ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Artificial Intelligence, Machine Learning and Deep Learning

part 1

Luca Casini

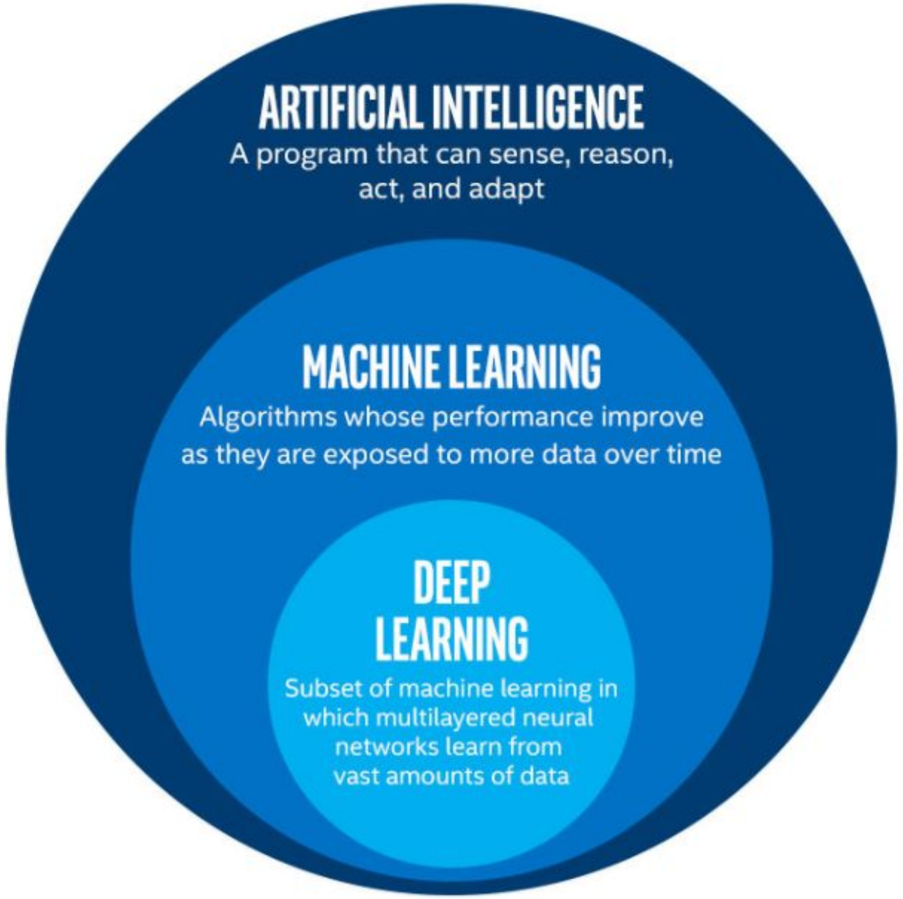
PhD in Data Science and Computation

INTRODUCTION

Introduction

- Artificial intelligence seems to be the greatest technological breakthrough of the third millennium
- Can be applied to virtually any field with incredible results
- Has enabled new applications that were unthinkable 20 years ago
- While many are enthusiastic about it, some fear it may be highly problematic
- Still AI has a very long history dating back to the 1950s
- Many different technologies have been called AI, the last one is deep learning
- AI has gone in and out of fashion many times throughout the years
- The terms AI, Machine Learning and Deep Learning seem to be used interchangeably but they should not

Introduction



- AI is broad term and describes anything that can has the goal of mimicking human behaviour
- Machine Learning is a field at the intersection of computer science, statistics and mathematics. The goal is creating models that can make predictions by learning from data
- Deep Learning is a sub-field of ML and is about using deep neural networks to push and extend the capabilities of computer beyond classical ML models
- **This course will mainly focus on Deep Learning and current state of the art**

Brief History of AI

1950–1956: The Beginning

- Turing was already imagining thinking machines. Science fiction was also starting to reference this.
- first experiments on computer programs that were able to reason and learn
- first ideas of artificial neurons and neural networks (McCulloch & Pitts)
- first AI games (Strachey, Samuel)

Chapter 3
Computing Machinery and Intelligence

Alan M. Turing

Editors' Note: The following is the article that started it all – the article by Alan Turing which appeared in 1950 in the British journal, *Mind*. Accompanying the article are three running commentaries by Kenneth Ford, Clark Glymour, and Pat Hayes of the University of West Florida; Stevan Harnad of the University of Southampton; and Ayse Pinar Saygin of the University of California, San Diego, designated respectively by the symbols: ♠, ♣, and ♥. A fourth commentary by John Lucas of Merton College, Oxford, is found in Chapter 4.

3.1 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

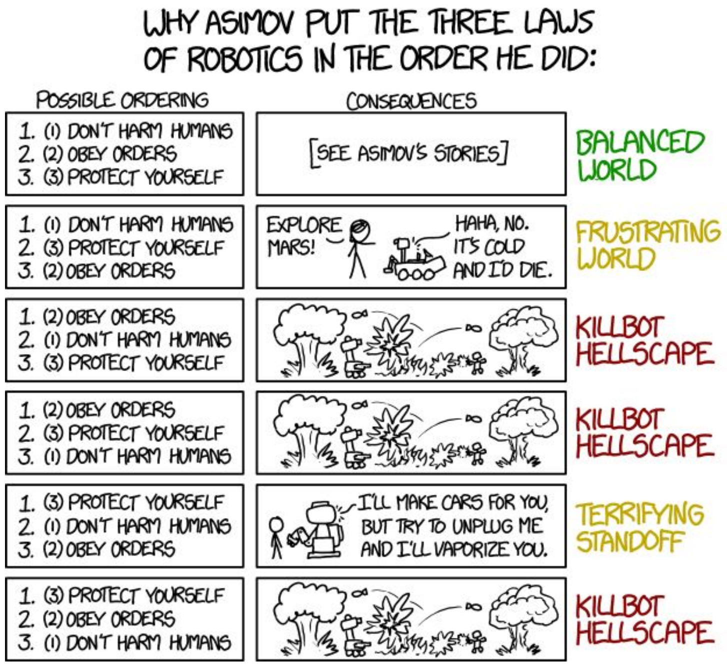
Manchester University

♣ Harnad: Turing starts on an equivocation. We know now that what he will go on to consider is not whether or not machines can think, but whether or not machines can do what thinkers like us can do – and if so, how. Doing is performance capacity, empirically observable. Thinking is an internal state. It correlates empirically observable as neural activity (if we only knew which neural activity corresponds to thinking!) and its associated quality introspectively observable as our own mental state when we are thinking. Turing's proposal will turn out to have nothing to do with either observing neural states or introspecting mental states, but only with generating performance capacity indistinguishable from that of thinkers like us.

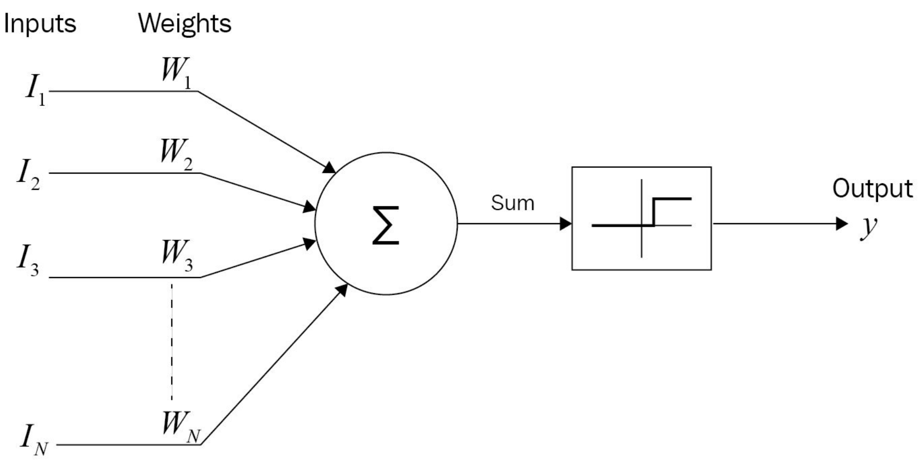
Brief History of AI

1950–1956: The Beginning

- Turing was already imagining thinking machines. Science fiction was also starting to reference this.
- first experiments on computer programs that were able to reason and learn
- first ideas of artificial neurons and neural networks (McCulloch & Pitts)
- first AI games (Strachey, Samuel)

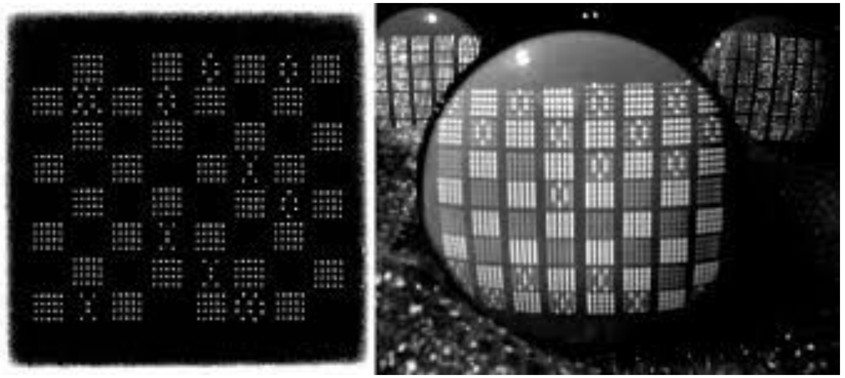


Brief History of AI

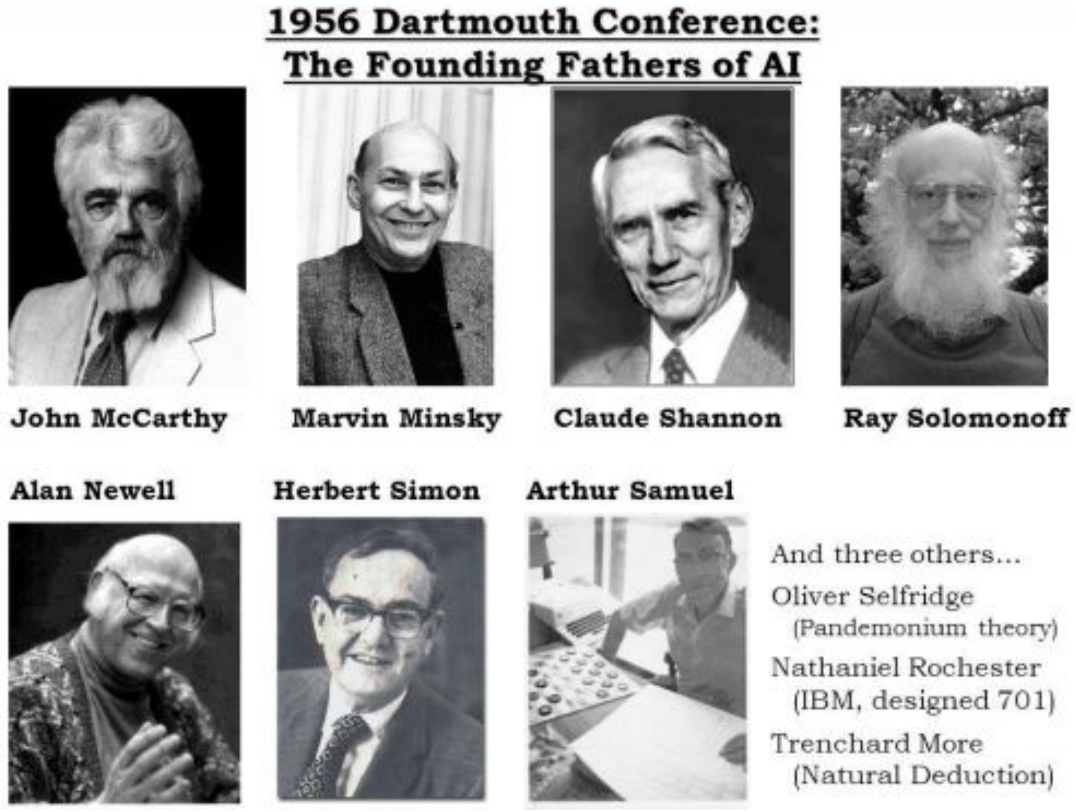


1950–1956: The Beginning

- Turing was already imagining thinking machines. Science fiction was also starting to reference this.
- first experiments on computer programs that were able to reason and learn
- first ideas of artificial neurons and neural networks (McCulloch & Pitts)
- first AI playing a game of Checkers (Strachey, Samuel)



1956 Dartmouth Conference



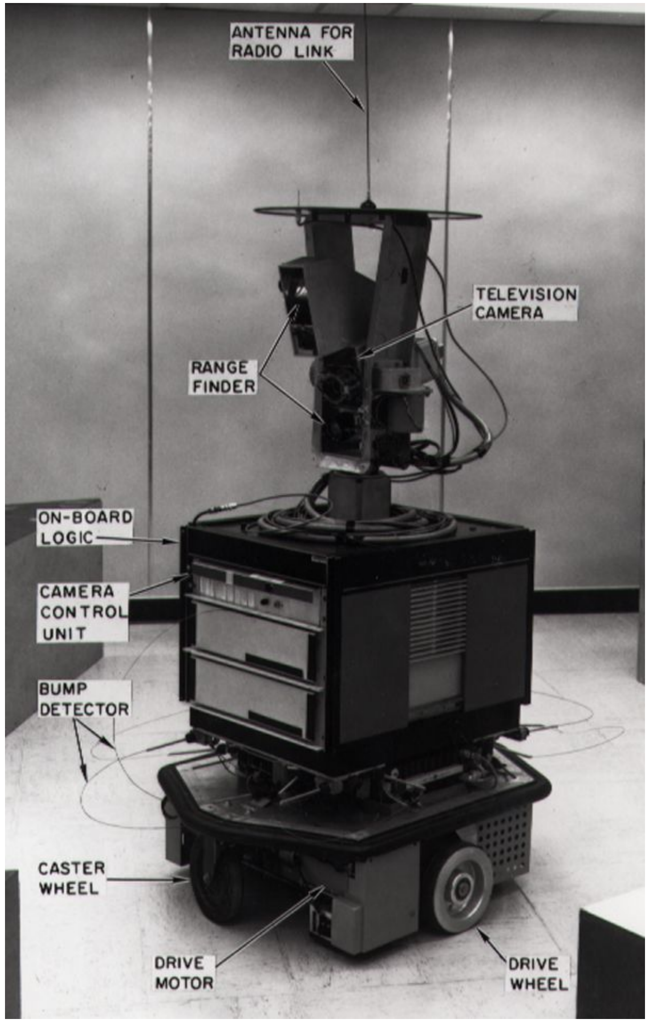
1956 Dartmouth Conference is where the term Artificial Intelligence was coined.

"We propose a study of artificial intelligence [..]. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it."

Brief History of AI

1956–1974: The Golden Age

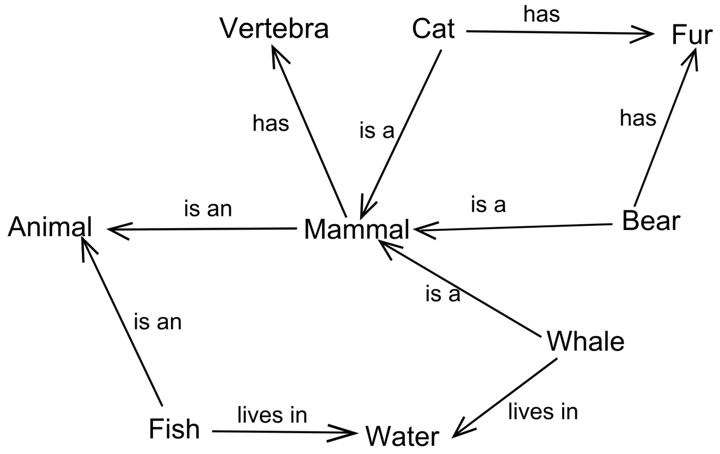
- GOFAI, symbolic representation and search, LISP language, semantic networks
- first experiments with language (first chatbot)
- government founding and high optimism
  - 1958, H. A. Simon and Allen Newell: "*within ten years a digital computer will be the world's chess champion*" and "*within ten years a digital computer will discover and prove an important new mathematical theorem.*"
  - 1965, H. A. Simon: "*machines will be capable, within twenty years, of doing any work a man can do.*"
  - 1967, Marvin Minsky: "*Within a generation ... the problem of creating 'artificial intelligence' will substantially be solved.*"
  - 1970, Marvin Minsky (in Life Magazine): "*In from three to eight years we will have a machine with the general intelligence of an average human being.*"

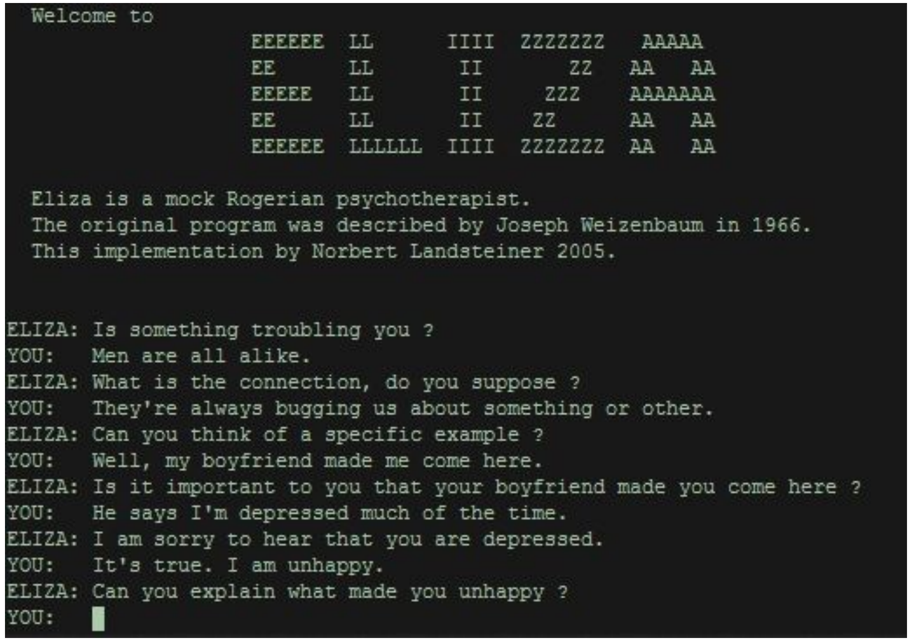


Brief History of AI

1956–1974: The Golden Age

- GOFAI, symbolic representation and search, LISP language, semantic networks
- first experiments with language (first chatbot)
- government founding and high optimism
    - 1958, H. A. Simon and Allen Newell: "*within ten years a digital computer will be the world's chess champion*" and "*within ten years a digital computer will discover and prove an important new mathematical theorem.*"
    - 1965, H. A. Simon: "*machines will be capable, within twenty years, of doing any work a man can do.*"
    - 1967, Marvin Minsky: "*Within a generation ... the problem of creating 'artificial intelligence' will substantially be solved.*"
    - 1970, Marvin Minsky (in Life Magazine): "*In from three to eight years we will have a machine with the general intelligence of an average human being.*"





Brief History of AI

1974–1980: First AI Winter

- funding was cut due inflated expectations and not enough results
- problem of complexity and lack of computational resources
- neural nets don’t work and are difficult to train

1980–1987: Commercial Success

- rise of the expert systems and knowledge representation, big business success
- in the meantime: revival of neural networks, backpropagation solves training

1987–1993: Second AI Winter

- Expert systems were expensive to create and maintain, and not always effective. Also general purpose computers were becoming more performing and available
- Over 300 AI companies had shutdown, gone bankrupt, or been acquired by the end of 1993
- many scientist started rejecting the symbolic approach

Brief History of AI

1993–Present: Modern AI

- Computational power greatly increased allowing for new milestones (1997 IBM Deep Blue, 2011 IBM Watson Jeopardy!)
- The field slowly become more rigorous, integrating more sophisticated mathematical tools
- With the growth of the internet and availability of (big) data in the 21st century machine learning started gaining popularity. Neural networks showed potential, especially in vision related tasks
- The use of NN with many neurons stacked spawned the field of Deep Learning. DL keeps pushing the state of the art in most fields of application

Brief History of AI

1993–Present: Modern AI

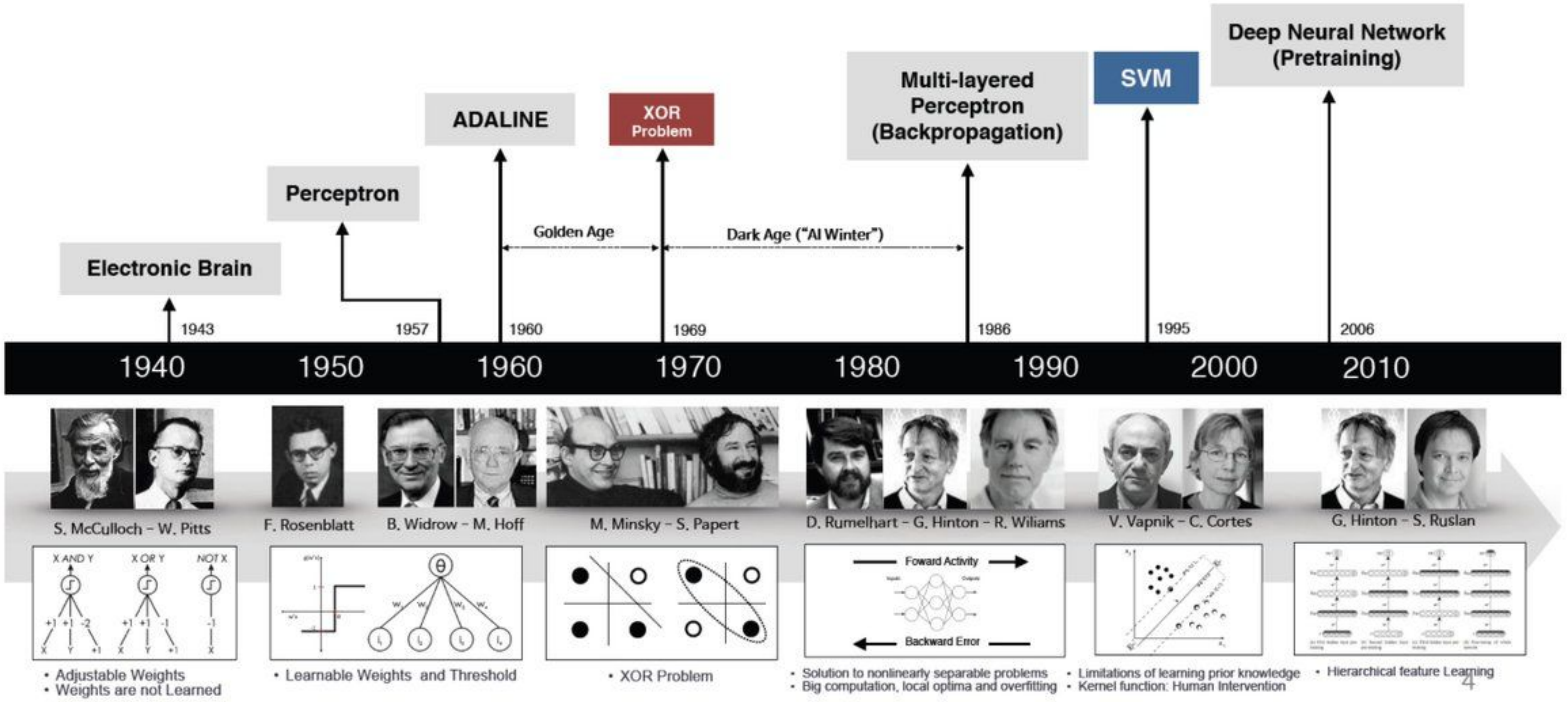
- Computational power greatly increased allowing for new milestones (1997 IBM Deep Blue, 2011 IBM Watson Jeopardy!)
- The field slowly become more rigorous, integrating more sophisticated mathematical tools
- With the growth of the internet and availability of (big) data in the 21st century machine learning started gaining popularity. Neural networks showed potential, especially in vision related tasks
- The use of NN with many neurons stacked spawned the field of Deep Learning. DL keeps pushing the state of the art in most fields of application

Brief History of AI

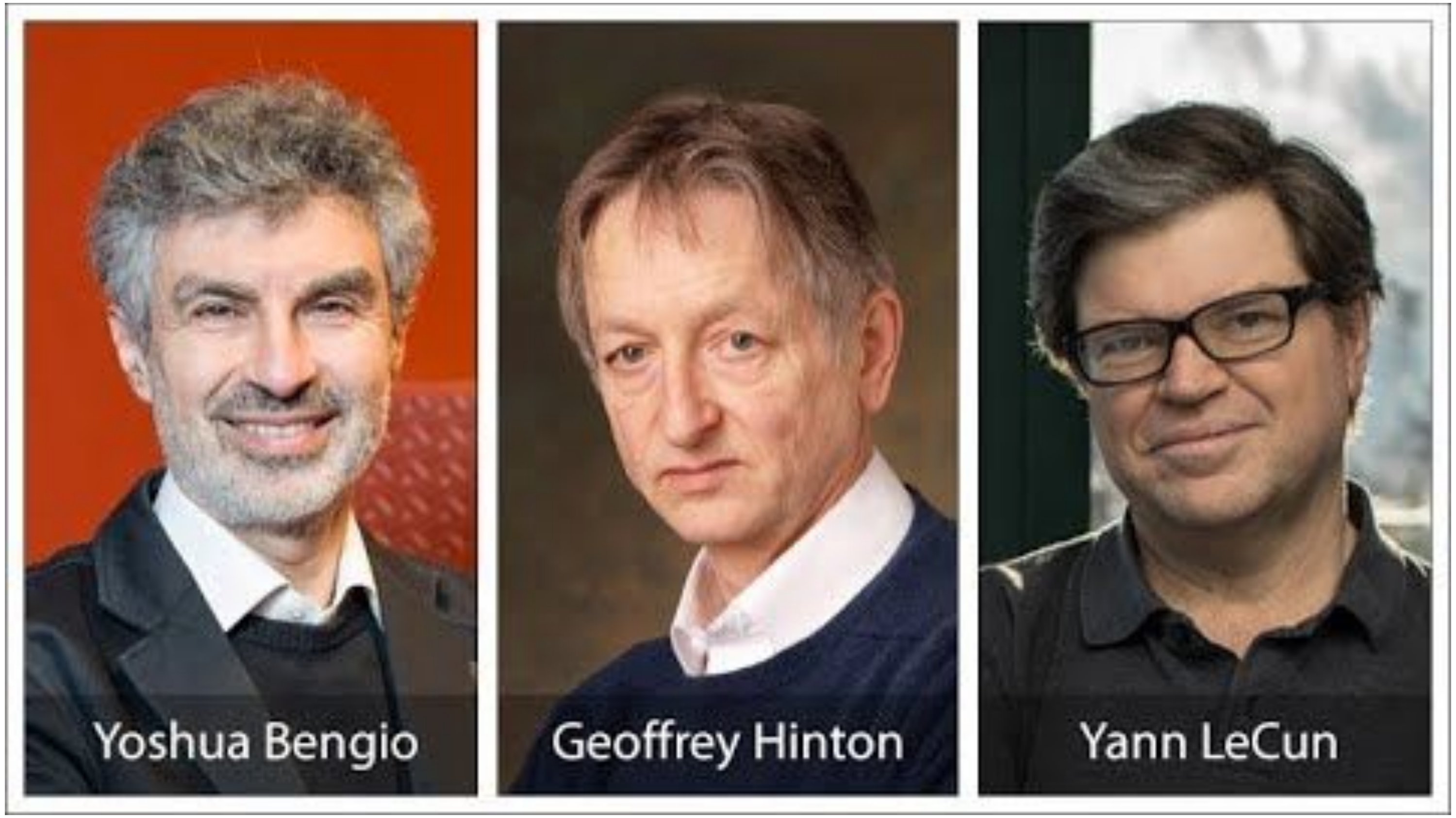
1993–Present: Modern AI

- Computational power greatly increased allowing for new milestones (1997 IBM Deep Blue, 2011 IBM Watson Jeopardy!)
- The field slowly become more rigorous, integrating more sophisticated mathematical tools
- With the growth of the internet and availability of (big) data in the 21st century machine learning started gaining popularity. Neural networks showed potential, especially in vision related tasks
- The use of NN with many neurons stacked spawned the field of Deep Learning. DL keeps pushing the state of the art in most fields of application

Brief history of neural nets



2018 Turing Prize



"For conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing"

MACHINE LEARNING
&
DEEP LEARNING

What is Machine Learning?

Machine Learning is a field of Computer Science focused on the creation of deterministic algorithms that are able to extract some kind of knowledge from data.

A machine learning problem can be defined formally as triplet **`<T,P,E>`:**

- A task **T** to complete
- A performance metric to be maximized **P**
- The experience **E** accumulated by the system

A machine is then said to be learning if as **E** increases so does **P**

Machine Learning

In other words:

- The task **T** is obviously the problem we want to solve. e.g. Classification, Regression, etc.
- The performance metric **P** is usually an error measure to be minimized, also called **Loss.**
- The experience **E** is represented by iterating over the **dataset.**
    - Good data is key to obtaining good performance. Volume is not the only factor.

There are many machine learning algorithms and techniques, all based on linear algebra, probability and statistics. For example:

- Linear Regression is plain linear algebra: find a curve that fits the data minimizing the error.
- Other models are based on probability and try to learn the distribution of the data and consequently give outputs that are probabilistic.

We are going to focus on **Neural Networks,** a biologically inspired method that can be seen as a universal function approximator.

Supervised vs. Unsupervised

Supervised learning:

- data must be labeled
- clearly defined problem and metrics
- e.g. classification, regression
- works great but depends on the quality and quantity of data

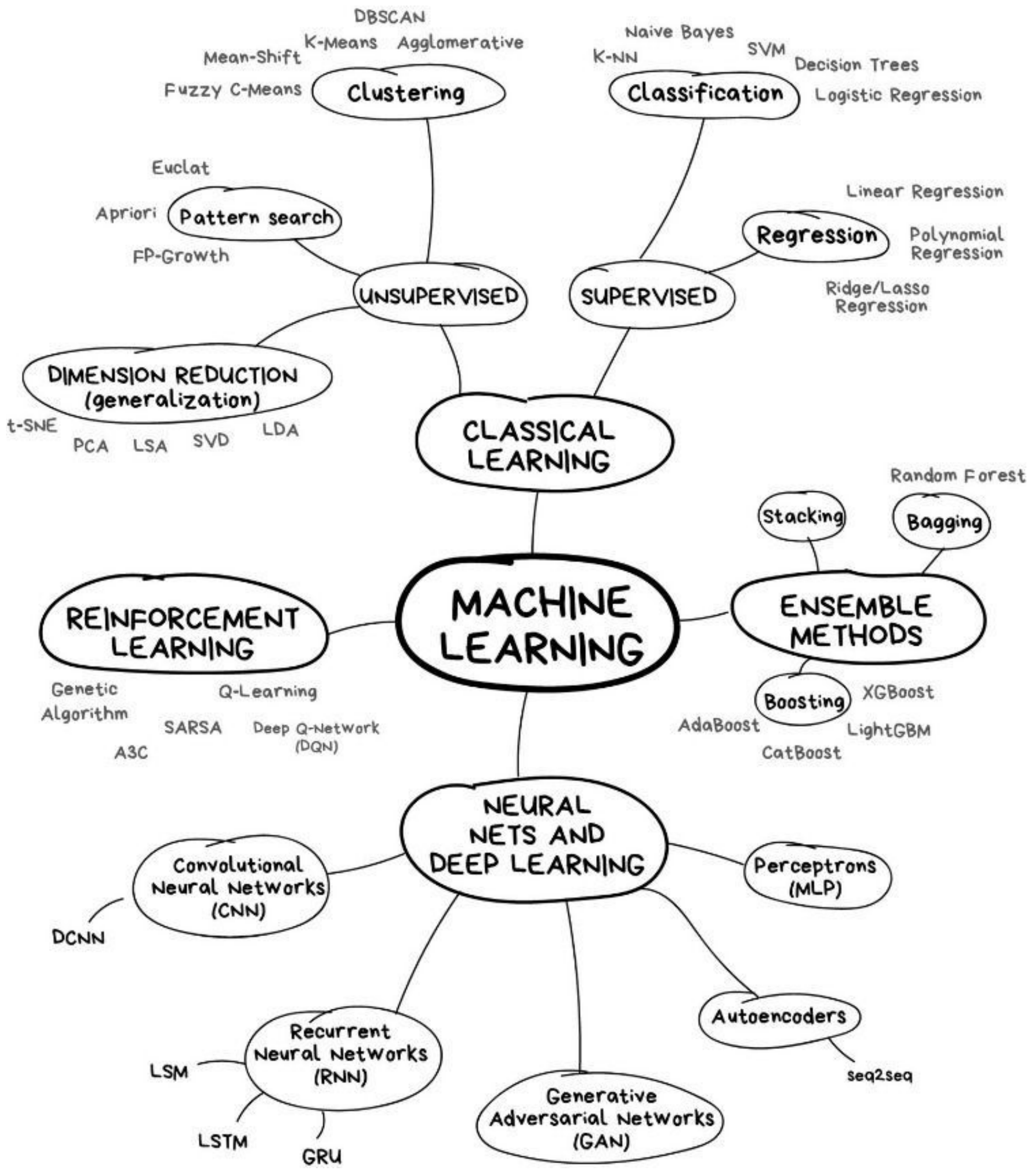
Unsupervised Learning

- data is unlabeled
- we are trying to discover relationship in the data, if any.
- e.g. clustering
- more flexible but many need a lot of data/computational power

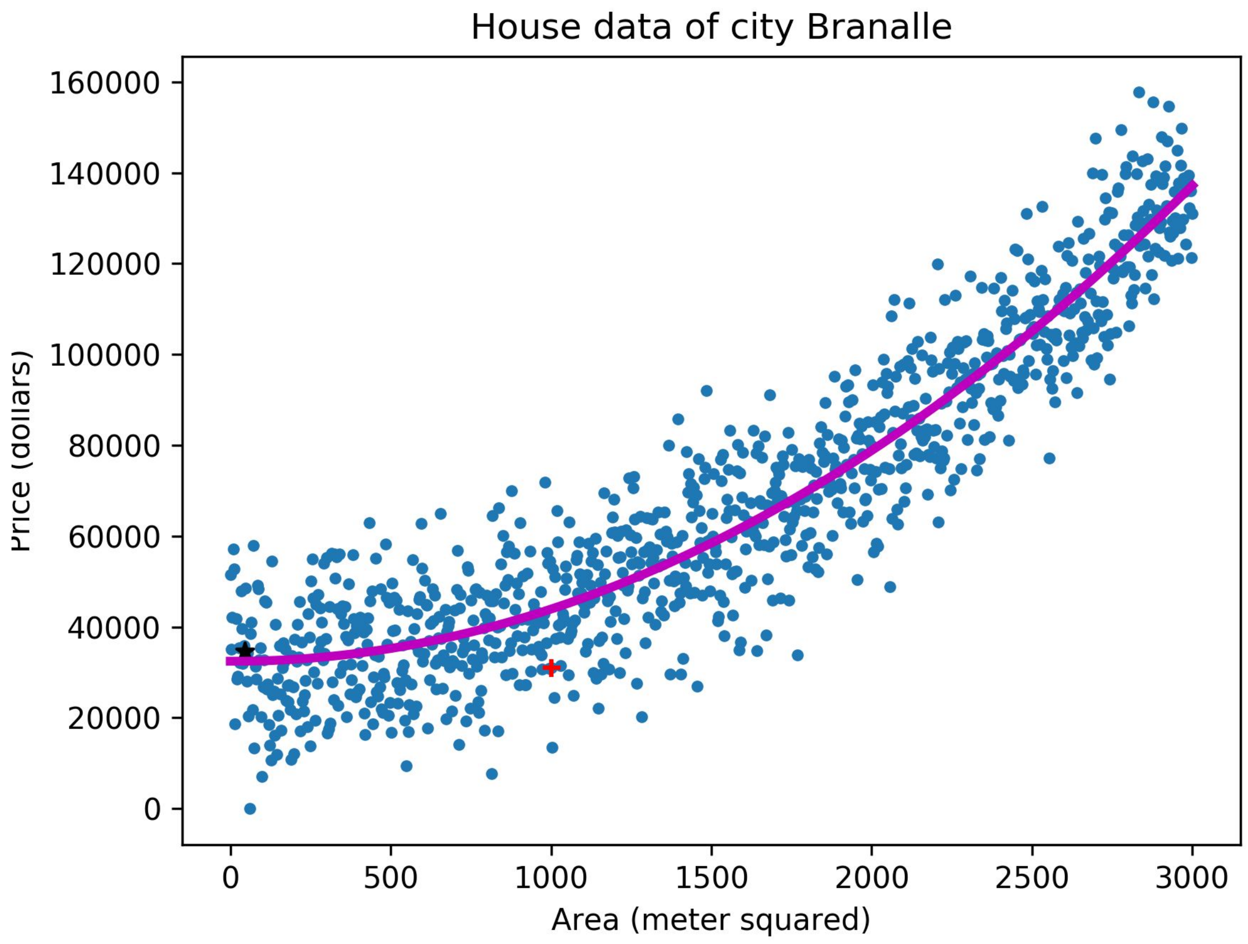
Reinforcement Learning

- the goal is learning to perform an action in the world
- when the action is performed correctly a reward is given
- used for games and robots, similar to how mice are trained

Map of Machine Learning

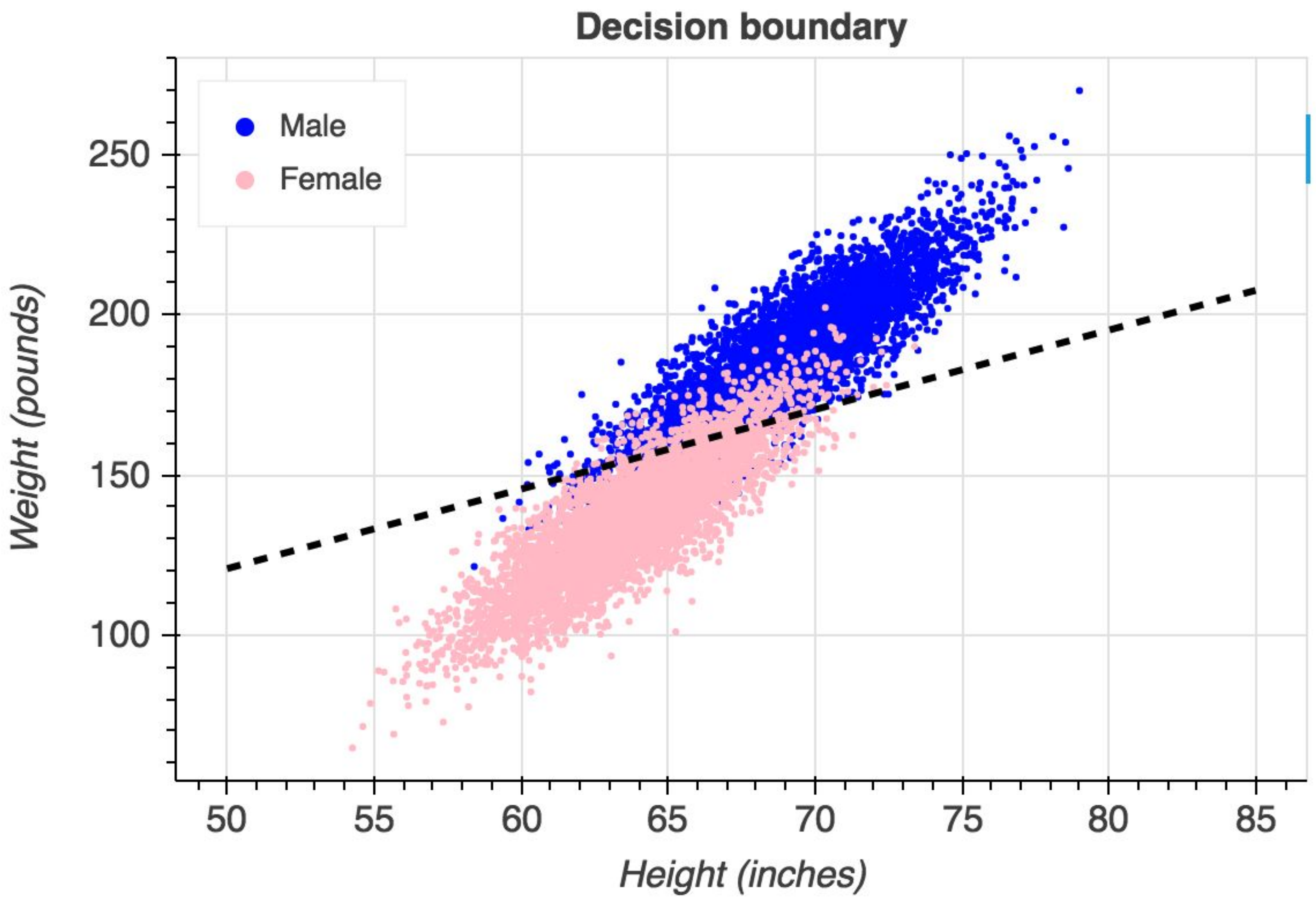


Regression



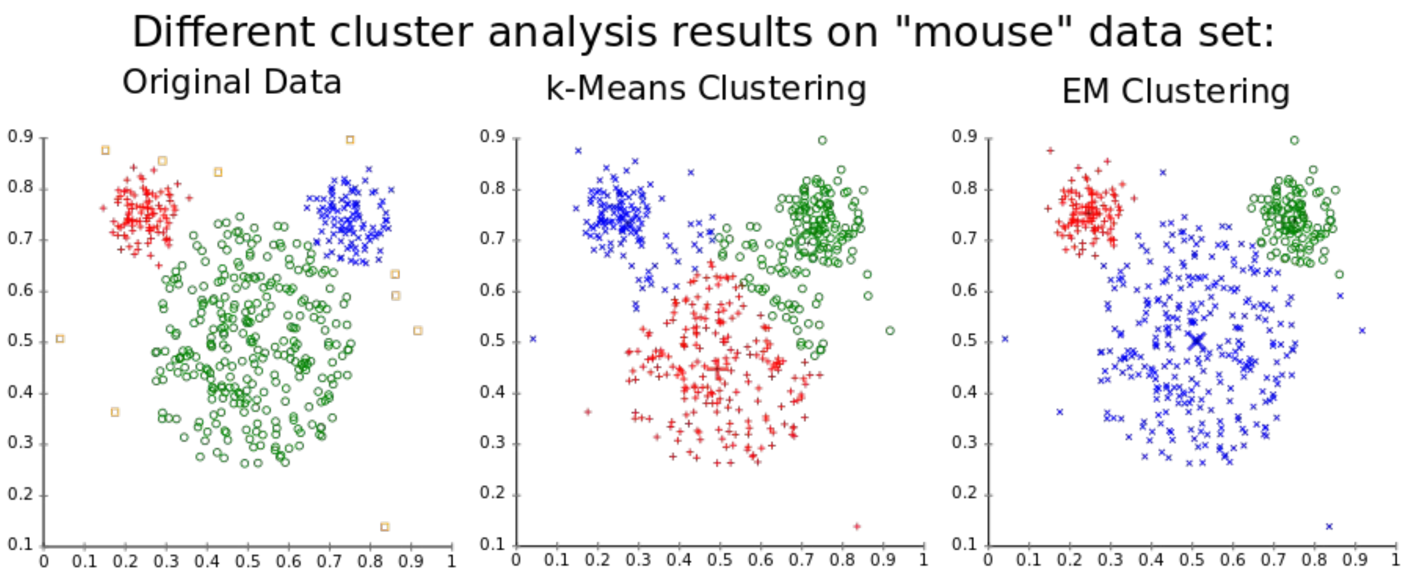
- Maybe the most common task of all
- given a dataset we try to learn the trend of the output
- when a new input is presented we can predict the variable of interest
- e.g: given the area of a house we can predict its price within a margin of error

Classification



- Classification is actually quite similar to regression
- This time we are trying to learn the line that best divides the data in two (or more) categories
- e.g: given the height and weight of a person we can predict the sex within a margin of error

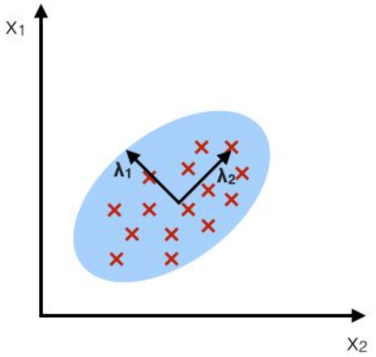
Clustering



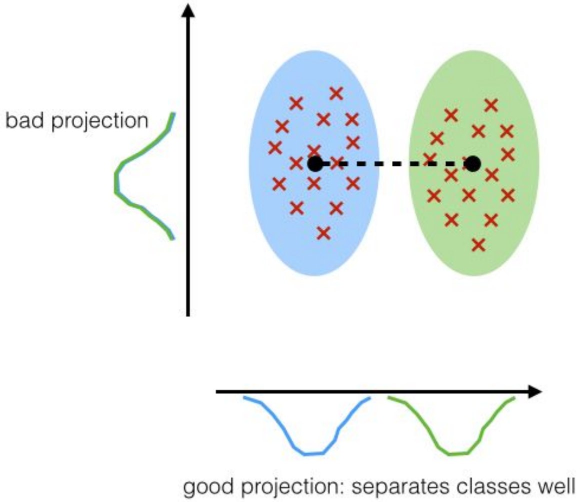
- Clustering algorithms are used to group similar data points based on their variables in an unsupervised way.
- A practical example is trying to classify user behaviour in an online shop

Dimensionality Reduction

PCA:
component axes that maximize the variance

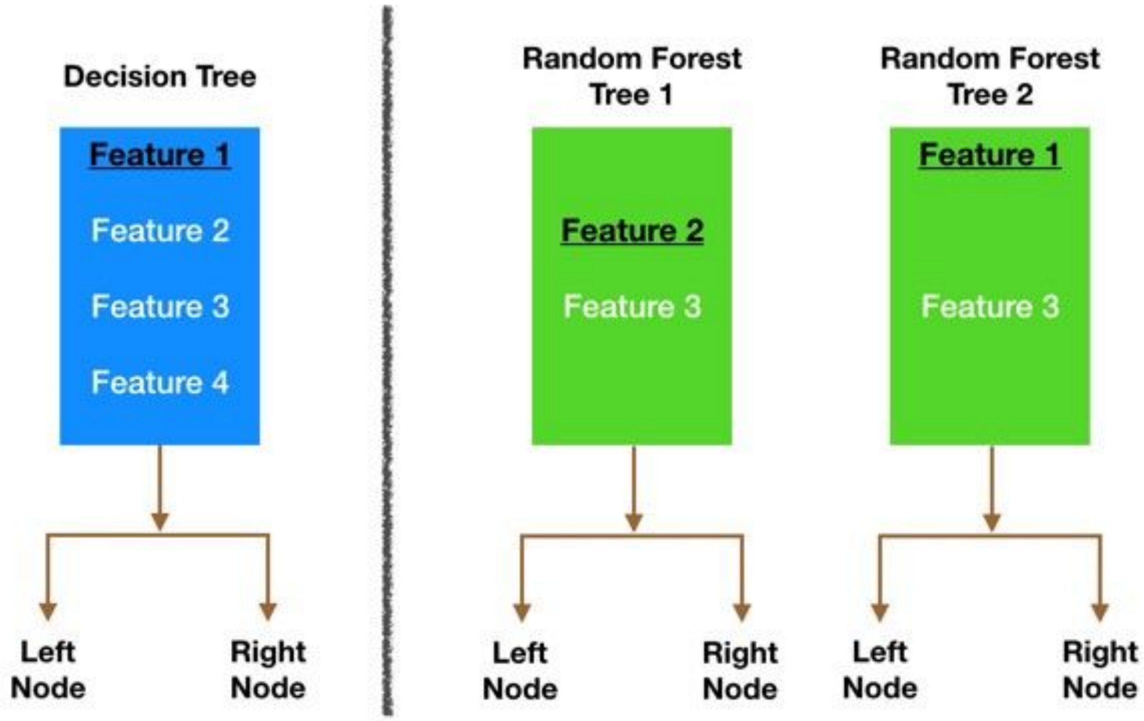


LDA:
maximizing the component axes for class-separation



- Sometimes the features in a dataset are not very useful
- To many dimensions are also difficult to manage
- Dimensionality reduction techniques have the goal of removing/combining variables or manipulating their space in order to create a better dataset

Ensemble Methods



- Ensemble methods are those models composed of many, small, weaker, models
- their predictions are combined in order to obtain the final output of the model
- each predictor is trained on a different subset of the data
- the results is that each model is independent and can actually cover the errors of the others

Reinforcement Learning

- When it come to Reinforcement learning a video is worth a thousand words (and images)
- The basic idea is having an agent perform the same action thousand of times and rewarding good behaviour
- in time the agent should converge towards the ultimate goal
- e.g. parking the car

Generative vs. Discriminative Models

Another way to divide machine learning models is between generative and discriminative:

- A **discriminative model** learns how the data is **separated.**
  - The model is effectively learning what it means to be part of one class or another.

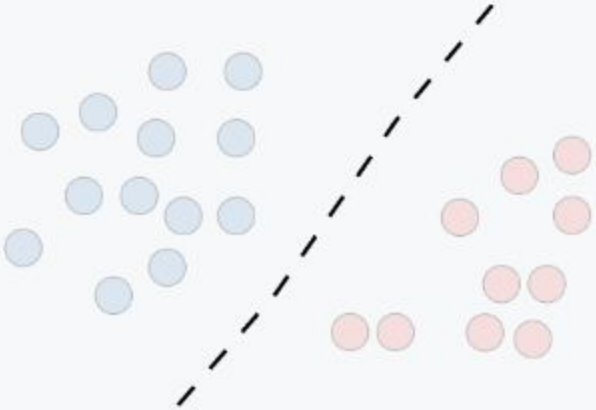
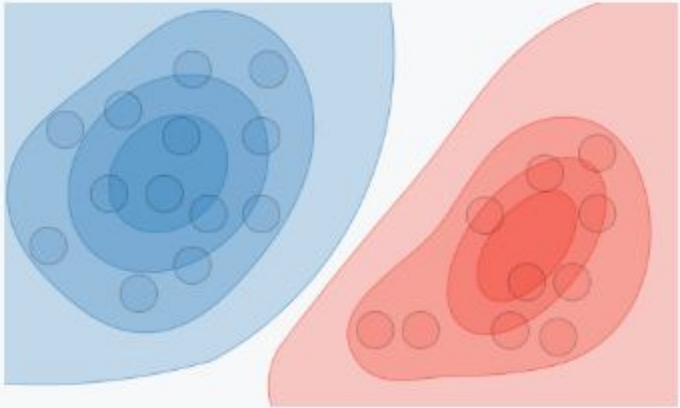
- A **generative model** is a model that learns how any specific kind of data is **distributed**.
  - This means that we can sample a specific distribution to generate something similar to what we learned

In tasks like classification, discriminative models work best because, given a large enough amount of data, they approximate directly $P(Y|X)$ without assumption on the distribution.

On the other hand generative models make assumption on the distributions of X and Y, usually gaussians, trading off precision for the ability to sample data.

http://ai.stanford.edu/~ang/papers/nips01-discriminativegenerative.pdf

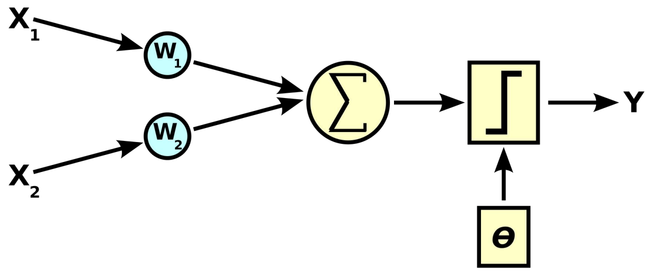
Generative vs. Discriminative models

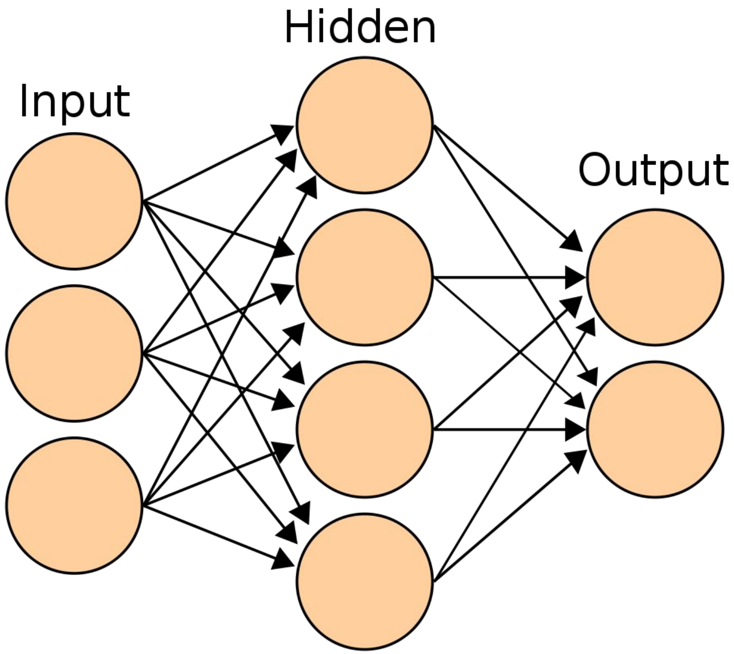
	Discriminative model	Generative model
Goal	Directly estimate $P(y\|x)$	Estimate $P(x\|y)$ to then deduce $P(y\|x)$
What's learned	Decision boundary	Probability distributions of the data
Illustration		
Examples	Regressions, SVMs	GDA, Naive Bayes

https://stanford.edu/~shervine/teaching/cs-229/cheatsheet-supervised-learning

What about Neural Nets?

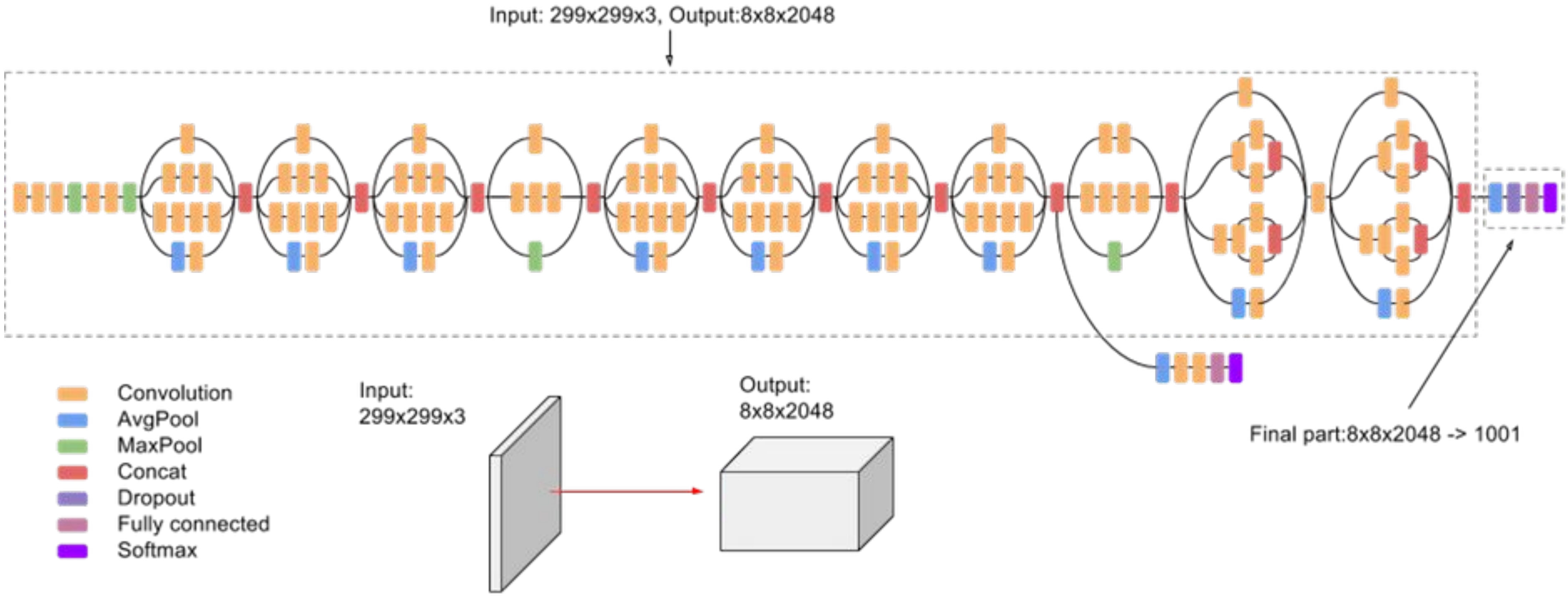
- Neural networks are one of the tools of machine learning
- **Artificial Neural Networks (ANN)** started in the 40s with introduction of a model of the neuron and further work led to the **Perceptron Model.**
  - Each neuron combines its inputs and passes them through and activation function
- In 1974 **Multilayer** version goes under the radar.
- The **backpropagation** was developed in 1986 algorithm allowed efficient training.
  - in the **forward** pass. Inputs are propagated and the outputs is used to compute the error (loss function)
  - in the **backward** pass errors are backpropagated to each neuron for **gradient descent**





why deep learning?

because it is!

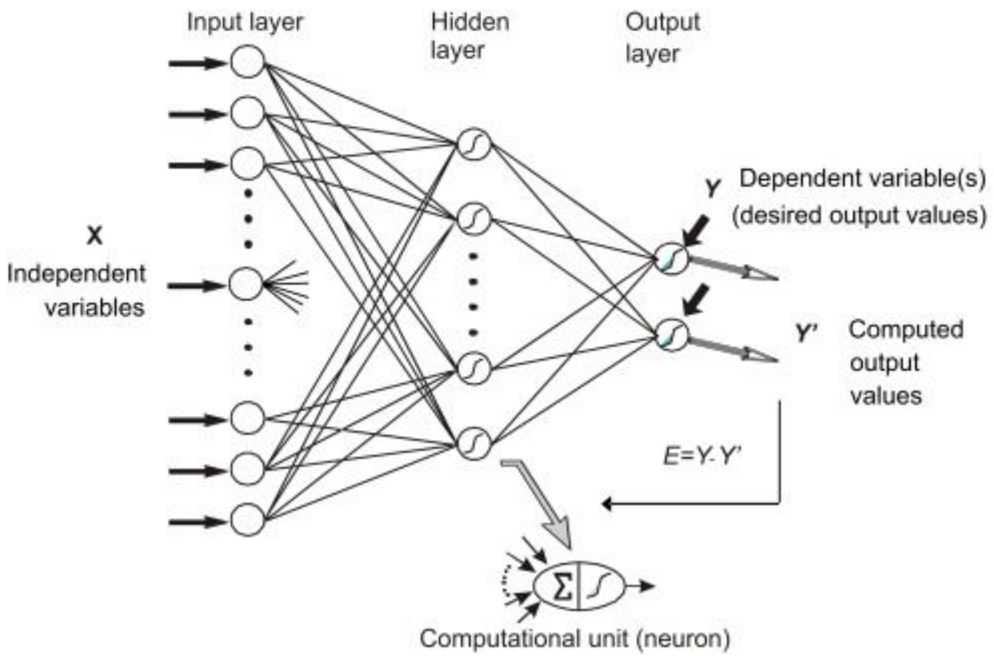


Types Of Neural Networks

There are many types of neural networks and architectures, we will briefly see the most common:

- Multilayer Perceptron
- Convolutional Neural Networks
- Recurrent Neural Networks
- (Variational) Autoencoders
- Generative Adversarial Networks
- Transformers

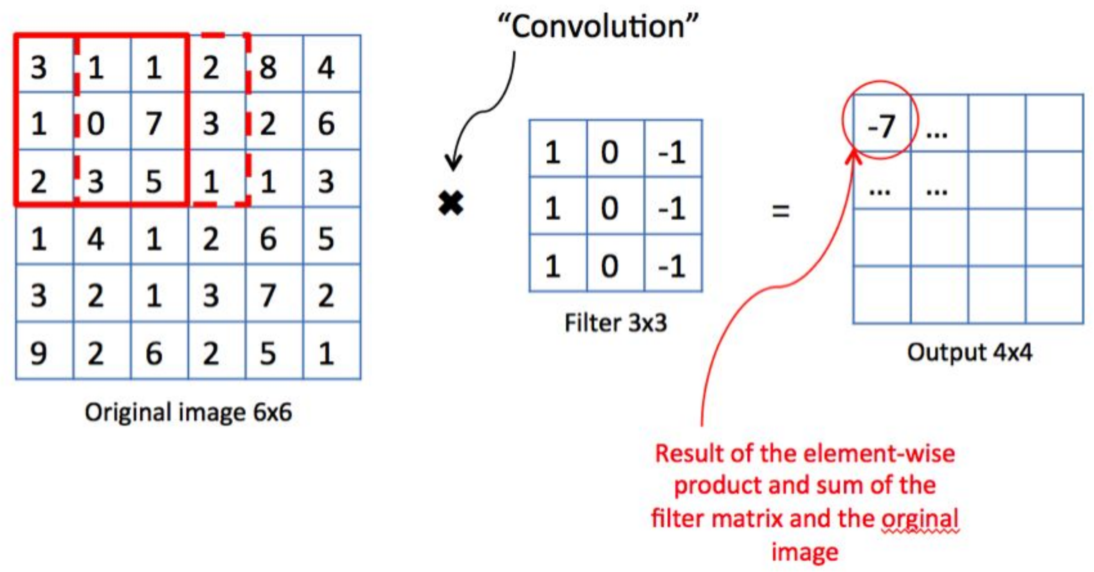
Multilayer Perceptron



- every neuron is connected
  - for this reason is also called a Dense Layer
- is the most basic form of neural network
- it becomes difficult to use when:
  - the input has too many variables/dimensions (images)
  - It has relationship through times (text, audio)
- the limitation of the MLP architecture led to the development of Convolutional Nets and Reccurent Nets

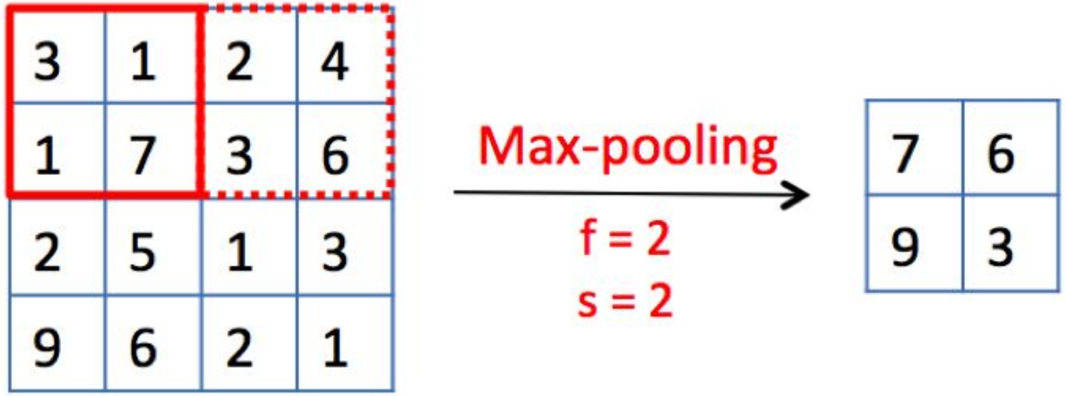
Convolutional Neural Network

- CNN were introduced in the 90s and have become the staple of image recognition
- They get their names from the convolution operation.
    - **Basically a filter (kernel) is slid across the image. stride and size are parameters**
    - after that max-pooling is used to reduce dimensionality
- Good results can be achieved with any kind of data if represented in matrix form.



Convolutional Neural Network

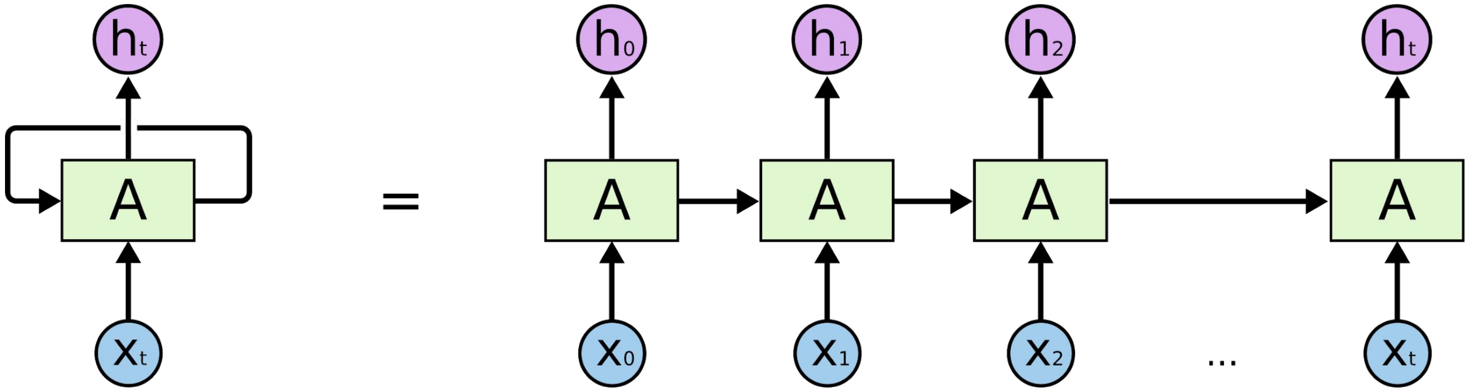
- CNN were introduced in the 90s and have become the staple of image recognition
- They get their names from the convolution operation.
  - Basically a filter (kernel) is slid across the image. stride and size are parameters
  - **after that max-pooling is used to reduce dimensionality**
- Good results can be achieved with any kind of data if represented in matrix form.



[vdumoulin/conv_arithmetic: A technical report on convolution arithmetic in the context of deep learning]()

Recurrent Neural Networks

- Temporal sequences are problematic for simple feed-forward networks. They need a way to "remember" what happened before.
- Recurrent Neural Networks tried to solve the problem by feeding back the output the input after each timestep
- Accumulation of input lead to the "vanishing gradient" problem. Long Short-Time Memory architecture (LSTM) solved the problem introducing gated inputs and outputs



[Understanding LSTM Networks]

Transformers

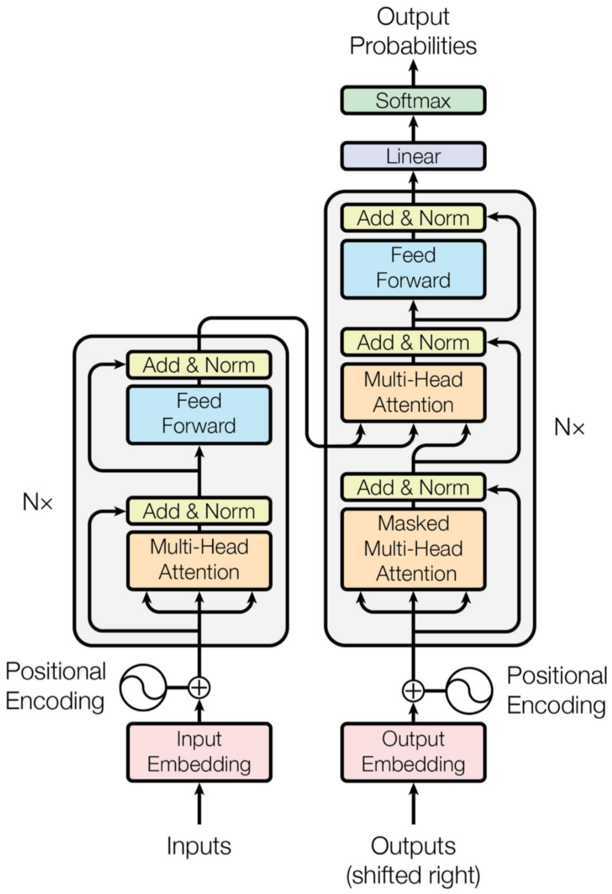


Figure 1: The Transformer - model architecture.

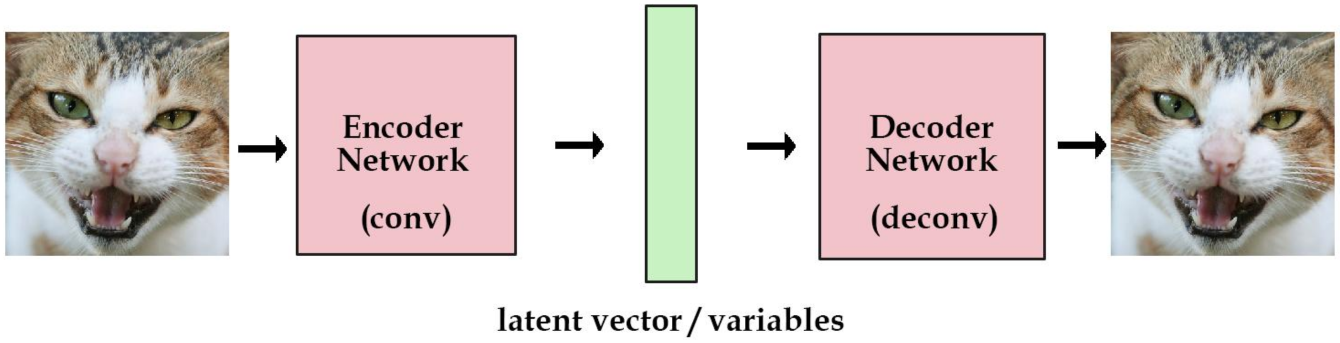
- State-of-the-art for sequence data like text, music, etc.
- Instead of recurrent inputs it uses a mechanism called Attention that learn where in the sequence to look
- Different from RNN it has the possibility of being stacked and parallelized very easily, allowing for massive scale
- The most famous application of it may be GPT-2, a model for text generation that is extremely believable in its outputs

[Music Transformer: Generating Music with Long-Term Structure]

[Better Language Models and Their Implications]

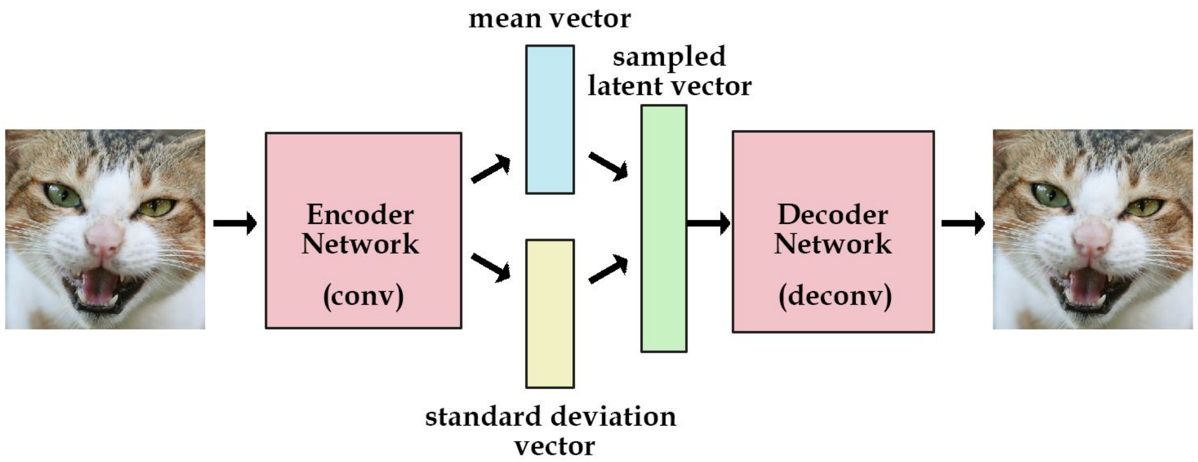
Autoencoders

- Autoencoders, or siamese networks, are neural networks with to specular parts: The first half maps the input to a lower-dimensional tensor and the last part tries to reconstructs the input from it.
- It essentially works as a compression algorithm
- learned latent space can be analyzed and leveraged for classification and other tasks
- If the output target is different from the input il can be used for translation tasks, for languages as well as other semantically connected inputs



Variational Autoencoders

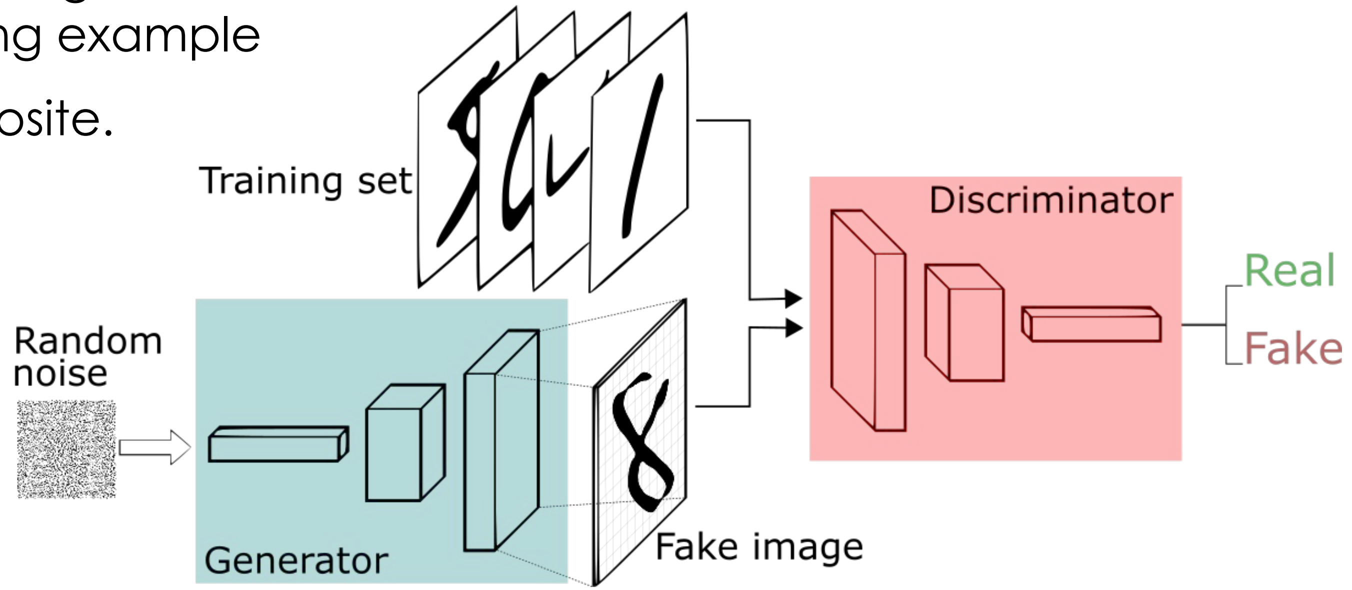
- Variational Autoencoder (VAE) are becoming very popular as generative models.
- They are a combination of traditional Autoencoders with Bayesian Variational Models
- Autoencoders are neural networks that take an input, reduce it's dimensionality to a latent space, and then learn how to reconstruct the original input starting from there
- After the training the second half can be used to reconstruct any input. The first half works like a compression algorithm
- VAEs are interesting because they learn to approximate how data is distributed. They can create new data in addition to recreate training examples
- The latent space is also semantically structured



http://kvfrans.com/variational-autoencoders-explained/

Generative Adversarial Networks

- GANs are relatively new models and attracted a lot of attention for their ability to generate very convincing images
- The basic idea is to have to networks fight against each other:
  - The generator creates data from random noise
  - The discriminator has to decide if the generated data is real or fake against a training example
- The two networks loss function are opposite. it's a zero-sum game
- During training is better to optimize in turns
- results are "sharper" than VAEs



https://skymind.ai/wiki/generative-adversarial-network-gan

Deep Learning Frameworks

The popularity and success of deep learning is also due to the availability of open source software implementation of popular neural networks and algorithms.

- **Tensorflow** is Google's framework and one of the most popular, it's widely adopted in the industry as well as research. It's based on static computational graphs
    - **Keras** is an high level interface for tensorflow and allows for fast prototyping
- **Pytorch** is a framework developed by Facebook and it has become increasingly popular with researchers for its ease of use and speed. Different from tensorflow, it uses dynamic computational graphs that make debugging easier

There are other niche frameworks like Caffe, MXnet and gluon that all work basically in the same way.

Deep Learning Applications

In the next lesson we will see some application to the following fields:

- Image classification, segmentation, captioning
    - application to medicine, self-driving cars, ecc.
- Audio classification and analysis
    - recommender systems for music
- Text Generation,Text-to-Speech, Automatic Translation, Chatbots
    - personal assistants like alexa
- Image and Music Generation, Arts
- **Your Suggestions?**

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Thank You!

Luca Casini
PhD in Data Science and Computation

luca.casini7@unibo.it