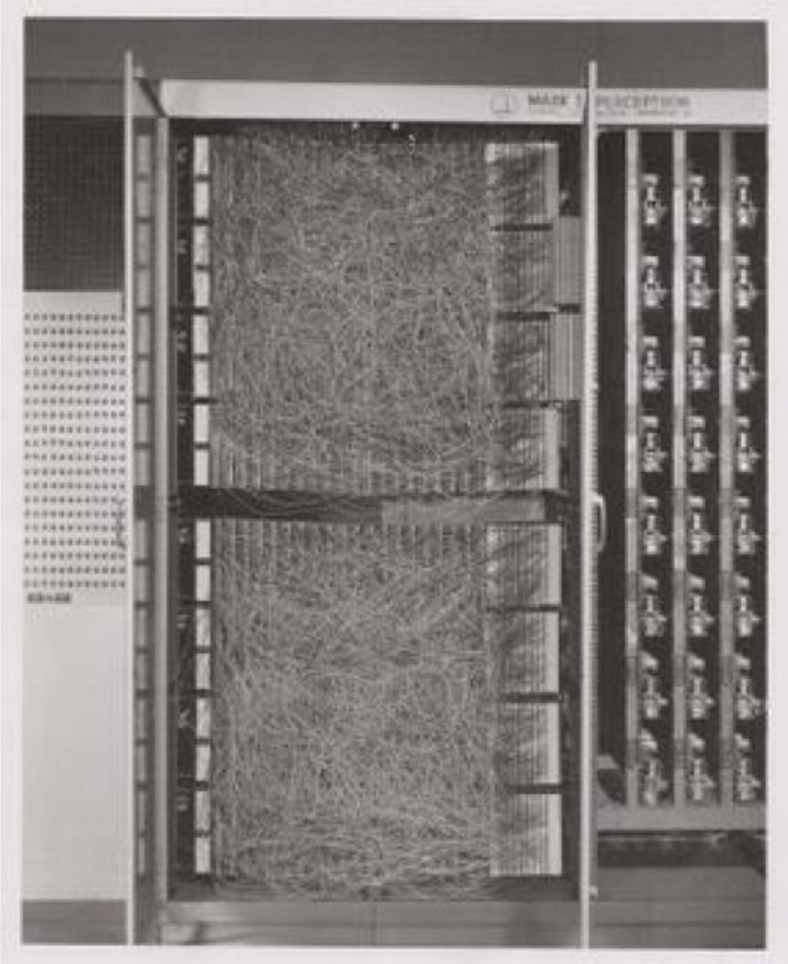


The most natural way of thinking about machine learning

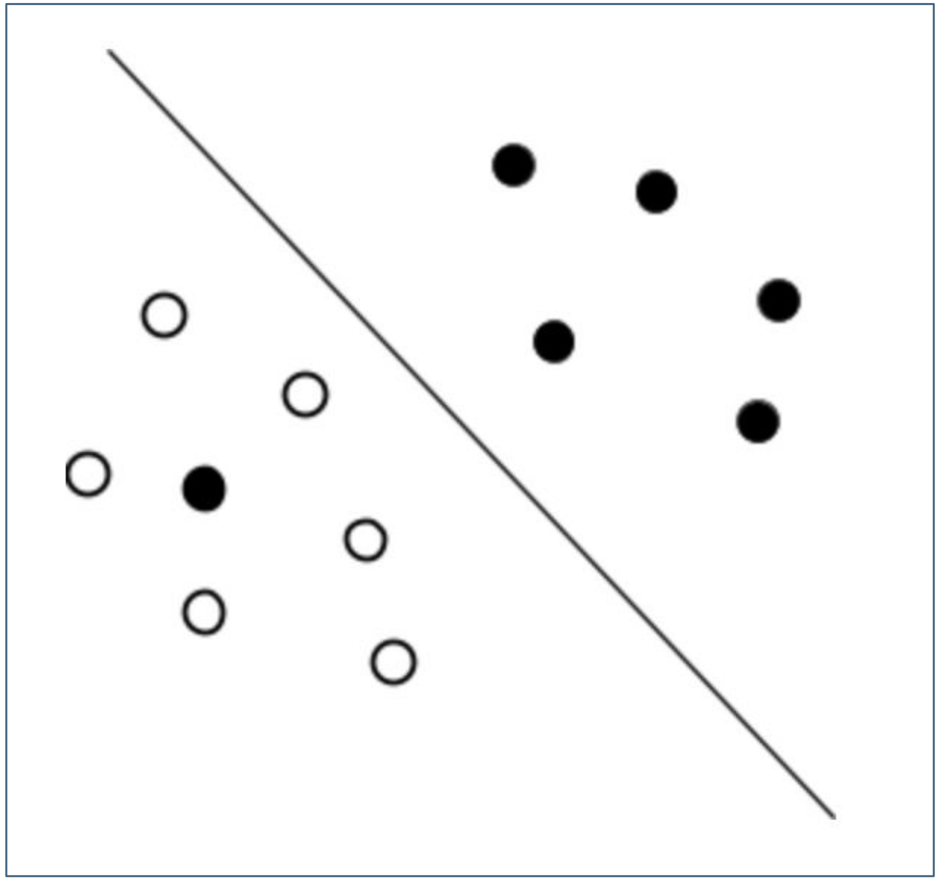
Perceptron - Rosenblatt (1958)

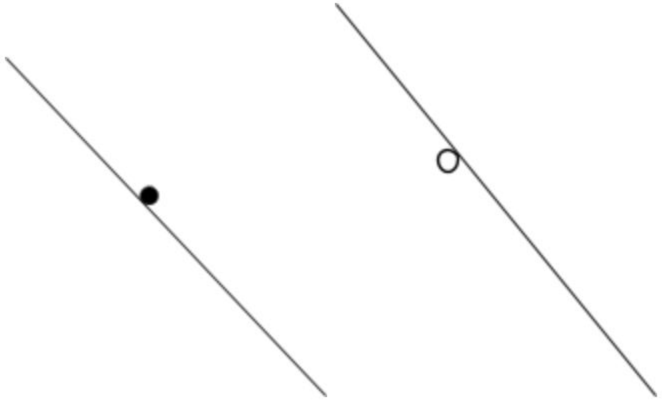
- Linear Model
- Incremental Learning
- Classification Problem





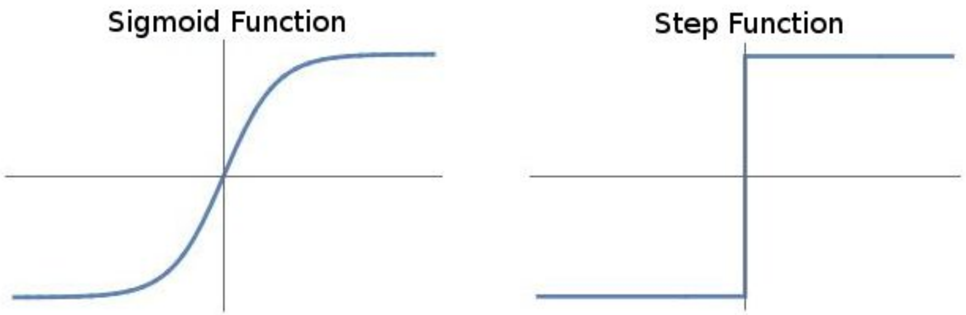
Perceptron - Rosenblatt (1958)





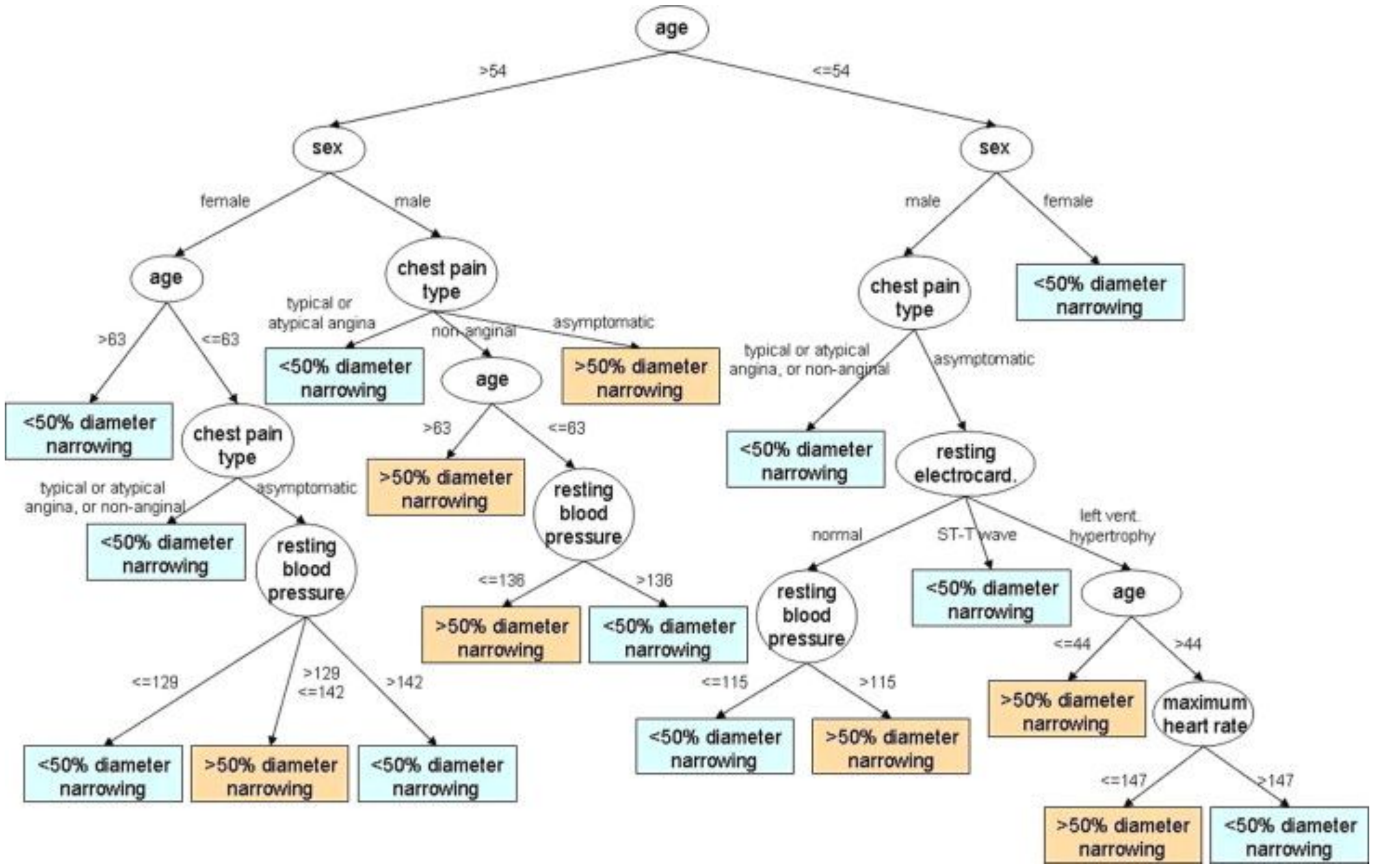
Smoothing problem

Logistic Regression



80’s

Decision Trees



- Leo Breiman: “The two cultures”
- Automating the creation of decision rules
- (are unstable with small changes of data)

First Neural Networks

LeCun - Currently Facebook

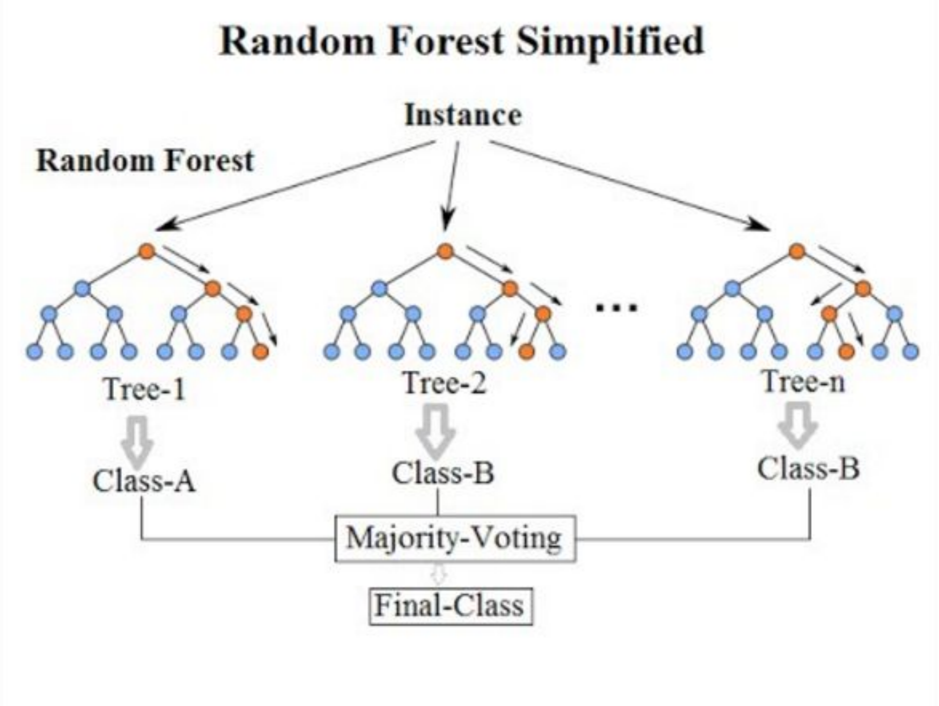
Hinton - Currently Google Brain

Bengio - Currently University Montreal

90’s

More techniques and more theory

- Random Forests (1995), Boosting (1996): currently highly used
- Vapnik's Machine Learning Theory
- Probably Approximate Correct - Valiant's theory (1984)
- Vapnik's Support Vector Machines
- Winter AI (neural networks)



2010’s

More techniques and more theory

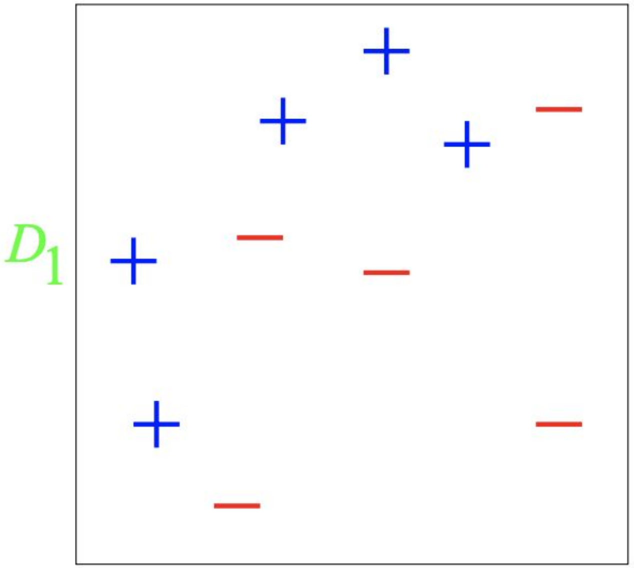
- Deep Learning
- Reinforcement Learning

Classical Models

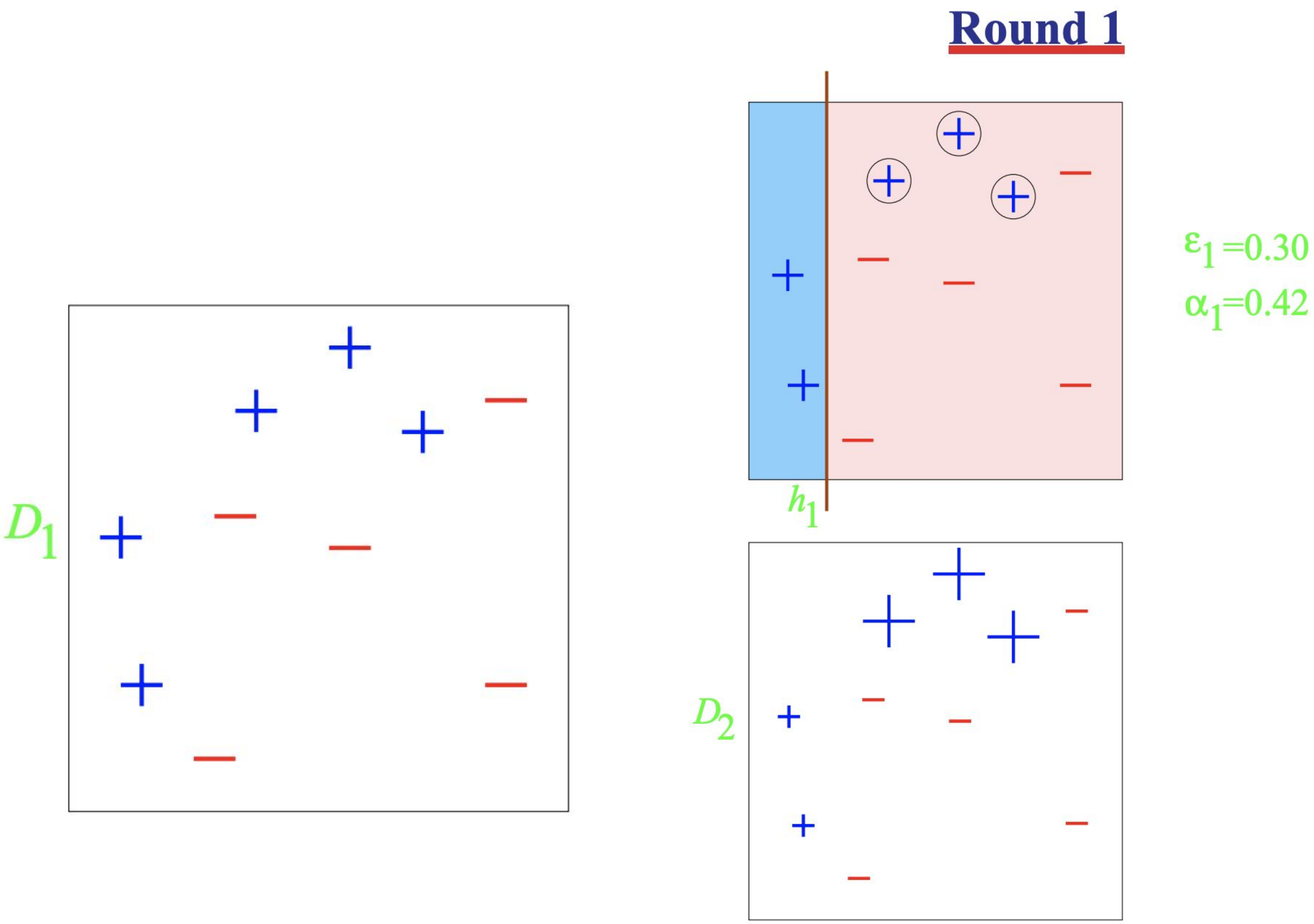
Building decision trees

http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

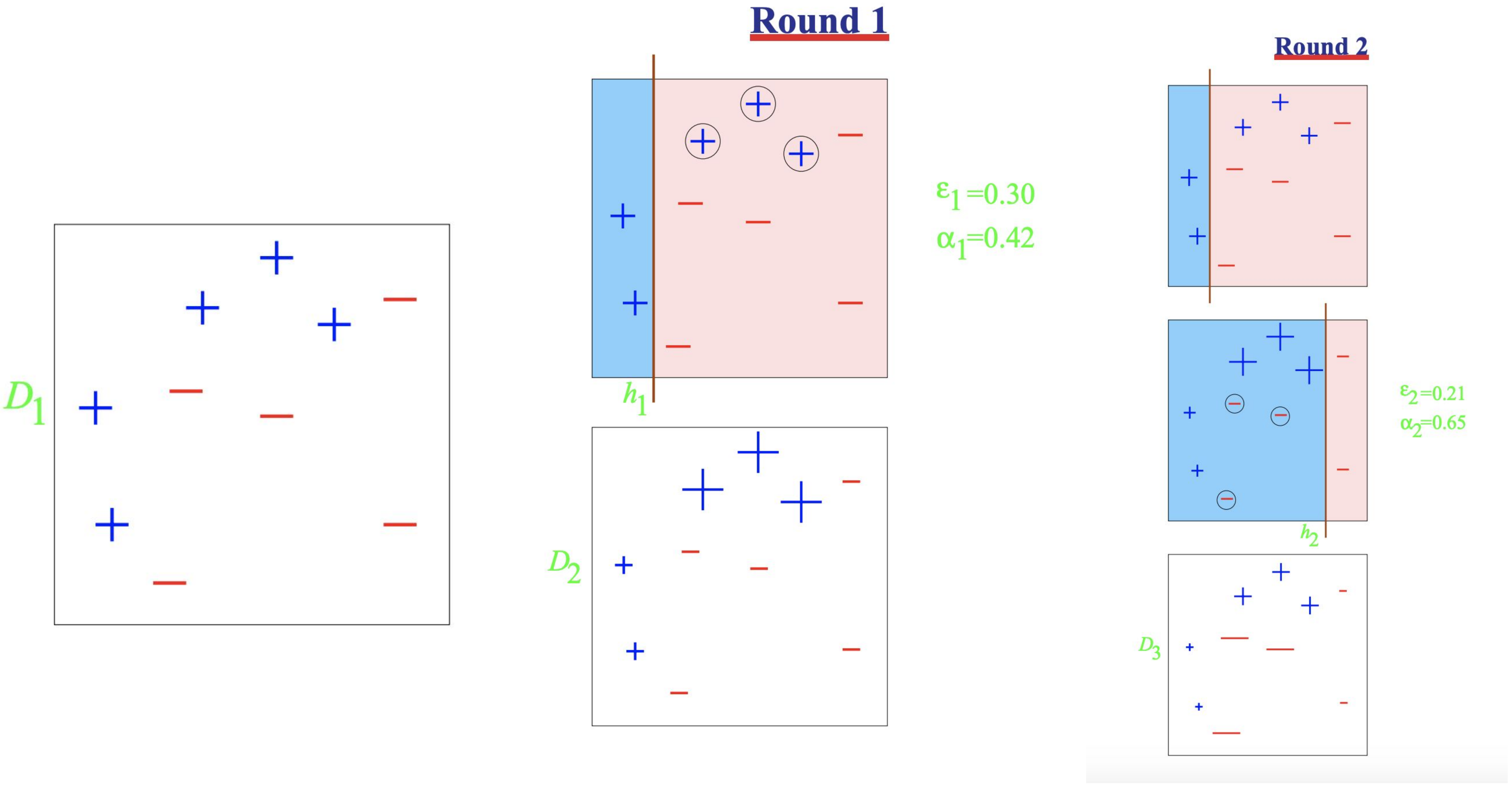
Boosting



Boosting



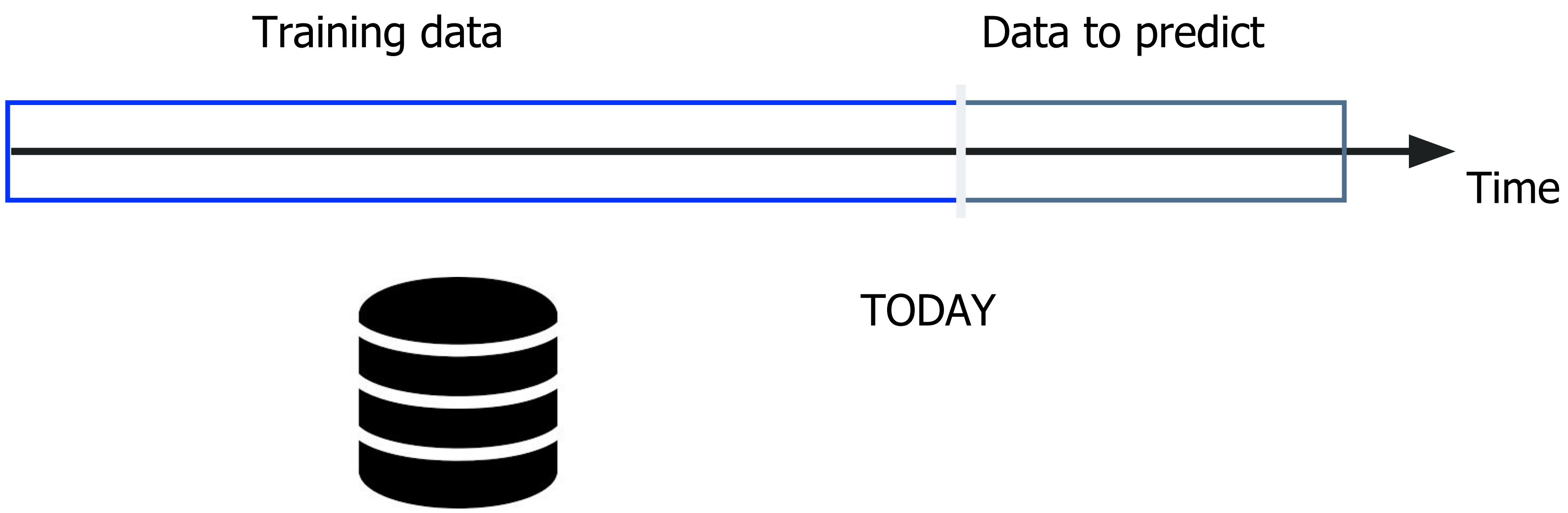
Boosting



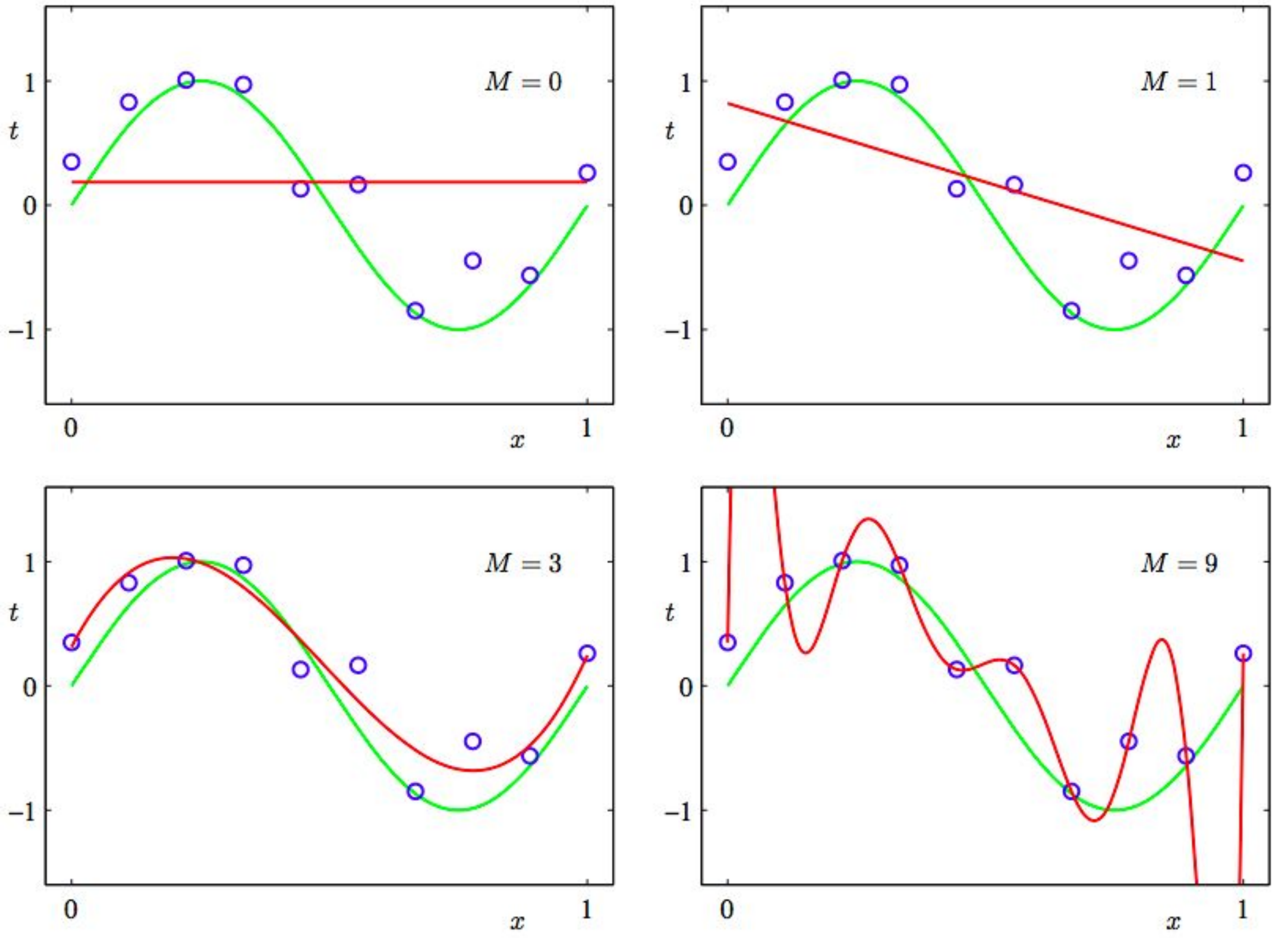
D_1
Round 1
h_1
D_2
ε_1=0.30
α_1=0.42
Round 2
h_2
D_3
ε_2=0.21
α_2=0.65

Machine Learning Theory

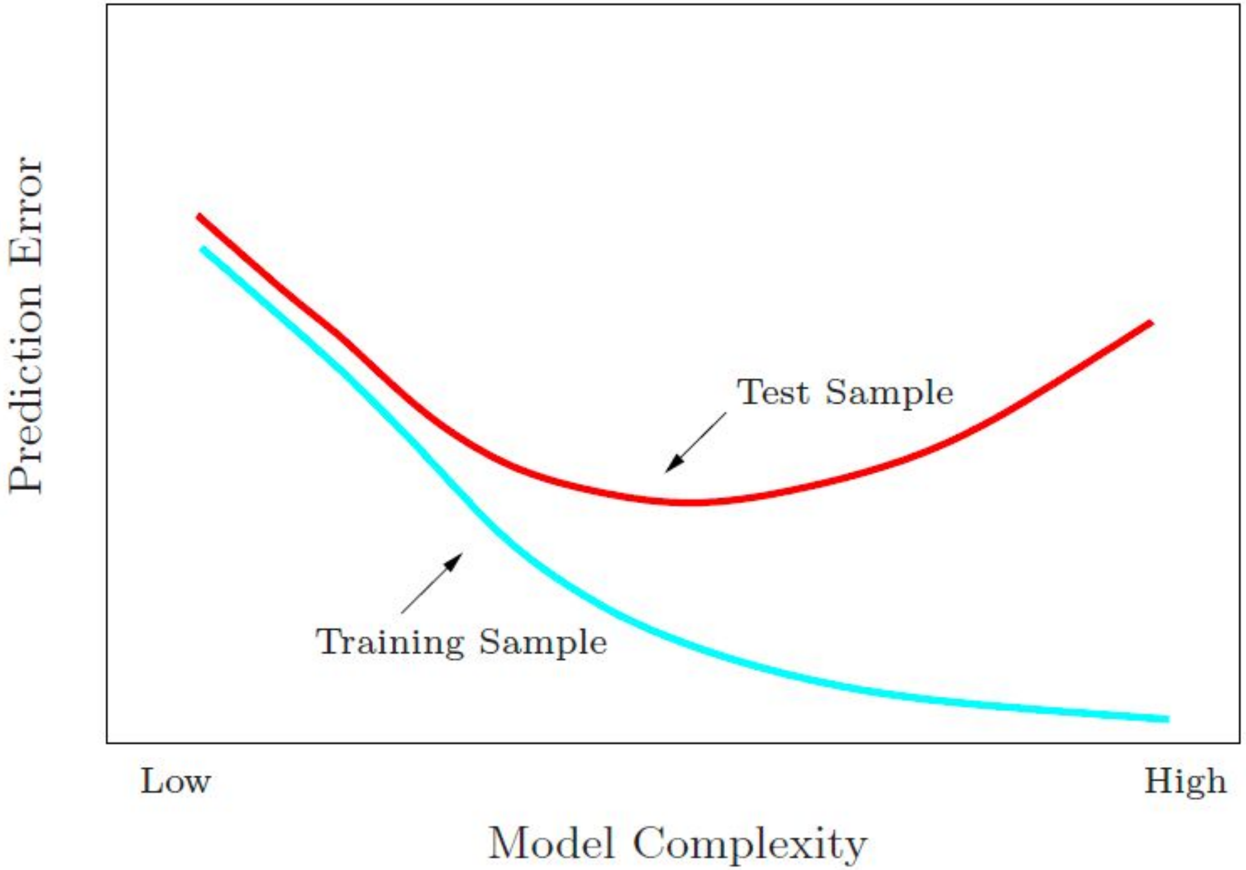
How do we use a model? How can we evaluate it?



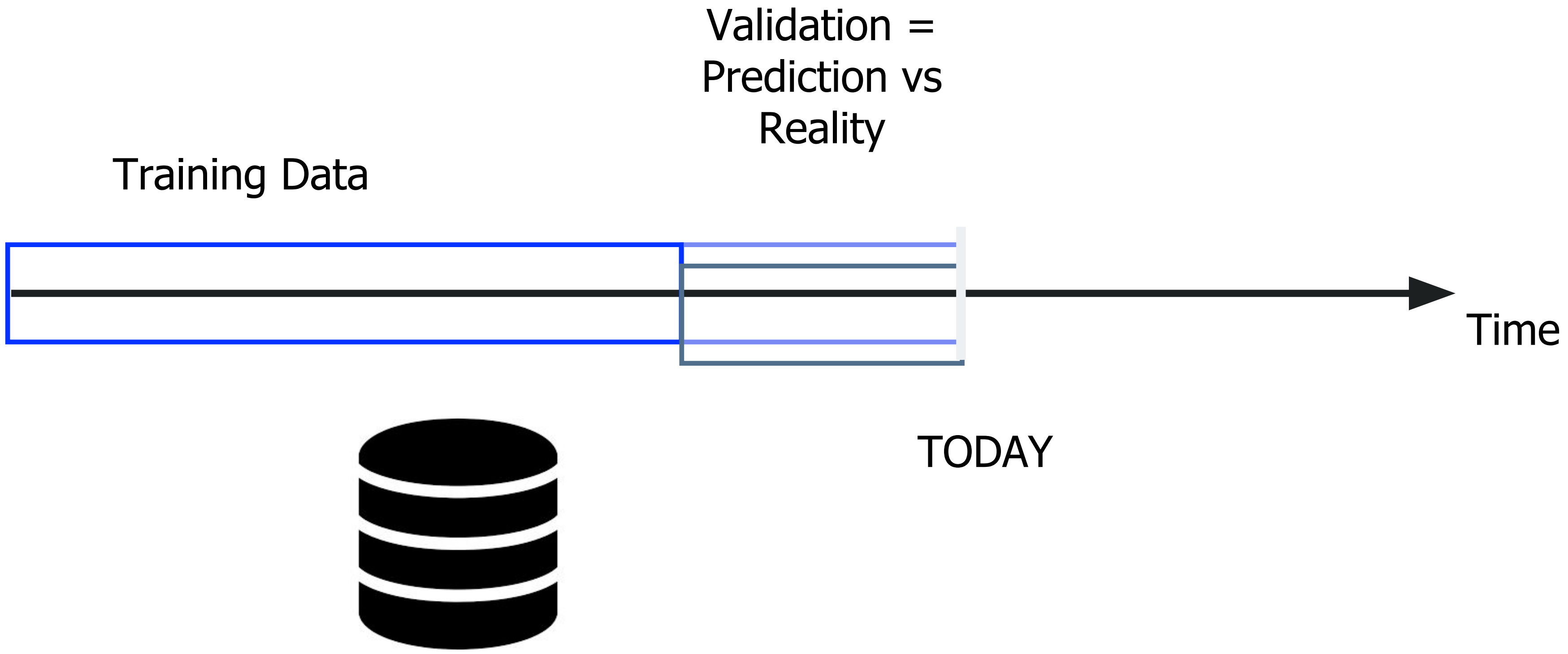
Model Complexity - Parsimonia Principle



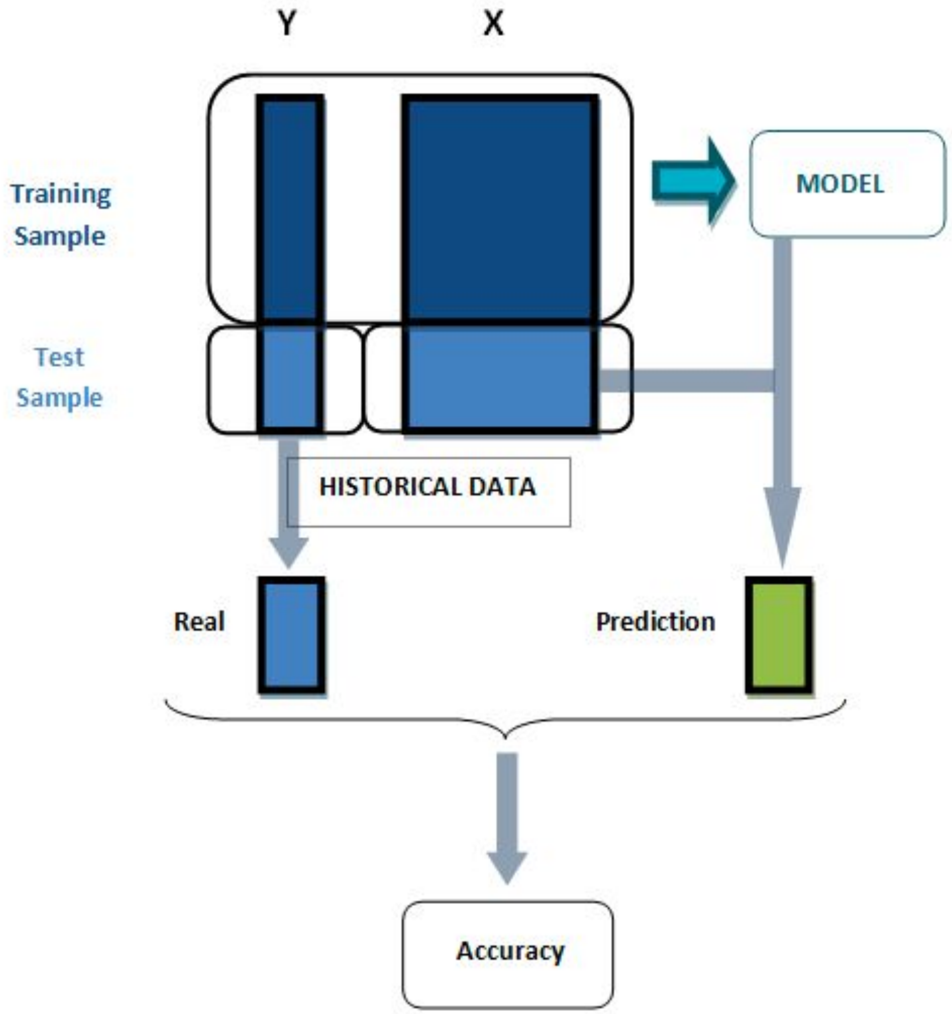
Error evolution



Cross validation: simulating the future



Simple Cross Validation



Machine Learning Theory

60's-90's Vapnik & Chervonenkis Theory

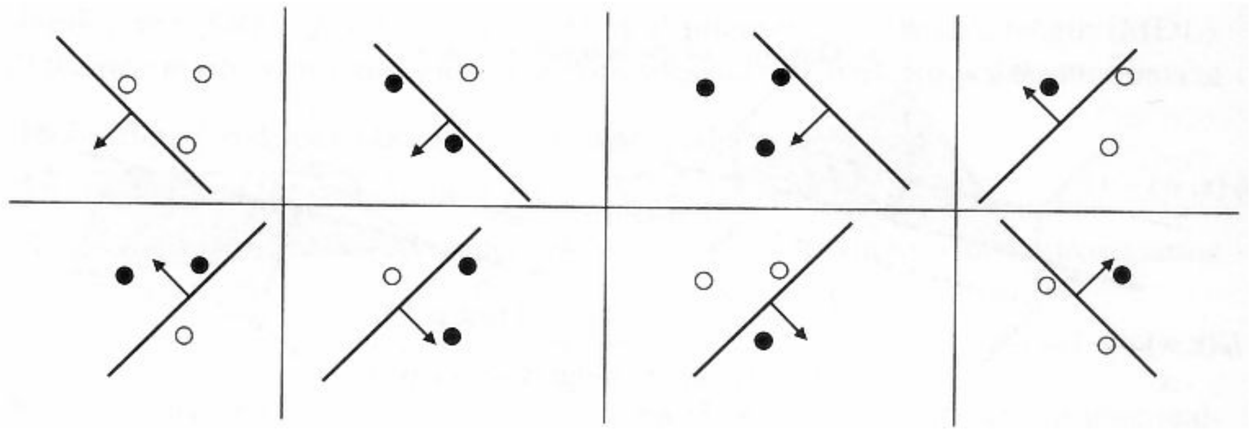
- No assumption on the distribution (different than traditional statistics)
- Main assumption: data is independent identically distributed

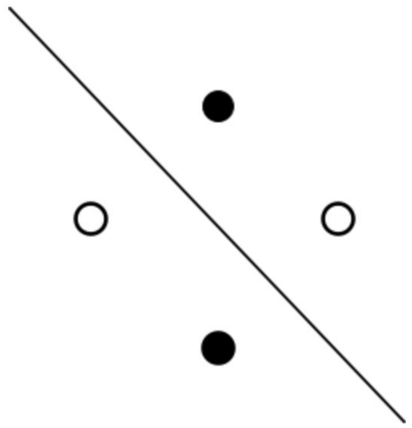
<u>The past & the future have the same distribution</u>

Occam's Razor Theorem

VC Dimension

- Number of variables in linear regression
- Magnitude of coefficients in linear regression (lasso)
- Depth in decision trees
- Number of parameters in polinomial interpolation
- Number of parameters in neural nnetworks





Occam's Razor Theorem

Confidence interval relating error on the data set, error in the future data, complexity of your models and sample size

- err_e: error training set
- err_v: error future set (validation)
- d: dimension (complexity) of the family model
- n: sample size
- C: constant

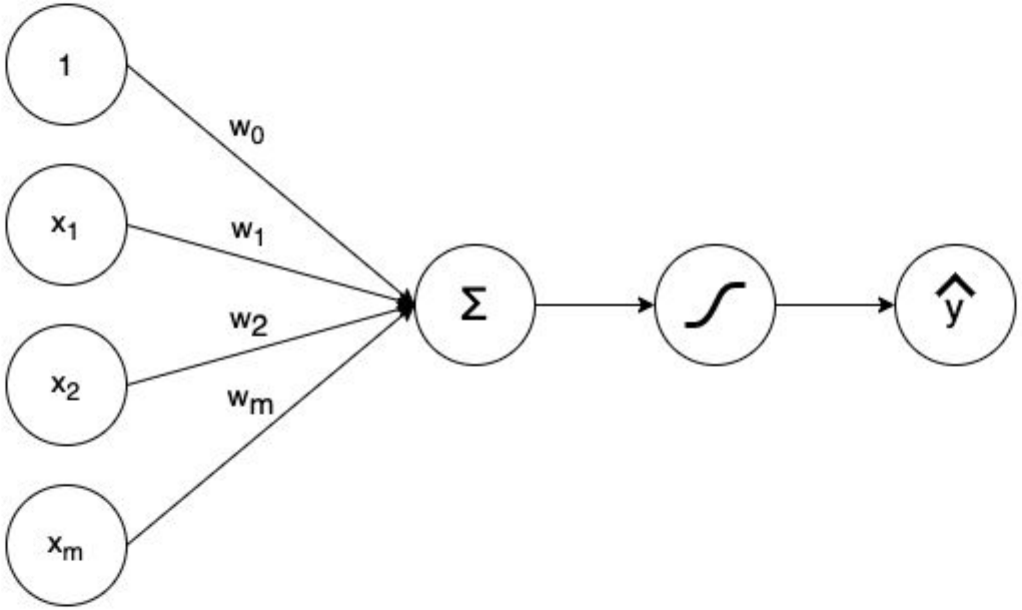
$$\boxed{err_e - \sqrt{Cd/n} \le err_v \le err_e + \sqrt{Cd/n}}$$

No Free Lunch Theorem

For any machine learning algorithm you can think of, you can always create a data set where it fails miserably

Deep Learning

Perceptron



https://dev.to/_akshaym/neural-network-basics-the-perceptron-3c4

Feedforward Neural Network

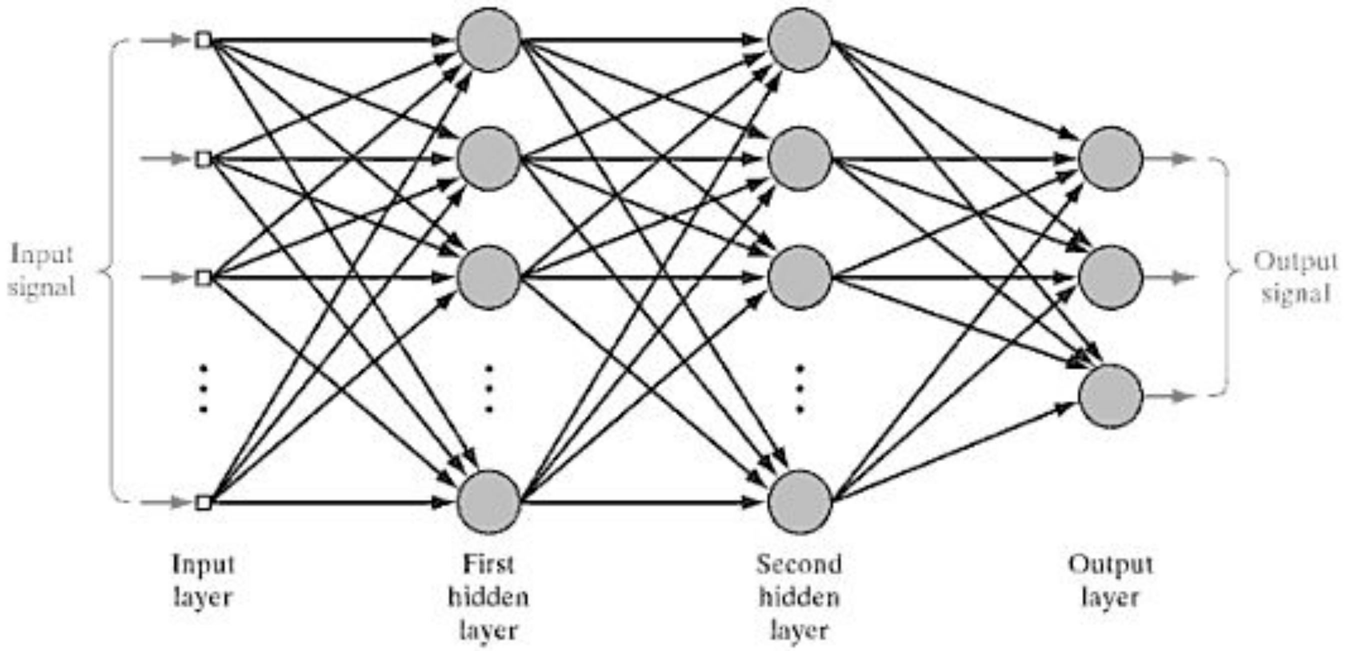
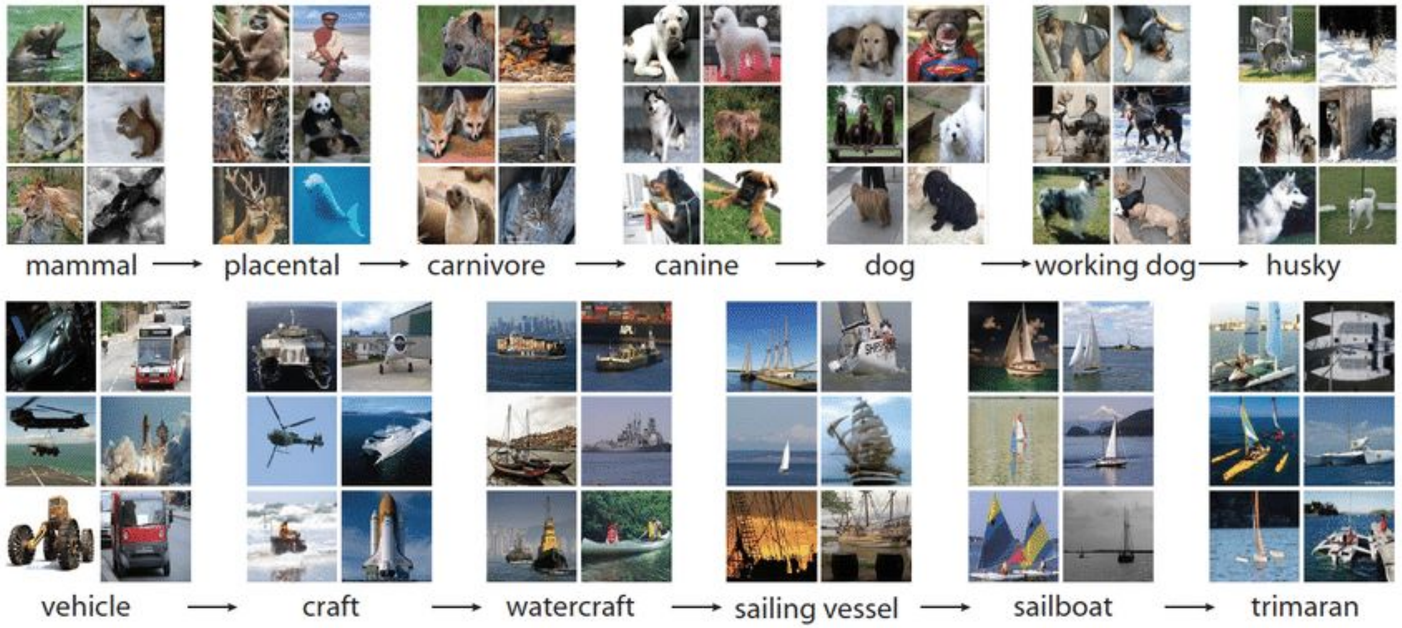
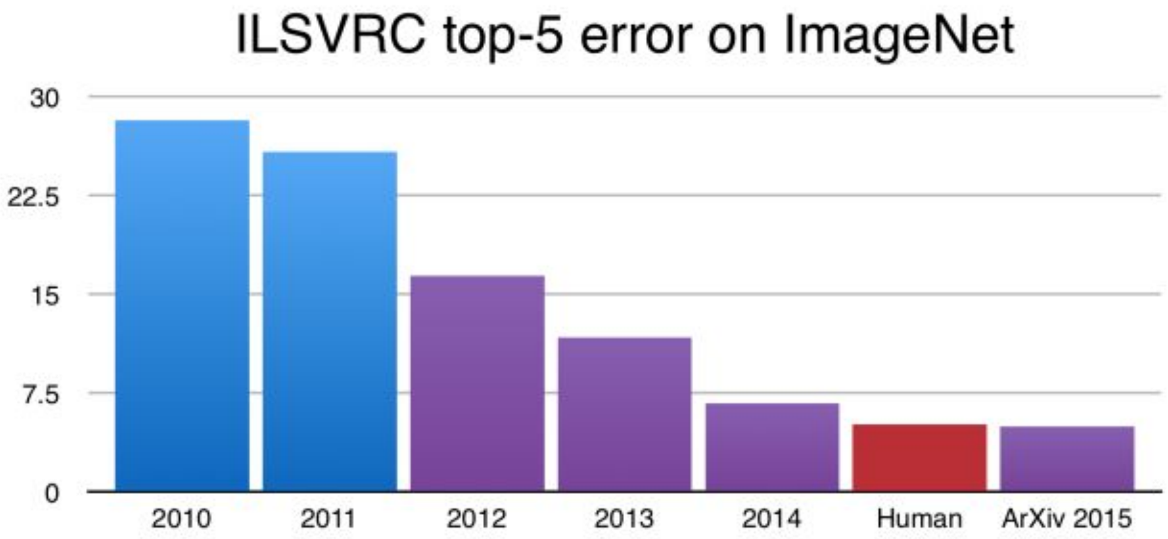


FIGURE 4.1 Architectural graph of a multilayer perceptron with two hidden layers.

http://www.cs.us.es/~fsancho/?e=135

ILSVRC & ImageNet 2012

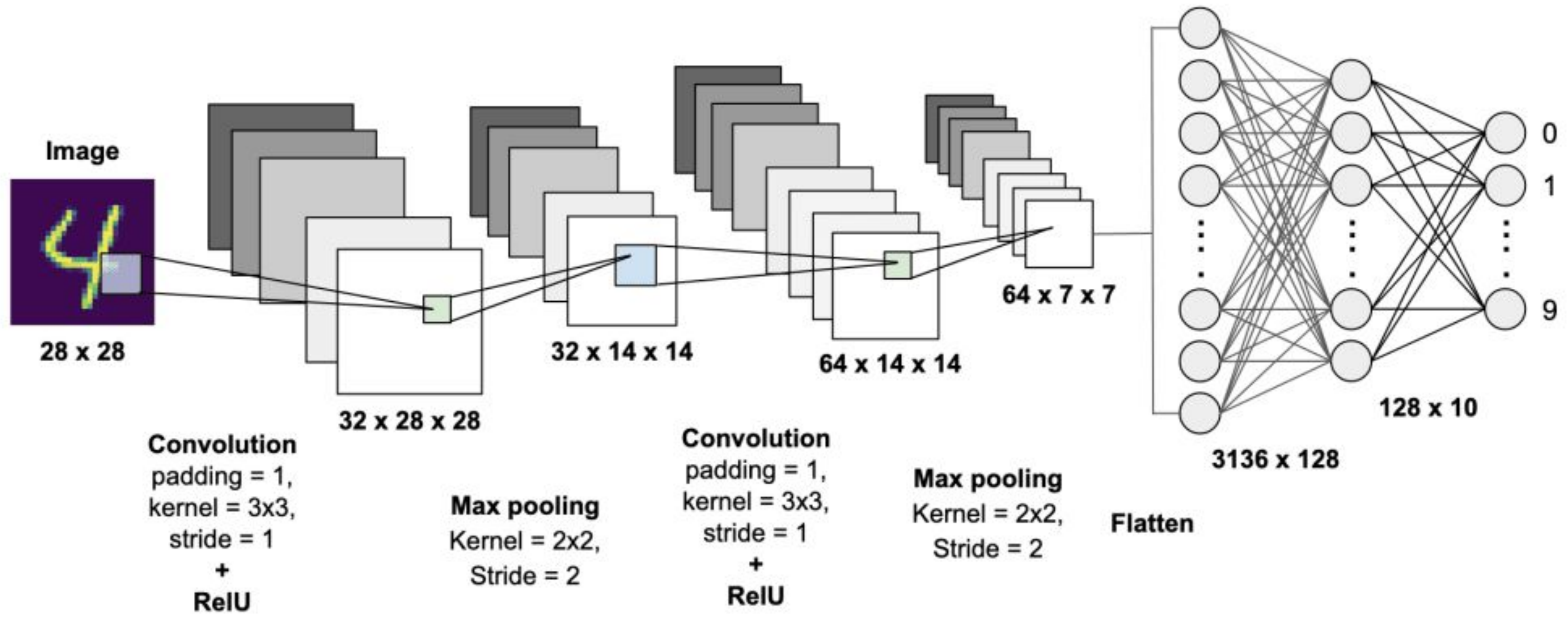




Dealing with images

0	0	0	0	0	0	0	0
0	**1**	**1**	**1**	**1**	**1**	**1**	0
0	0	0	0	0	0	**1**	0
0	0	0	0	0	**1**	0	0
0	0	0	0	0	**1**	0	0
0	0	0	0	**1**	0	0	0
0	0	0	0	**1**	0	0	0
0	0	0	0	0	0	0	0

Convolutional Neural Networks



https://dev.to/afrozchakure/cnn-in-a-brief-27gg

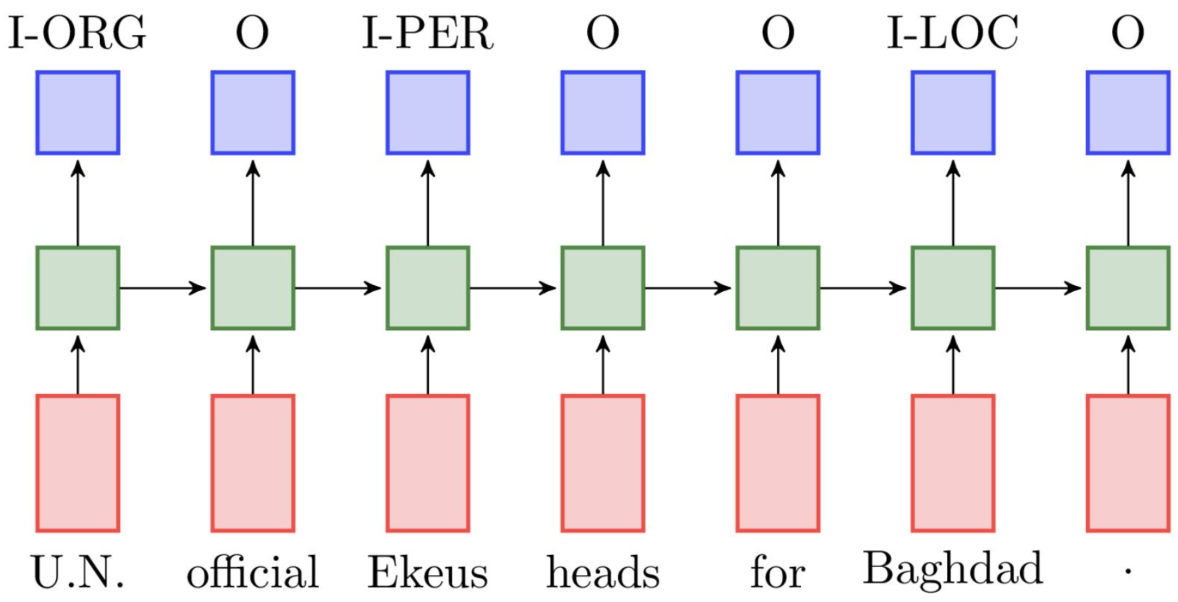
Dealing with text

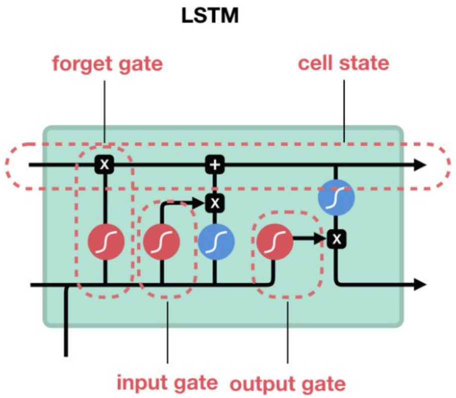
for	you	and	for	me	
1	0	0	1	0	2
0	1	0	0	0	1
0	0	1	0	0	1
0	0	0	0	1	1

Word Level

Sentence Level - Bag of Words

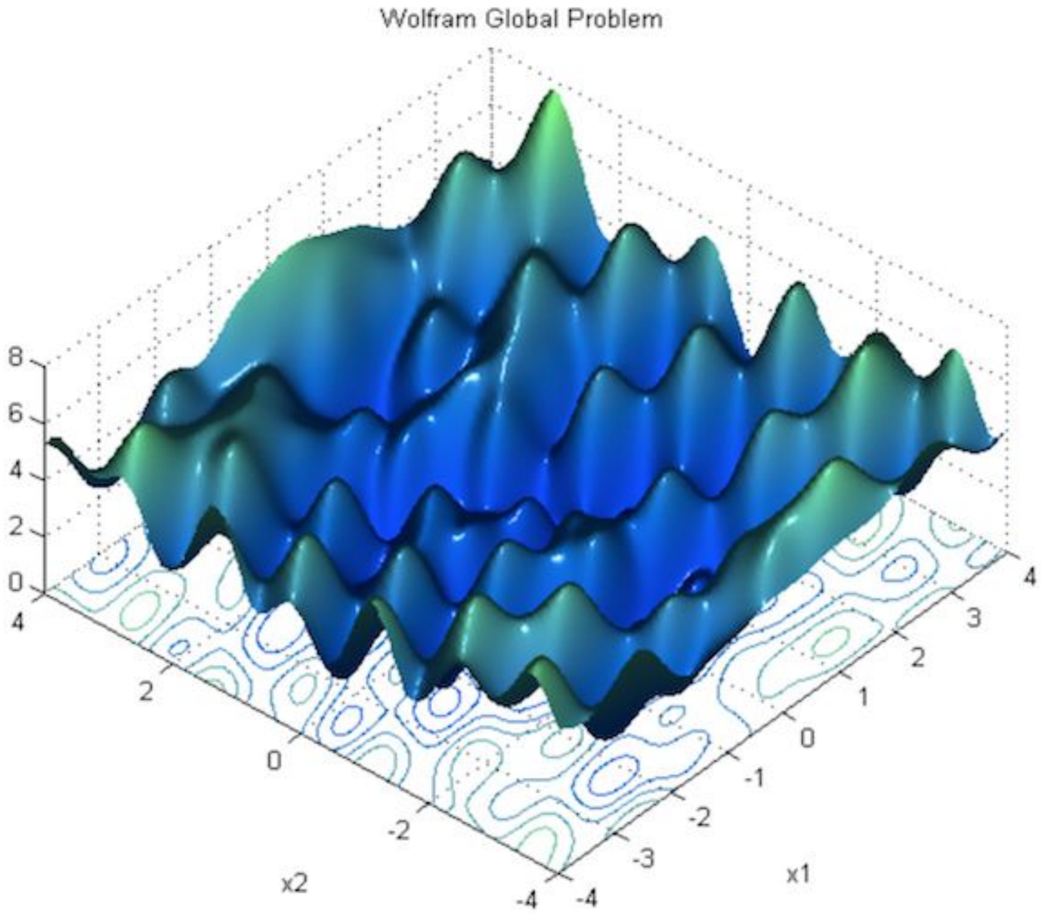
LSTM





https://confusedcoders.com/data-science/deep-learning/how-to-build-deep-neural-network-for-custom-ner-with-keras
https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21

Hard to train



https://ml4a.github.io/ml4a/how_neural_networks_are_trained/

Comments about Deep Learning

- Usually it is harder to code - closer to computer engineering
- Very large model take lots of time, implying high costs (BERT NLP model costed several million US dollars
- Specially designed for image, text, sound, ...

Advanced Machine Learning

Reinforcement Learning

- Exploring new alternatives
- You need to be able to simulate
- Suggest new states based on the past (predicting its reward using machine learning)