Deep Learning

Giuseppe Serra

Associate Professor - University of Udine

http://ailab.uniud.it/

About me

- 2006 - Laurea Degree of Computer Engineering University of Florence
- 2006 - Visiting Scholar - Carnegie Mellon University (CMU), Pittsburgh (U.S.A.)
- 2010 - Ph.D. in Computer Engineering, Multimedia and Telec. (University of Florence)
- 2011 - Visiting Scholar - TELECOM ParisTech, Paris, France
- 2011-2014: Postdoctoral Fellow (University of Florence and Modena and Reggio Emilia)
- 2014 - 2016 - Assistant Professor of Computer Engineering (University of Modena and Reggio Emilia)
- 2017 - Assistant Professor of Computer Science with tenure track (University of Udine)
- 2019 – Associate Professor of Computer Science (University of Udine)

Artificial Intelligence LAB (AILAB-UDINE)

See our website:

http://ailab.uniud.it/

Carlo Tasso

Full Professor

Artificial Intelligence - Software Engineering

Giuseppe Serra

Associate Professor

Machine Learning - Deep Learning - Multimedia

Saida Mahmoud

PhD Student

Machine Learning - Deep Learning - Fuzzy Logic

Alex Falcon

PhD Student

Deep Learning - Predictive Maintenance

Beatrice Portelli

PhD Student

Machine Learning - Fake News Detection - Adverse Durg Event Extraction

Simone Scaboro

Graduate Research Fellow

NLP - Adverse Drug Event Extraction

AILAB on Linkedin

Office Hours and Materials

Office Hours:

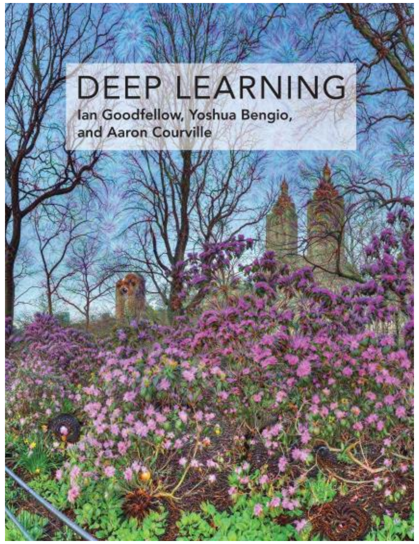
- Tuesday 11:00-12:00 (please send me an email to book a slot)

Contents:

- Videos and slides are on Teams

References:

- CS230 Deep Learning (Stanford University)
- UVA Deep Learning (University of Amsterdam)
- Deep Learning Book, by I. Goodfellow, Y. Bengio and A. Courville
- Other links I will provide you

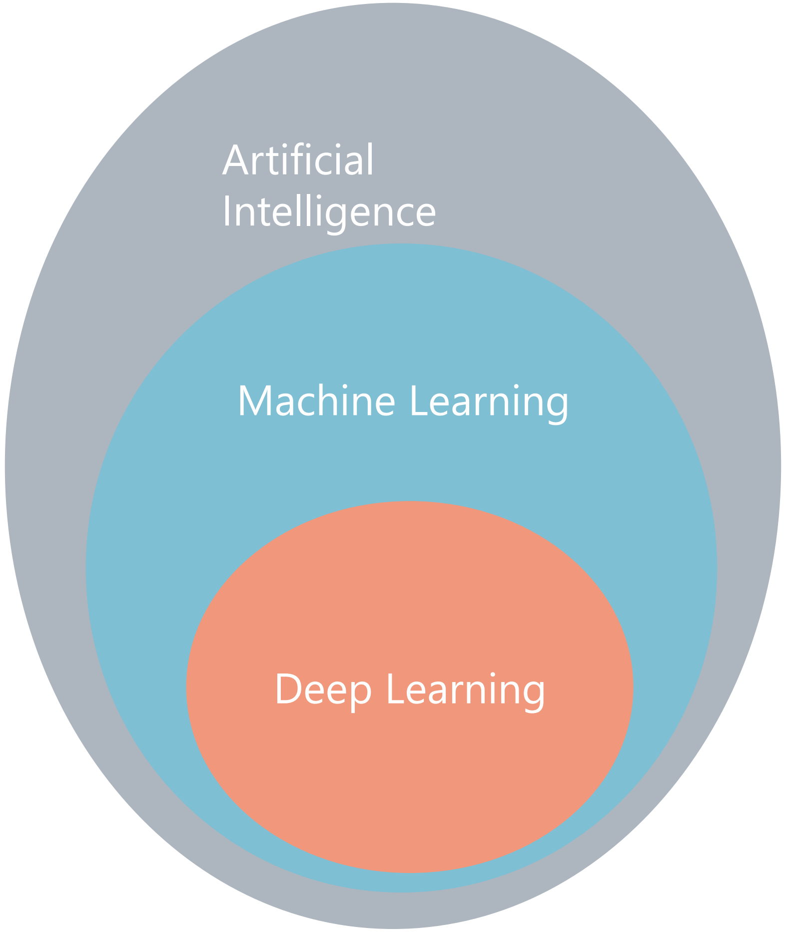


Terminology

Artificial Intelligence
Any technique which enables computers to mimic human behavior.

Machine Learning
Subset of AI techniques which use statistical methods to enable machines to improve with experiences without explicitly being programmed

Deep Learning
Extract patterns from data (or raw data) using neural networks



Machine and Deep Learning

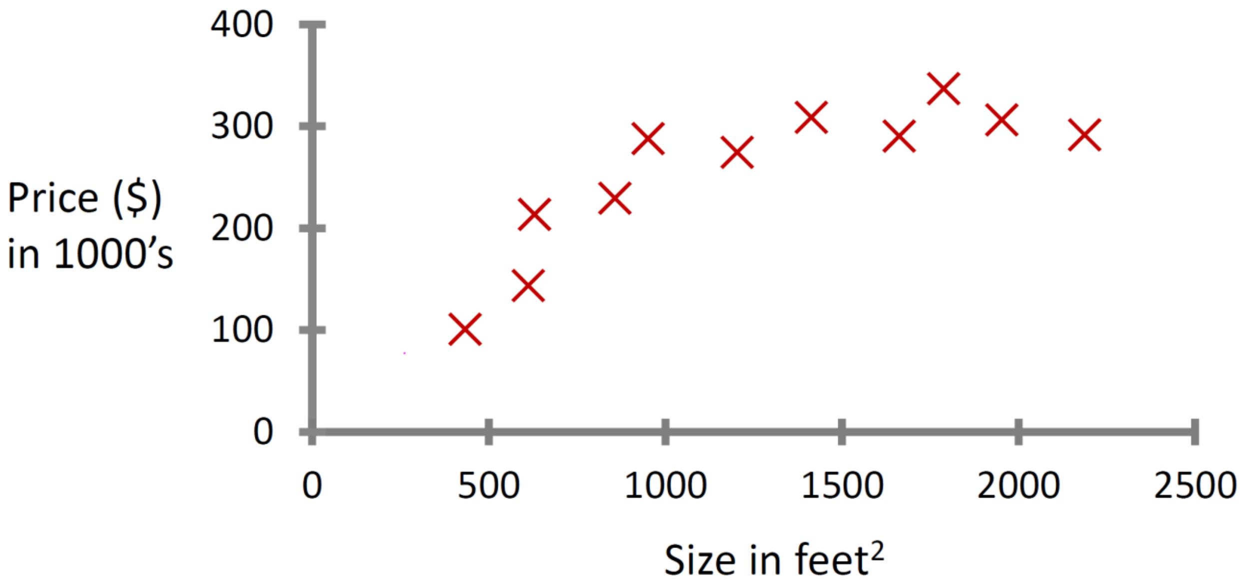
Machine learning algorithms can roughly divided in:

- Supervised learning
- Unsupervised learning
- Reinforcement learning
- Generative Learning

Deep Learning (DL) is cross-sectorial technique based on Neural Network.

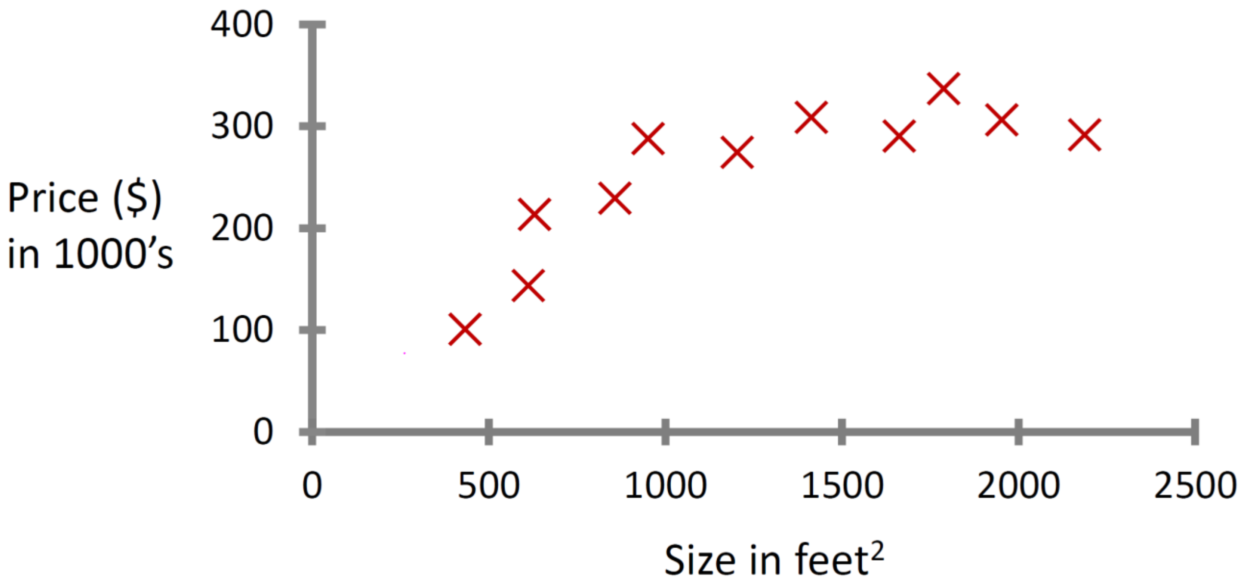
Supervised Learning - Regression

Housing price prediction



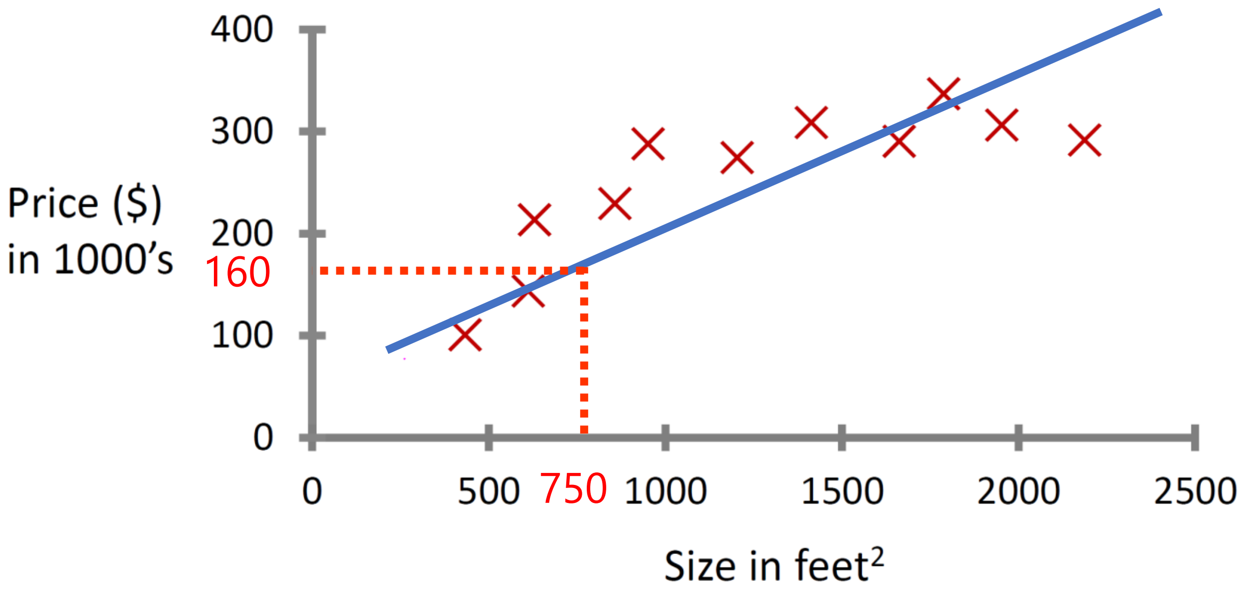
Housing price prediction

Suppose you have a friend with a house of 750 $feet^2$. How can a learning algorithm help you?



Housing price prediction

Suppose you have a friend with a house of 750 feet^2. How can a learning algorithm help you?



A learning algorithm can fit a straight line to the data. Based on that, it looks like maybe the house can be sold for about $ 150,000.

Supervised Learning

The **supervised learning** refers to the fact that we use a dataset in which **“the right answer” are given**.

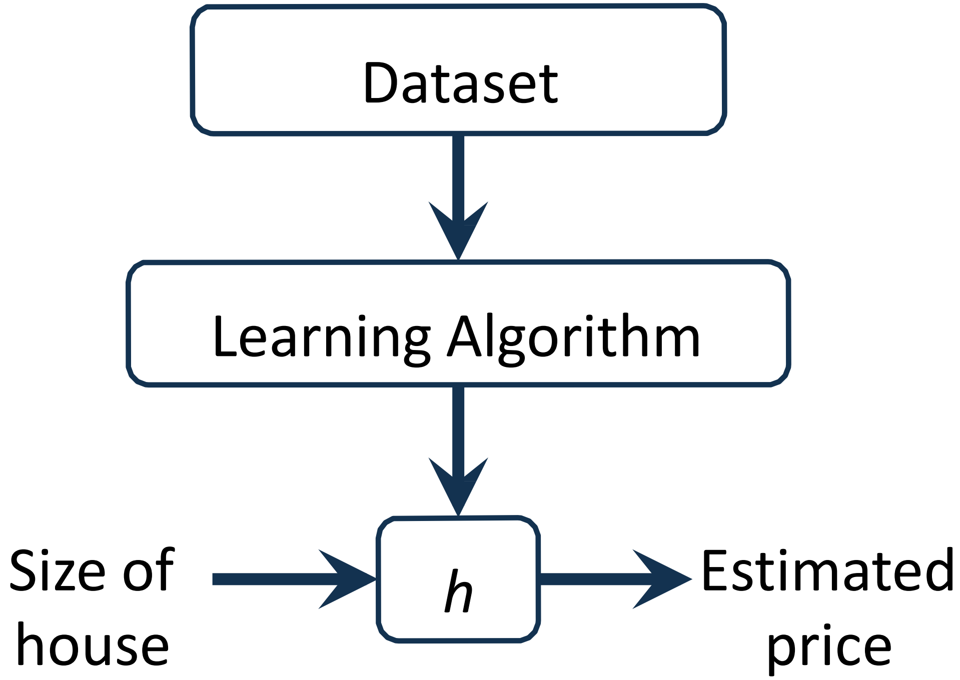
Dataset:

Size in feet^2 (x)	Price ($) in 1000's (y)
2104	460
1416	232
1534	315
852	178
…	…

$\boldsymbol{x}$'s = input variable /features

$\boldsymbol{y}$'s = «output» variable / «target» variable

Supervised Learning - Regression



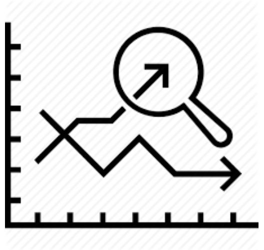
H = hypothesis (historical name). Today we can call it «model»

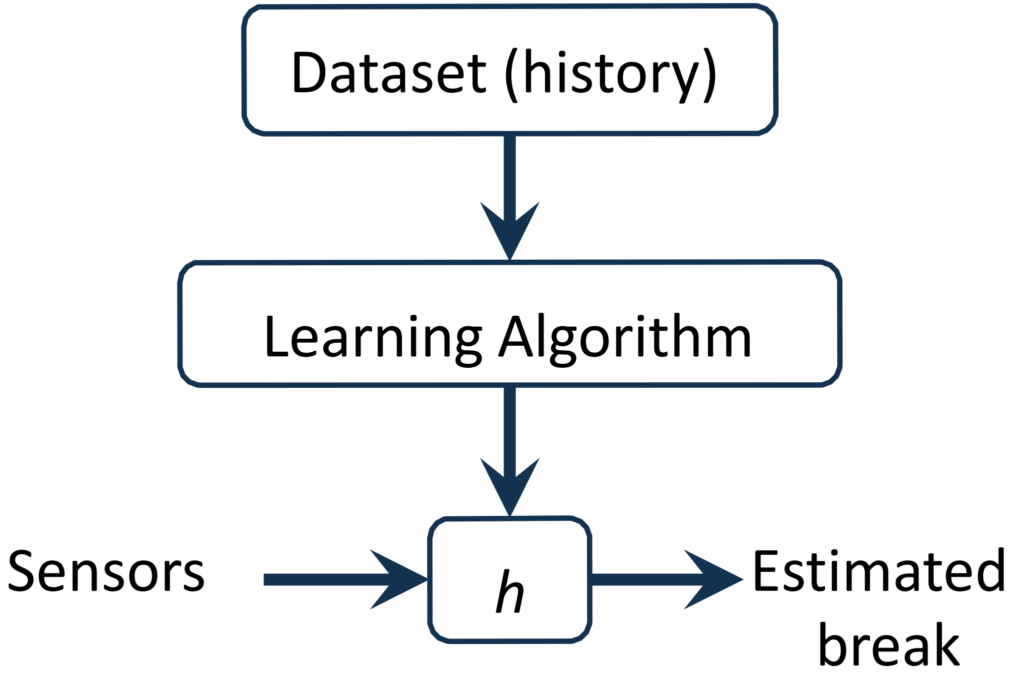
In regression the output of the model is a real and continuos value.

Regression with Multiple Features

Size (feet^2)	Number of bedrooms	Number of floors	Age of home (years)	Price ($1000)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
…	…	…	…	…

Examples – Predictive Maintenance





Supervised Learning - Classification

Classification

The **classification problem** is like the regression problem, except that the values **we want to predict take on only a small number of discrete values**.

Some examples:

Spam/ Not Spam

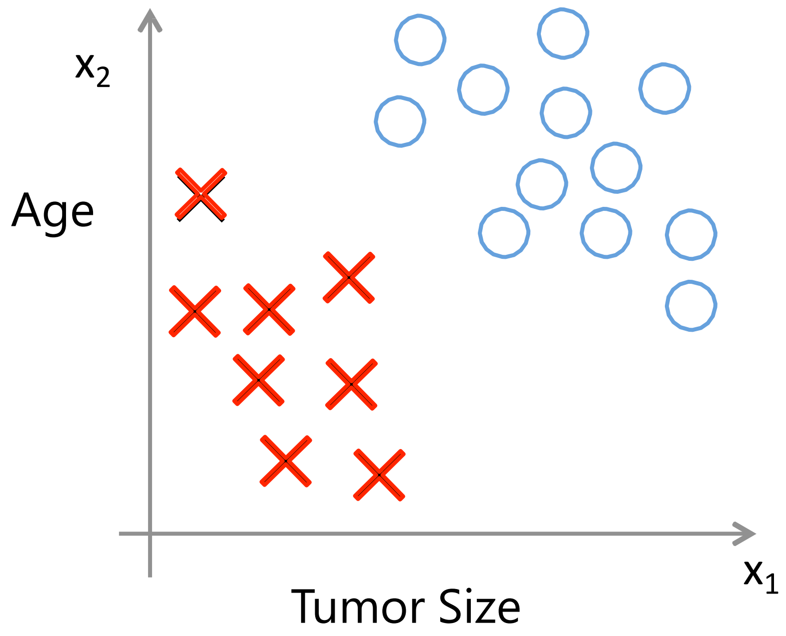
Image with a face / image without a face

Malignant tumor / Benign tumor

Binary classification problem: predictions can take on only two values, 0 (usually called “Negative Class”) or 1 (usually called “Positive Class”).

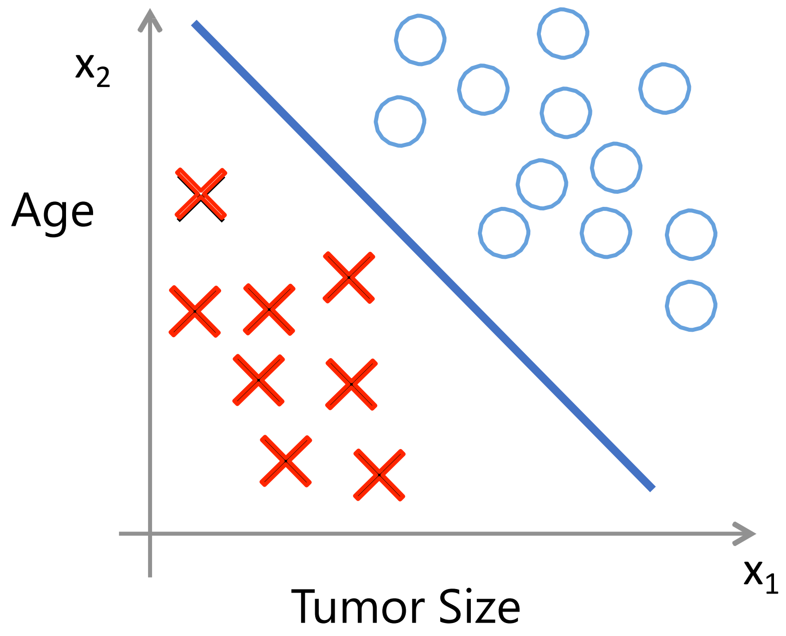
Classification

Suppose we have a tumor classification problem and we have as features both the age of the patients and the tumour sizes. The dataset could look like this:



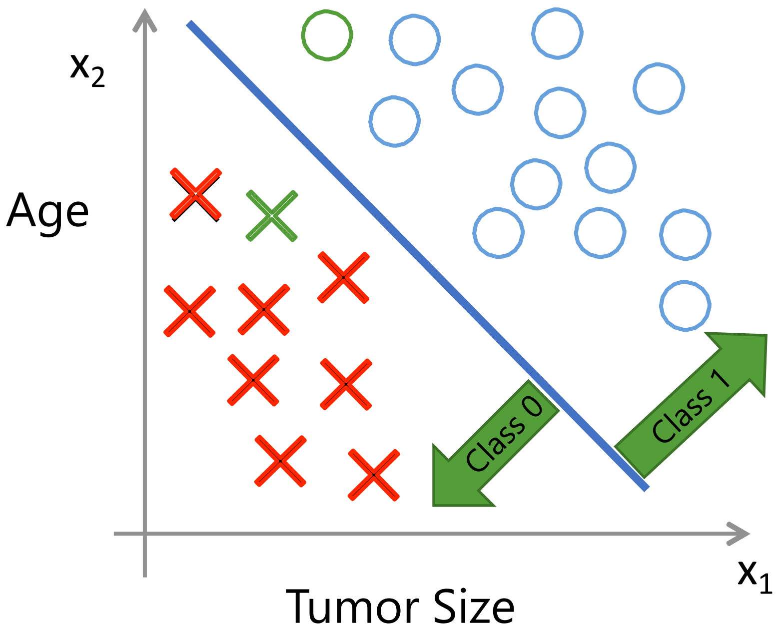
Classification

A Classification algorithm tries to estimate the decision boundary (i.e. the line that separates the data)

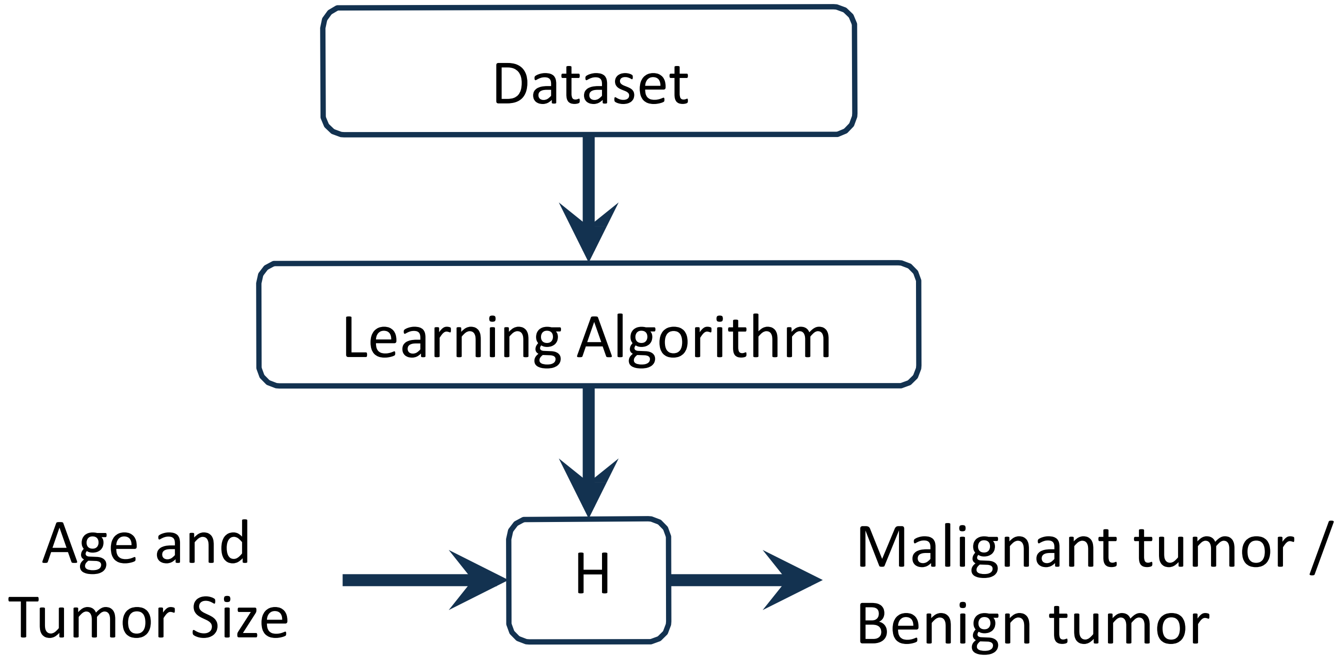


Classification

Once you have a classifier you can use it for prediction of new samples.

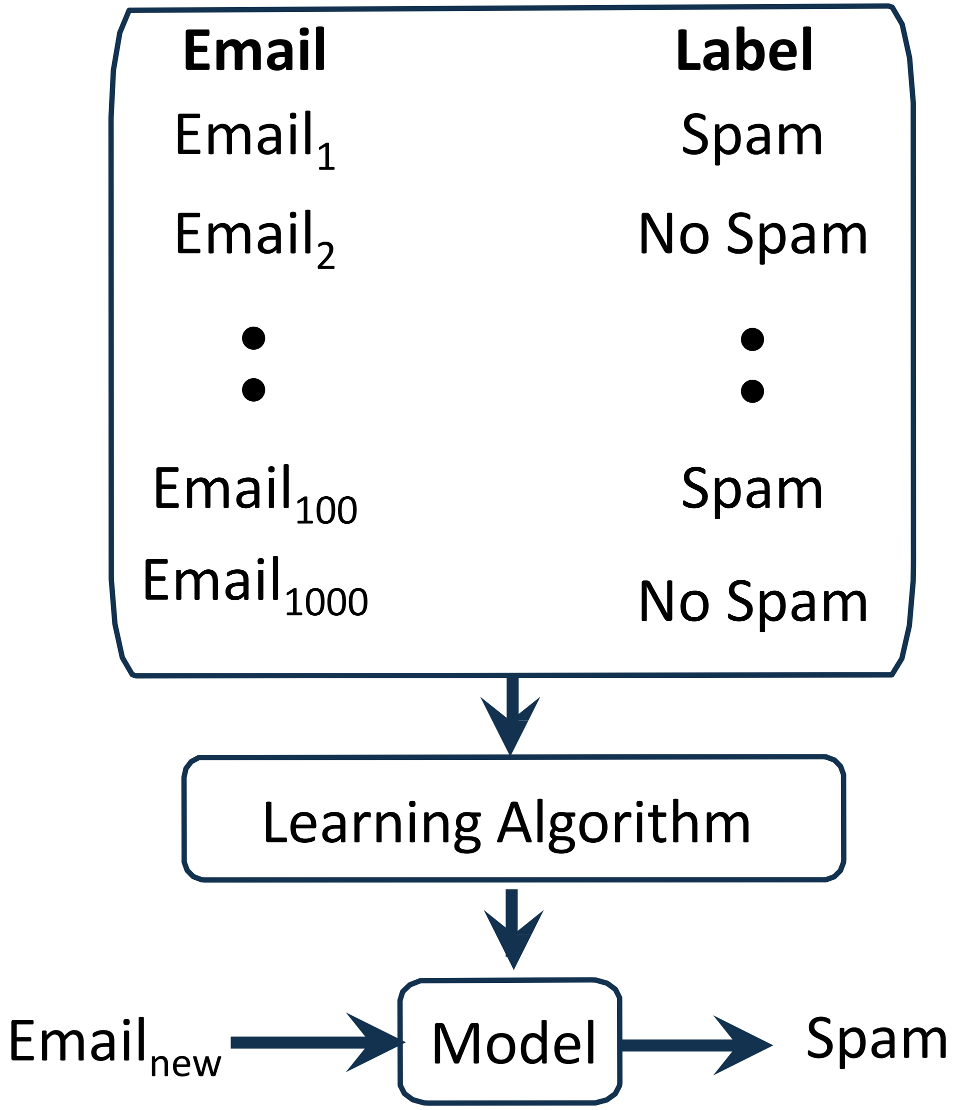


Classification



Examples – Spam Detection

Examples – Spam Detection



Examples – Spam Detection

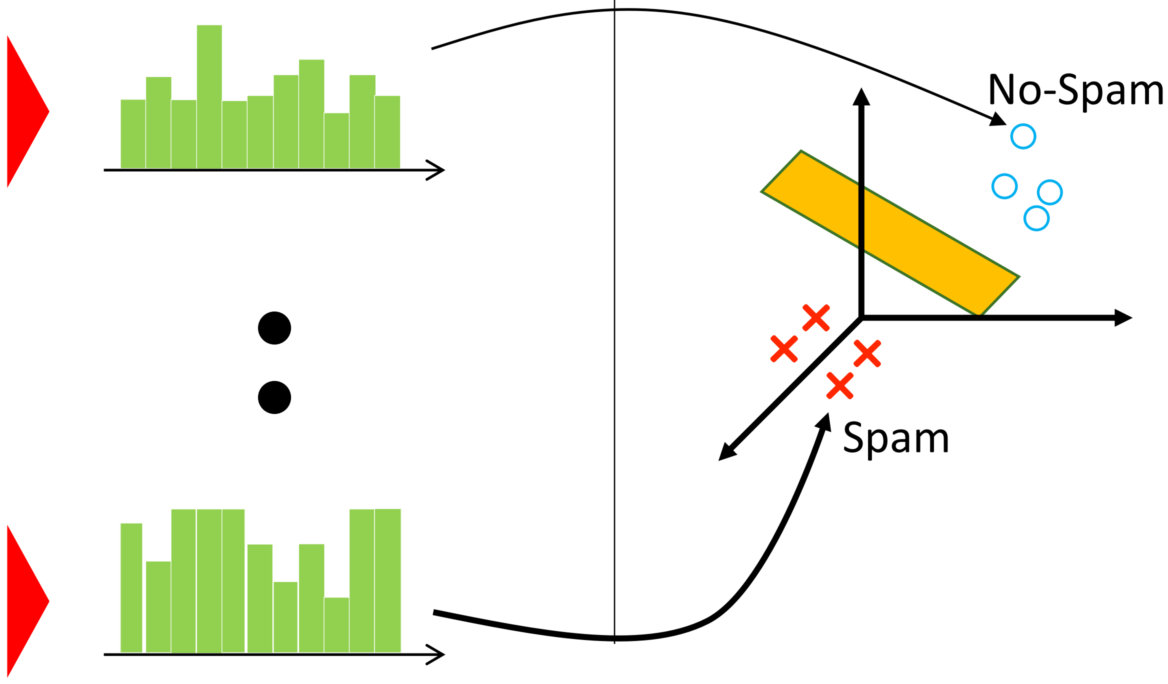
Email

Our photograph was forwarded to us as part of an article we are publishing for ….. your approval or any changes you would like.

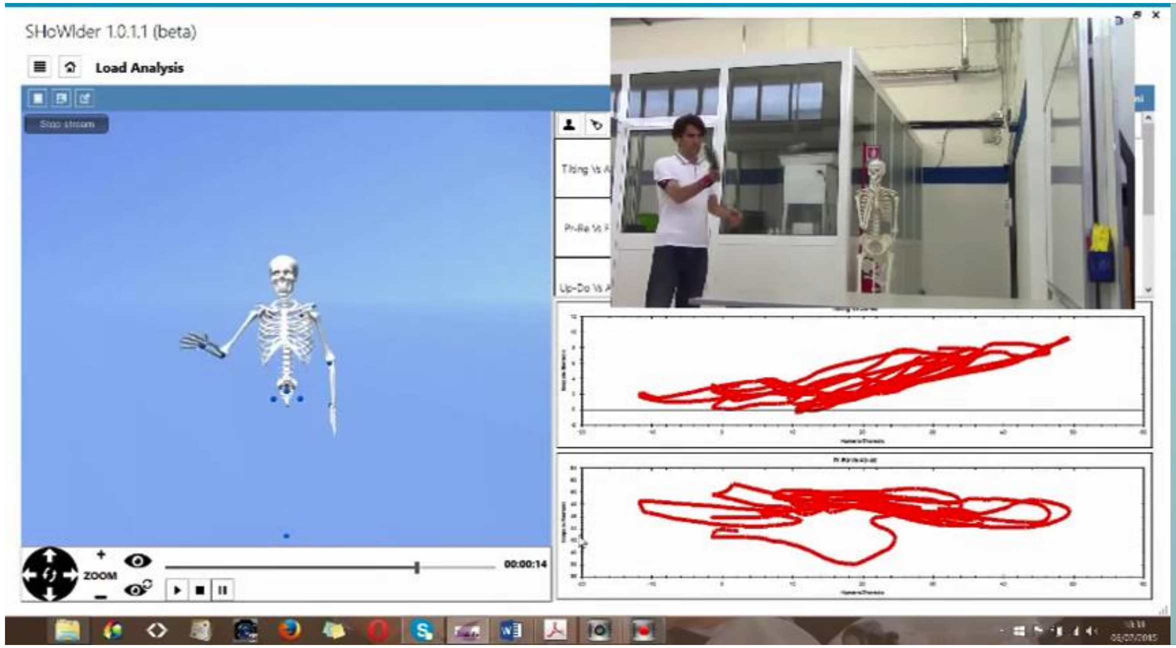
we have logged your IP-address on more than 40 illegal Websites. Important: Please answer our questions!

Numerical Representation

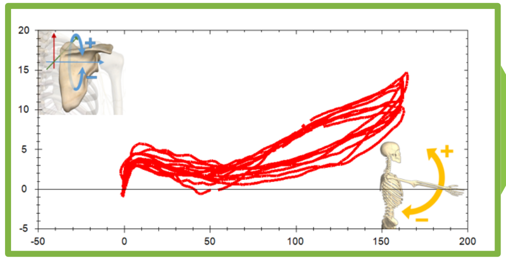
High dimensional Space



Examples - Arm Pathology Analysis



Examples - Arm Pathology Analysis



Artificial Intelligence Algorithm

"Pathology"

Prediction

"No Pathology"

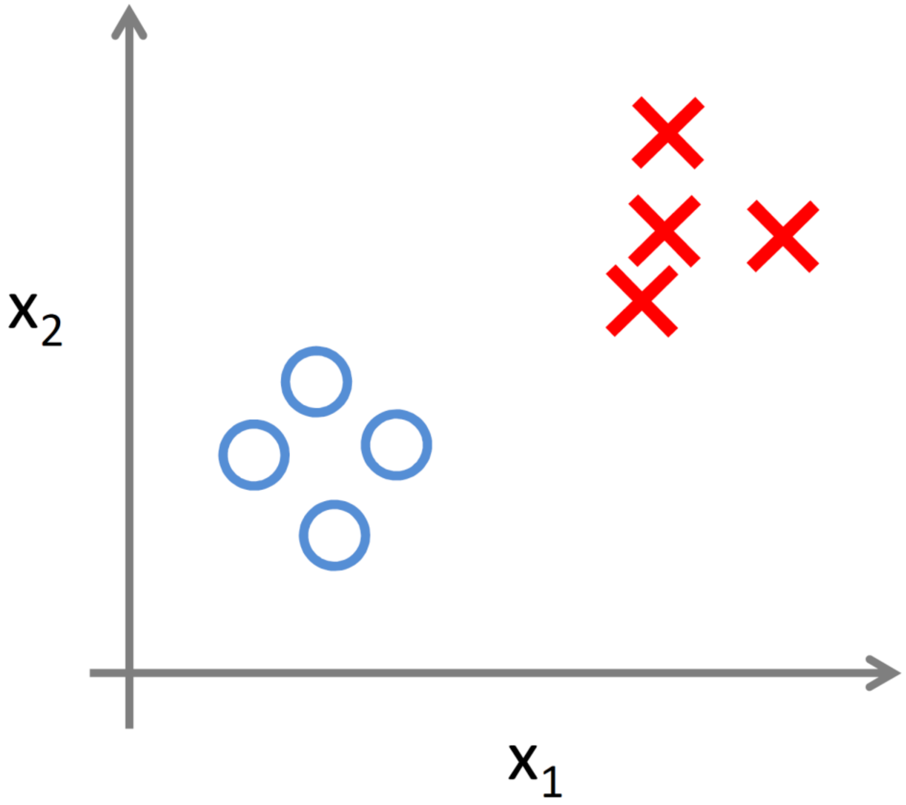
Supervised Learning

Input(x)	Output (y)	Application
Home features	Price	Real Estate
Ad, user info	Click on ad? (0/1)	Online Advertising
Image	Object (1,…,1000)	Photo tagging
Audio	Text transcript	Speech recognition
English	Chinese	Machine translation
Image, Radar info	Position of other cars	Autonomous driving

Unsupervised Learning

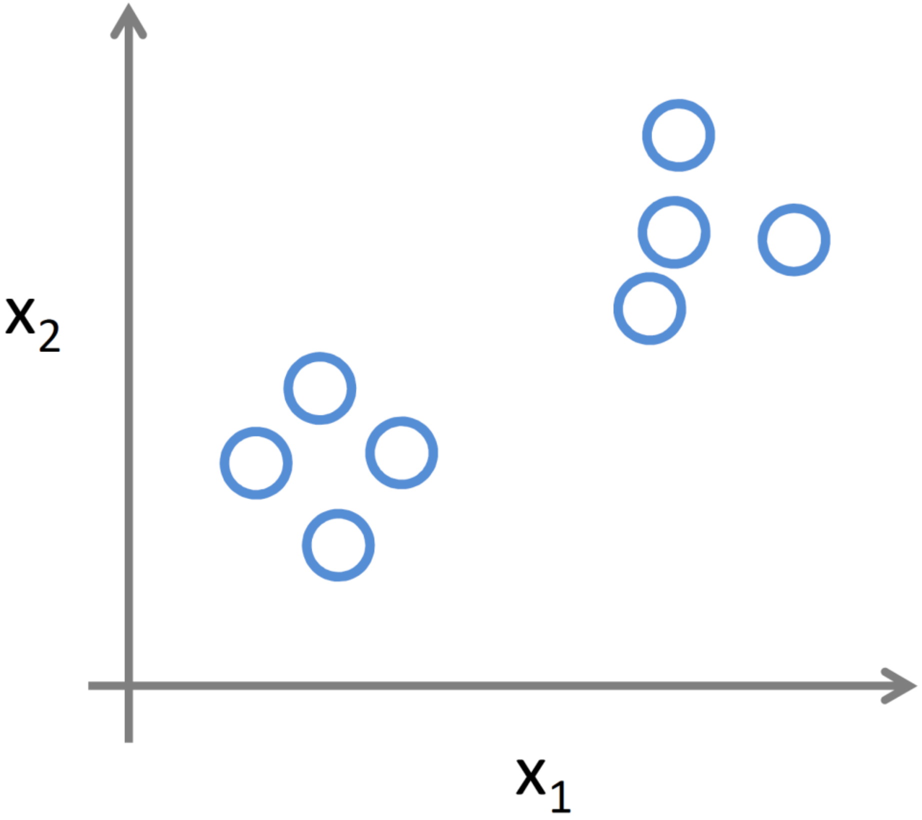
Classification

Recall, in supervised problem each example is labelled either as a positive or as negative



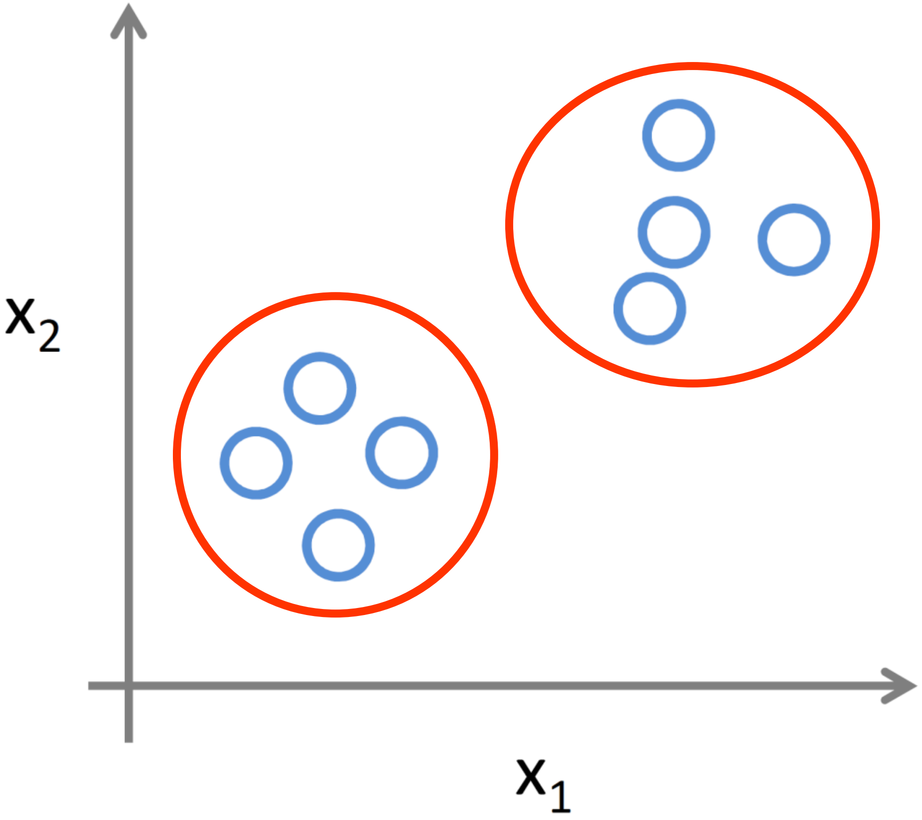
Unsupervised Problem

In unsupervised problem each example is NOT labelled either as a positive or as negative

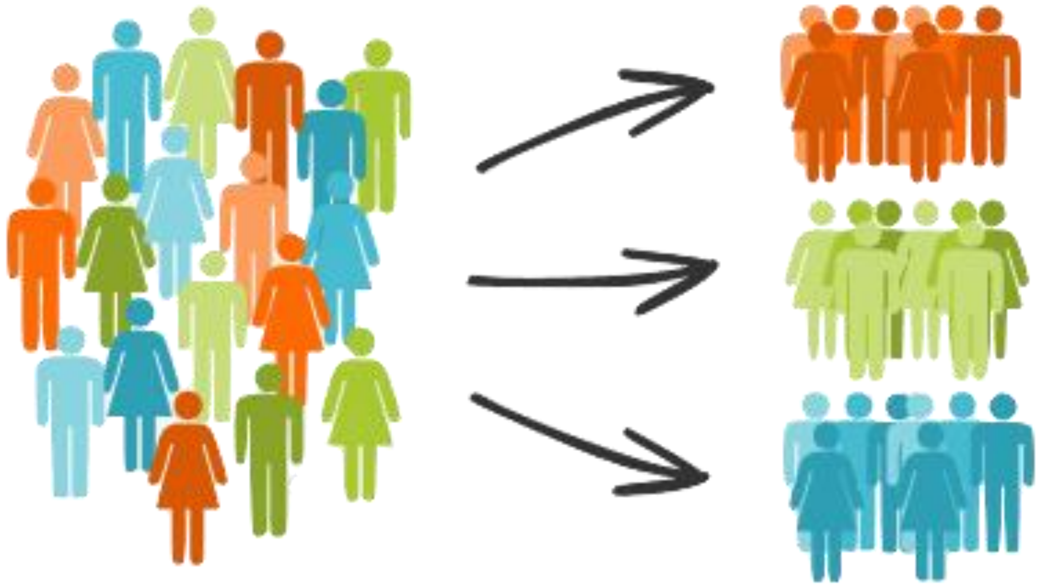


Unsupervised Problem

How can we find a structure of these data?
An unsupervised learning algorithm might decide that these data live in two different clusters (clustering algorithm)

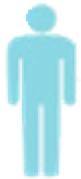


Example – Market Segmentation

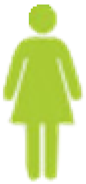


Example – Market Segmentation

Market Segmentation

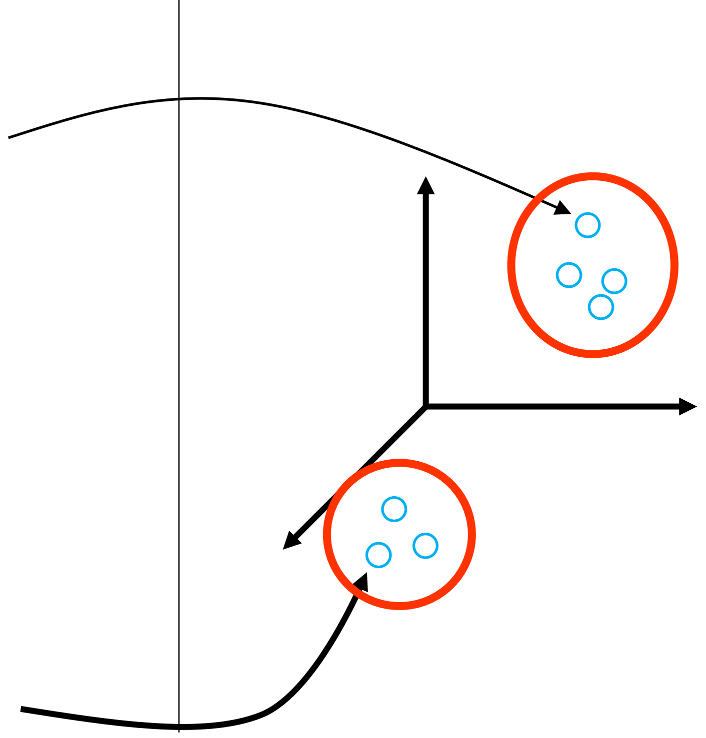


Age	Income	N. of children	...	N. of purchase in the last year
24	36000	1	...	40



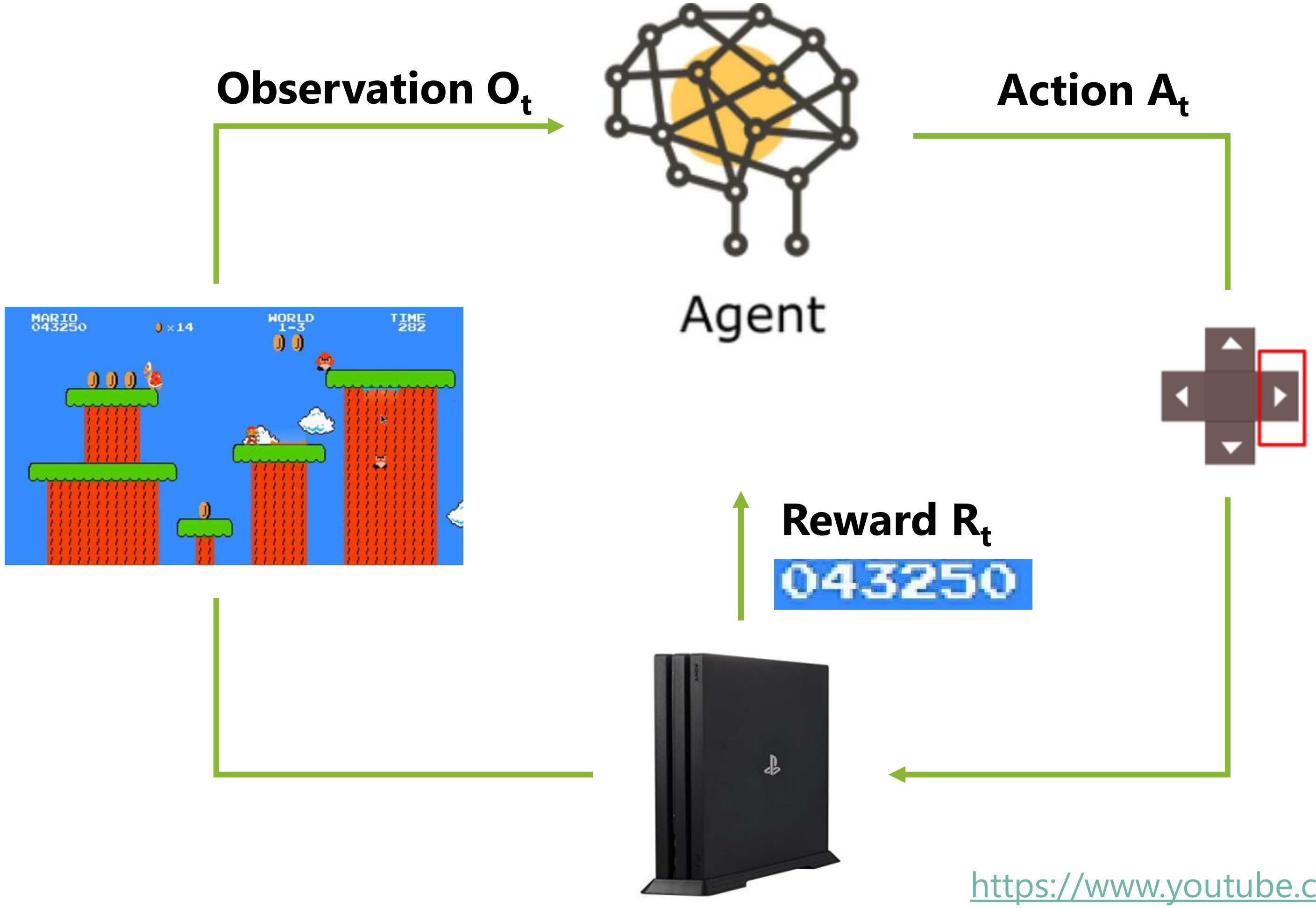
Age	Income	N. of children	...	N. of purchase in the last year
28	23000	2	...	10

High dimensional Space



Reinforcement Learning

Reinforcement Learning



https://www.youtube.com/watch?v=YUYnGA0Gqv4

Reinforcement Learning

Generative Learning

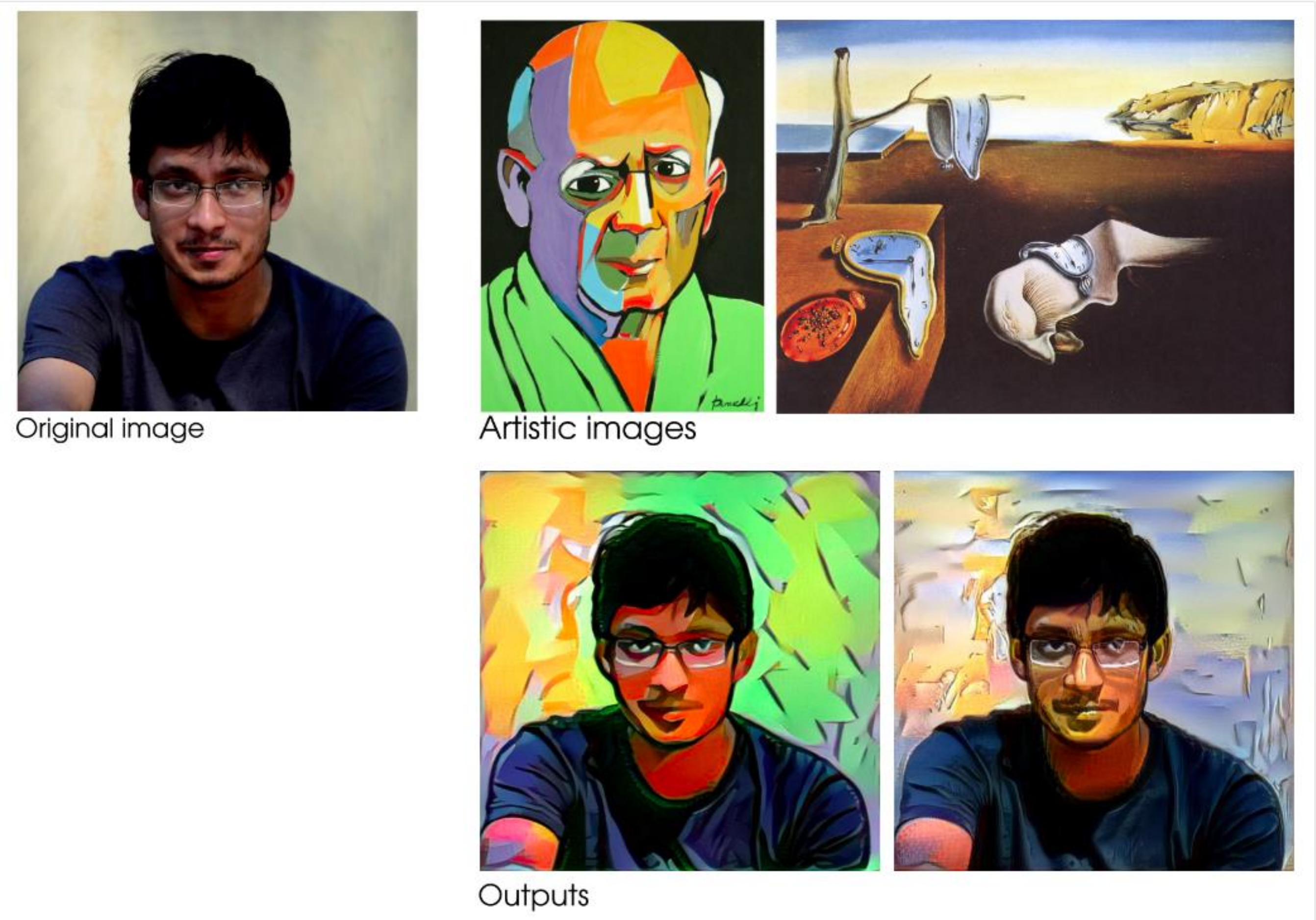
Generative Adversarial Networks

Generative Adversarial Networks (GANs) are an exciting recent innovation in Machine Learning.

GANs are *generative* models: they create new data instances that resemble your training data.

For example, GANs can create images that look like photographs of human faces, even though the faces don't belong to any real person.
These images were created by a GAN:

Example - Vincent Ai Artist



Example – DALLE 2

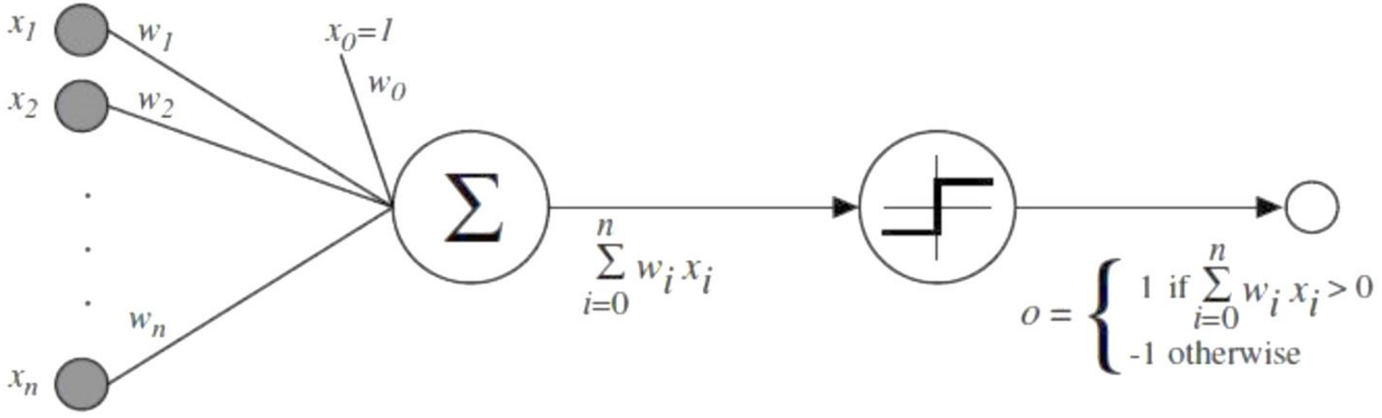
A brief history of Neural Networks & Deep Learning

Perceptron

- The perceptron algorithm was invented in 1958 at the Cornell Aeronautical Laboratory by Frank Rosenblatt
- The perceptron became the first model for binary classification.

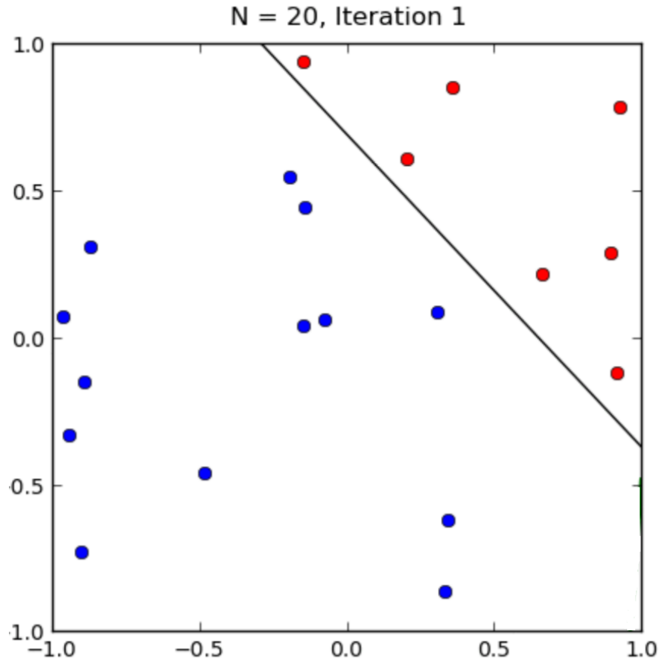
  - One weight w_i per input x_i;

  - If the result is larger than a threshold it returns 1, otherwise 0 (Non-Linearity)



Training a Perceptron

- Rosenblatt's innovation was mainly the learning algorithm for a perceptron
- Learning algorithm:
  - Initialize weights randomly
  - Take one sample x_i and predict y_i
  - For erroneous predictions update weights
  - If prediction $\hat{y} = 0$ and ground truth $y_i = 1$, increase weights
  - If prediction $\hat{y} = 1$ and ground truth $y_i = 0$, decrease weights
  - Repeat until no errors are made



Demo: https://lecture-demo.ira.uka.de/neural-network-demo/?preset=Rosenblatt%20Perceptron

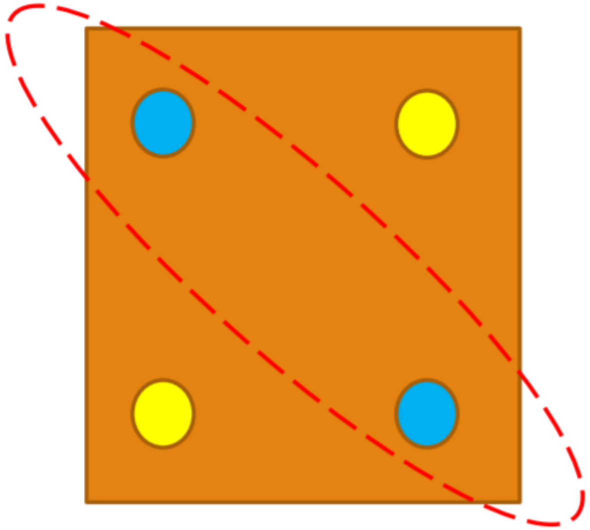
XOR & Single-layer Perceptron

- However, the exclusive or (XOR) cannot be solved by a perceptron [Minsky and Papert, “Perceptrons”, 1969]

A	B	Q
0	0	1
0	1	0
1	0	0
1	1	1

XNOR

The classification boundary to solve XOR is not a line!!



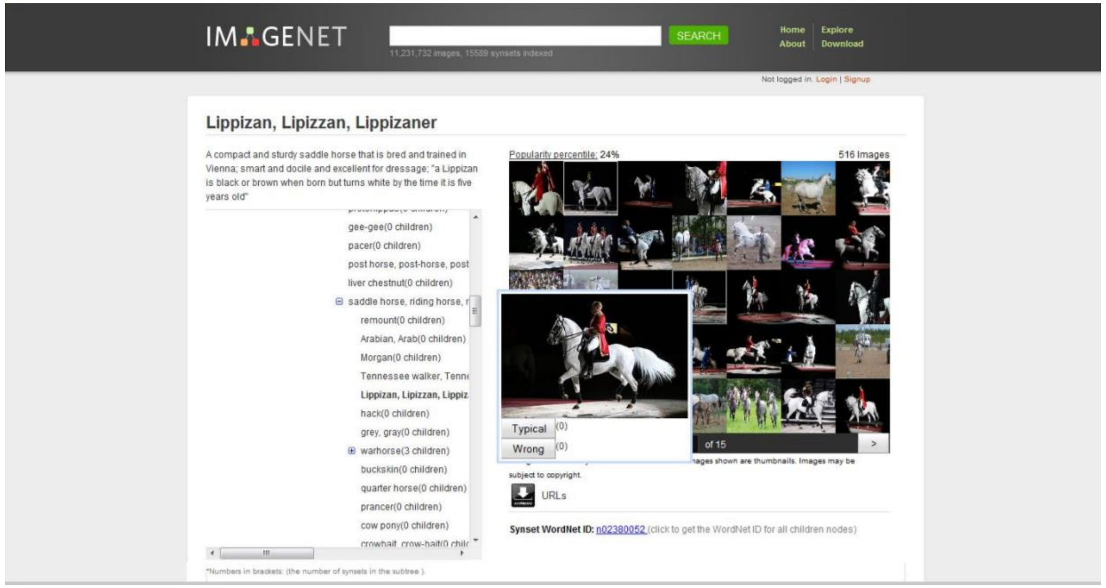
- Multi-layer perceptrons can solve XOR
- Problem: how to train a multi-layer perceptron? Rosenblatt’s algorithm not applicable (it expects to know the ground truth for each perceptron)

“AI winter”

- What everybody thought: “If a perceptron cannot even solve XOR, why bother?
- Results not as promised (too much hype!) -> no further funding -> AI Winter
- Still, **significant discoveries were made in this period**:
  - **Backpropagation** (Learning algorithm for MLPs) (see next classes)
  - **Recurrent Neural Networks** (see next classes)
- **Neural Network problems** a decade ago
  - **Lack of processing power**
  - **Lack of data**

ImageNet Challenge

- In 2009 the ImageNet dataset was published [Deng et al., 2009]
  - Collected images for each of the 100K terms in Wordnet (16M images in total)
  - Terms organized hierarchically: “Vehicle”->“Ambulance”
- ImageNet Large Scale Visual Recognition Challenge (ILSVRC)
  - 1 million images
  - 1,000 classes
  - Top-5 and top-1 error measured



Building ImageNet

Step 1:
Collect candidate images via the Internet

Step 2:
Clean up the candidate Images by humans

YAHOO!
picsearch™
flickr
Live Search
Google™ Image Search

amazonmechanical turk
beta
Artificial Artificial Intelligence

Deep learning at ImageNet classification challenge

CNN based, non-CNN based

2012 Teams	%error
Supervision (Toronto)	15.3
ISI (Tokyo)	26.1
VGG (Oxford)	26.9
XRCE/INRIA	27.0
UvA (Amsterdam)	29.6
INRIA/LEAR	33.4

2013 Teams	%error
Clarifai (NYU spinoff)	11.7
NUS (singapore)	12.9
Zeiler-Fergus (NYU)	13.5
A. Howard	13.5
OverFeat (NYU)	14.1
UvA (Amsterdam)	14.2
Adobe	15.2
VGG (Oxford)	15.2
VGG (Oxford)	23.0

2014 Teams	%error
GoogLeNet	6.6
VGG (Oxford)	7.3
MSRA	8.0
A. Howard	8.1
DeeperVision	9.5
NUS-BST	9.7
TTIC-ECP	10.2
XYZ	11.2
UvA	12.1

ImageNet 2012 winner: AlexNet

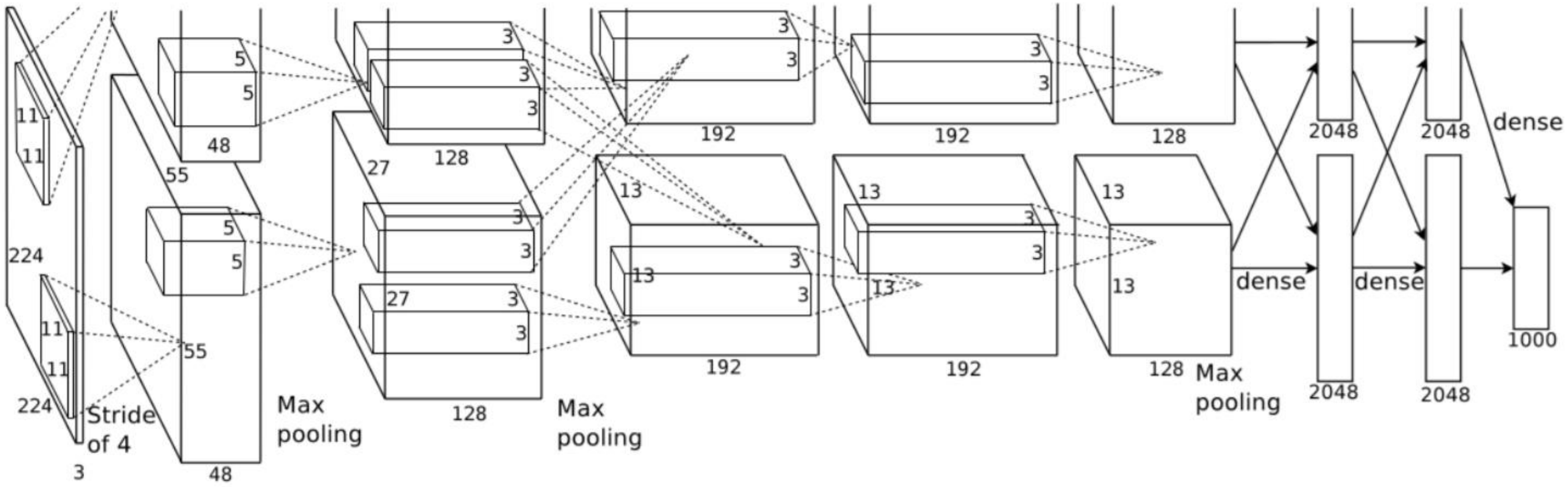


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

Krizhevsky, Sutskever & Hinton, NIPS 2012

The Game of Go

In 2016 AlphaGo (Deep Mind) won Lee Sedol, the world's top Go player.

See more info:

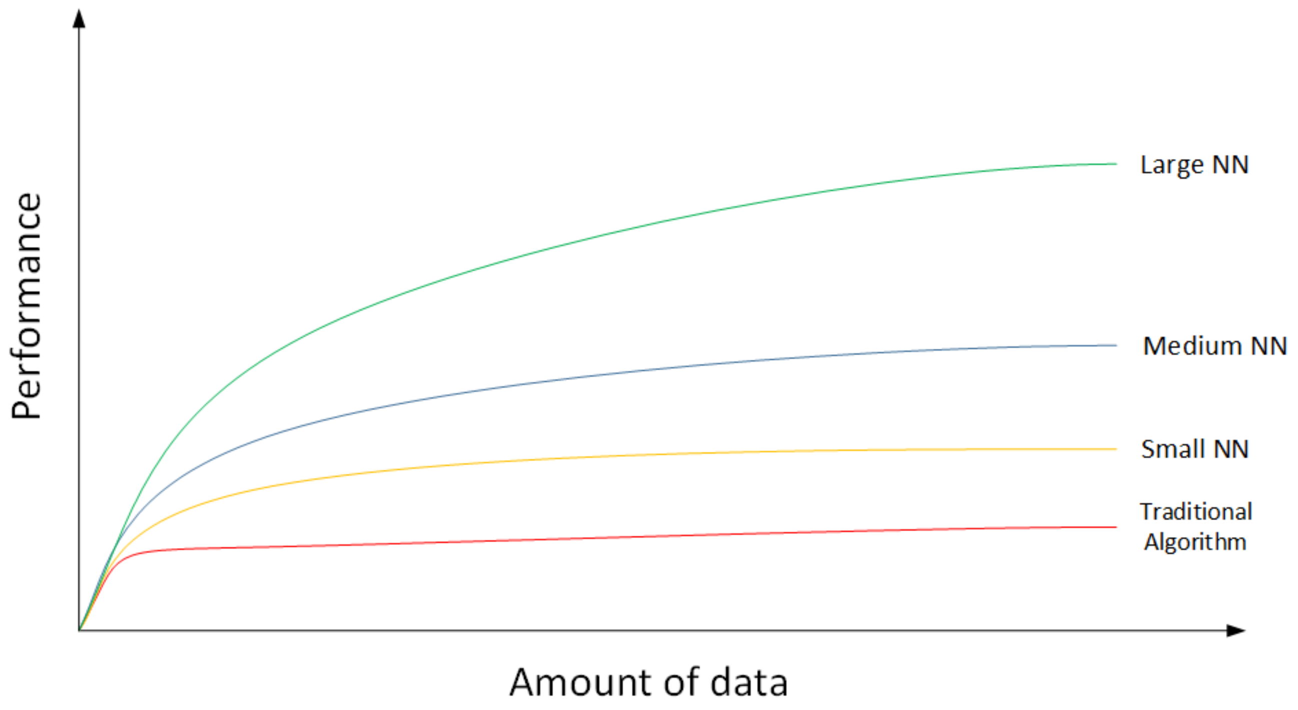
http://airesearch.com/tag/go/

[AlphaGo – The Movie (Full Documentary)]()

Scale drives Deep Learning progress

- When **a small amount of data** when the performance of **traditional learning model** (Logistic Regression, SVM, Decision Tree etc) is **better**.

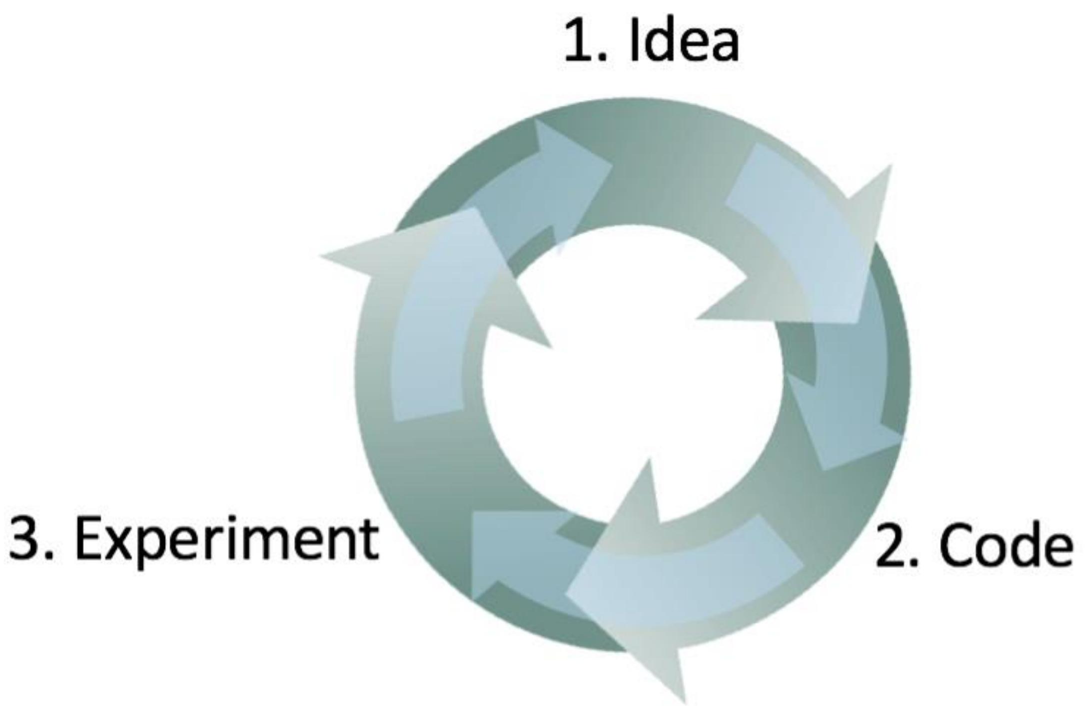
- **Deep learning** is the **first class of learning algorithms that is scalable**: performance just keeps getting **better** as you feed them **more data** (see the NLP Model GPT-3 with 175 billion machine learning parameters).



Scale drives Deep Learning progress

Three key factors:

- Data
- Computation/Hardware
- Algorithms



If faster computing, for each step, the less time-consuming, so the more the circulation can be efficiently performed.