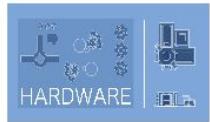


Datascience Bootcamp

Mónica Villas – Professional Career

IBM



1993-2002



1994



2002-2012



2005



2012



2012- Hoy



2016

2019

ODISEIA

Observatorio del impacto social y ético de la inteligencia artificial



Datascience bootcamp

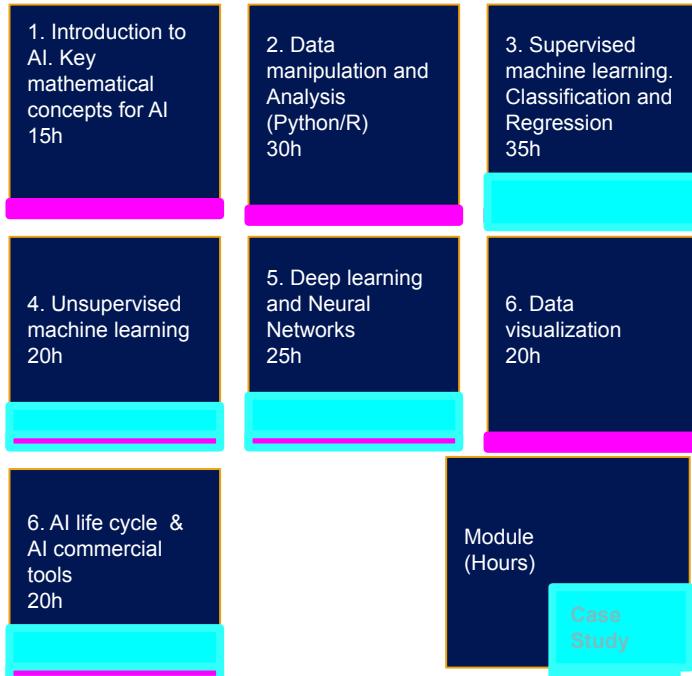
- Python course
- Content
- Calendar
- Teachers
- Additional content
- Capstone
- Tools

Python course

- Challenge for everyone when moving to on-line
- 30 hours + 18 hours home
- Additional exercises to be sent by Benjamin
- Additional resources to be sent in a document by Mónica
- Additional installation to be done as an exercise

Academic Structure

The program is hands-on, with a very practical focus learning Artificial Intelligence with Python (90%) and R (10%)



Theory

Data Science Modules

+

Practice

Case Studies

+

Human Sciences

Soft Skills Toolkit

+

Distributed in 7 modules, you will learn starting from AI foundations, to how to convert data into information and how to present it to finally analyze the future challenges of AI and Machine Learning.

Real world problems that will challenge you to apply all the theory in practice.

The IMMUNE methodology goes beyond technology and includes a Human Sciences perspective with aspects such as Data Visualization and Presentation Skills.

Calendar (April 4th, added)

Calendar

Data Science Developer Bootcamp + Python

Intake March 2020 | Part - Time | 39 sessions

March							April							May							June							July								
Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat		
1	2	3	4	5	6	7		1	2	3	4			1	2			1	2	3	4	5	6		1	2	3	4	5	6	7					
8	9	10	11	12	13	14	5	6	7	8	9	10	11	3	4	5	6	7	8	9	7	8	9	10	11	12	13	5	6	7	8	9	10	11		
15	16	17	18	19	20	21	12	13	14	15	16	17	18	10	11	12	13	14	15	16	14	15	16	17	18	19	20	12	13	14	15	16	17	18		
22	23	24	25	26	27	28	19	20	21	22	23	24	25	17	18	19	20	21	22	23	21	22	23	24	25	26	27	19	20	21	22	23	24	25		
29	30	31			26	27	28	29	30			24	25	26	27	28	29	30		28	29	30					26	27	28	29	30	31				

Classes
Holidays

<https://drive.google.com/open?id=1faGAVnJShgvZ5QqmJtmXckUzxYuQq6nR>

Human science topics

The brain (accommadating in 1.30)

Design thinking (need to be F2F)

Industry talk

Acommodating the new dates to try to do it f2f

INDUSTRY TALKS

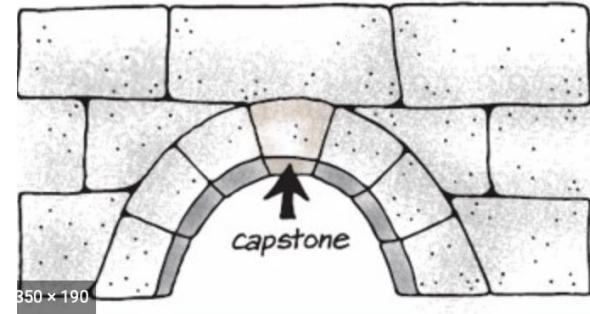
Retail | Rocio Rodriguez - Analiticae

Marketing | Ana Laguna - Zoundream

Utilities | Alicia. Mateo - Endesa

Insurance | Elena Gonzalez - Coverwallet

Capstone



- To be done in group , 2 people , 3 as maximum
- The objective is to practice what you have learned
- A proposal can be done by the team
- One mentor by team
- Deadline to send the proposal or to request help before April 17th -
Note to be sent to Mónica

Tools

- Collab
- Hangout for the calls
- Google classroom for documentation
- Some additional virtual machines if they are needed
- In progress : Virtual blackboard and way to work by peers
- Share documentation to install Anaconda

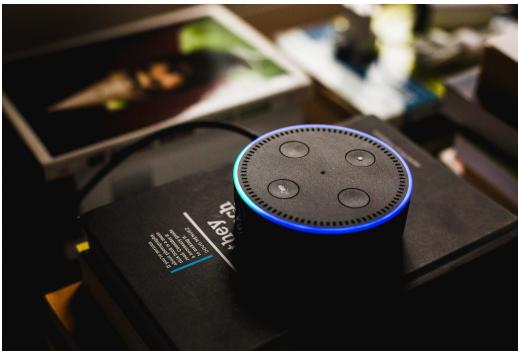
Artificial Intelligence Introduction

1. What is AI ?
2. AI evolution
3. AI classification
4. Why Python
5. Case Studies
6. The role of Datascience

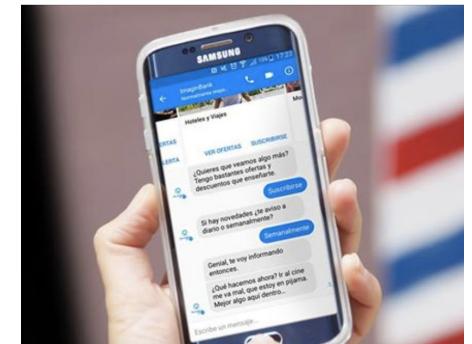
Is this AI?



Are you using AI ?



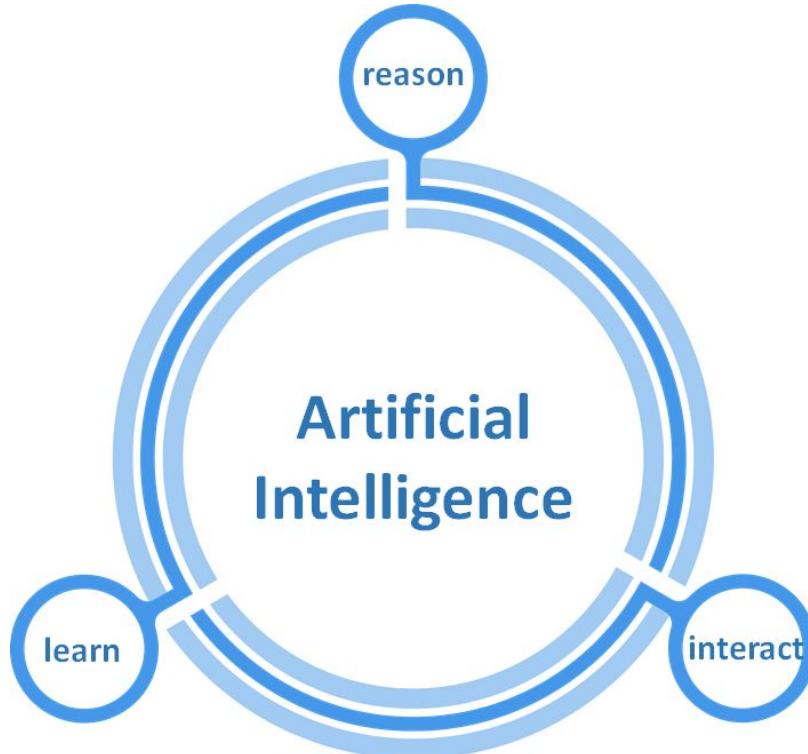
¿Necesitas planes para el fin de semana?
Pregúntale a Gina: nuestro [#chatbot](#) te informará de los descuentos que nada se te escape.
¡Anímate! ♦



Then....what is AI?



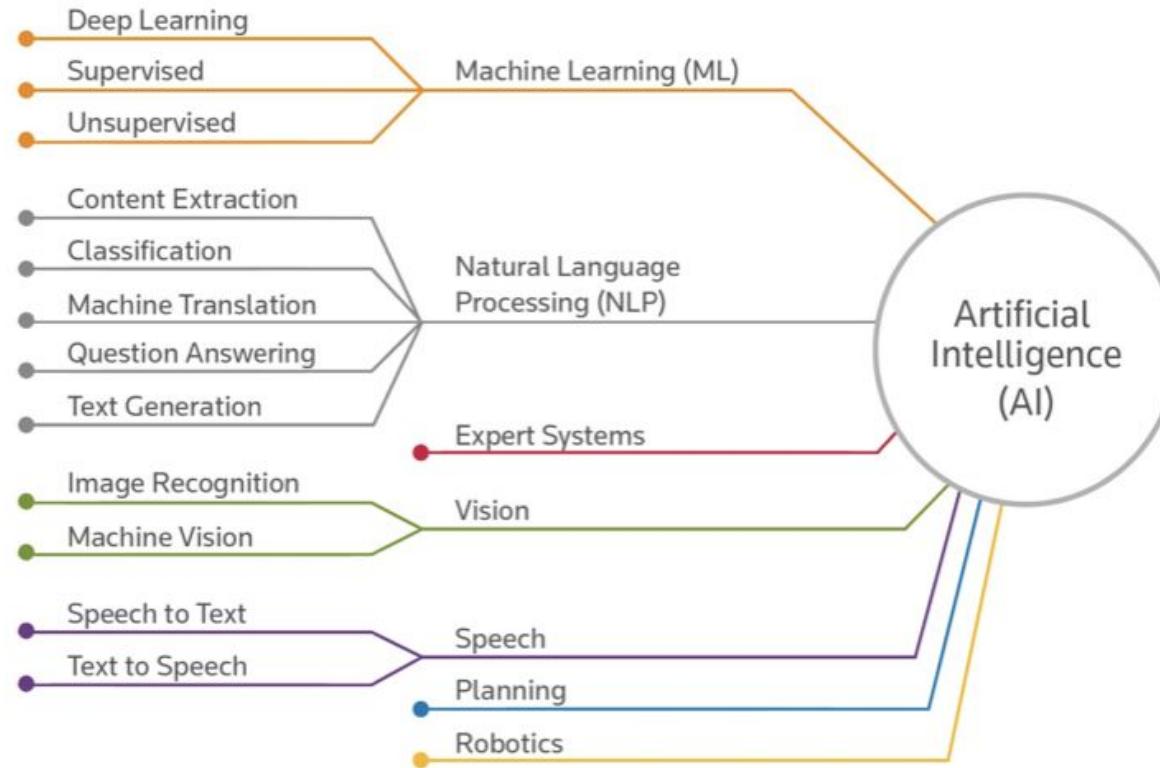
IA needs to



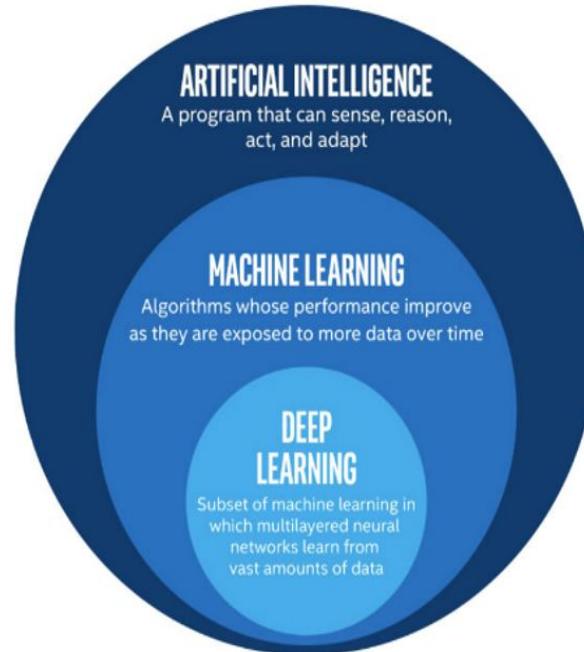
AI is technology that thinks and acts by simulating human behavior.

Narrative Science study, 2015

Artificial Intelligence

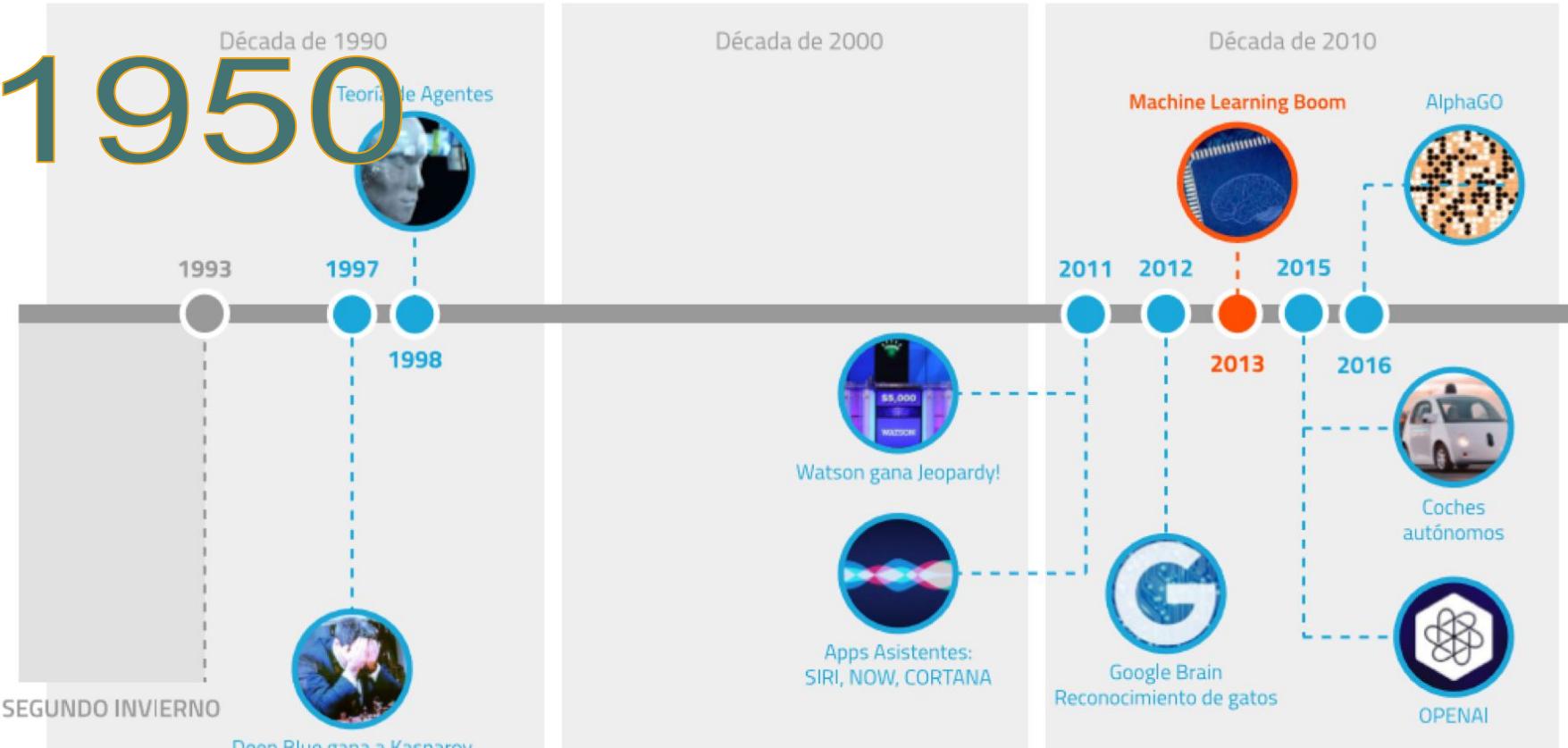


Artificial Intelligence



Fuente: Michael Mills, Artificial Intelligence in Law: The state of Play 2016, Thomson Reuters

When: Artificial intelligence evolution



Artificial intelligence now?



Machine Learning

Seven Steps in Machine Learning

Seven steps in Machine Learning

Step1 : Data (quality and quantity)

Step2: Training data

Step3 : Model to be used

Step4: Training

Step5: Evaluation

Step6: Tuning parameters

Step7: Predict

Differentiate between wine and Beer?



Scalability problem?

Which variable can we measure?

Alcohol content

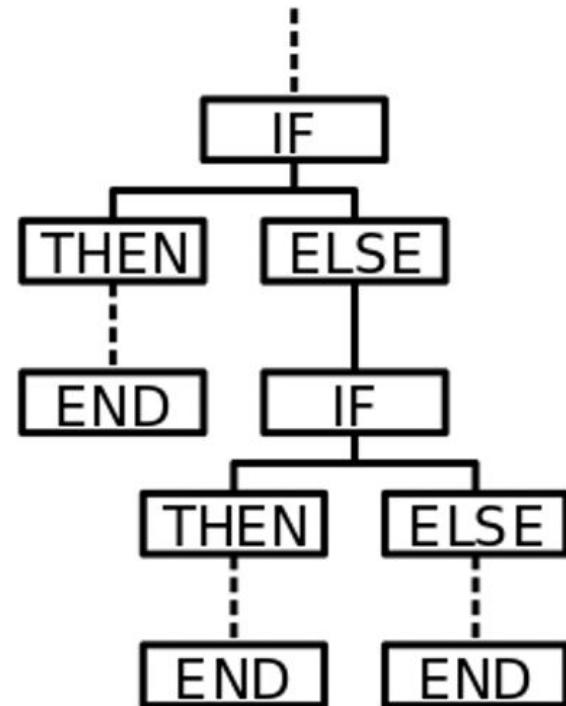
Colour

Acidity

.....

Traditional approach?

A program where I can
track all the variables

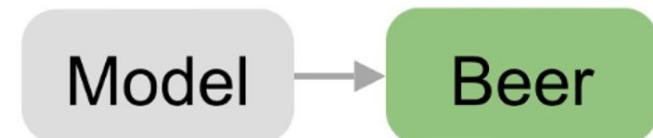


Create a model

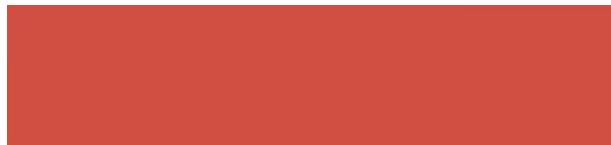
A model which responds to the question is it Beer or Wine ?



Train the model



Data: colour and alcohol content



Colour

13.5% Alc/volume

Alcohol

Obtain data



Step1: Data, quantity and quality

ALCOHOL	COLOUR	WINE?

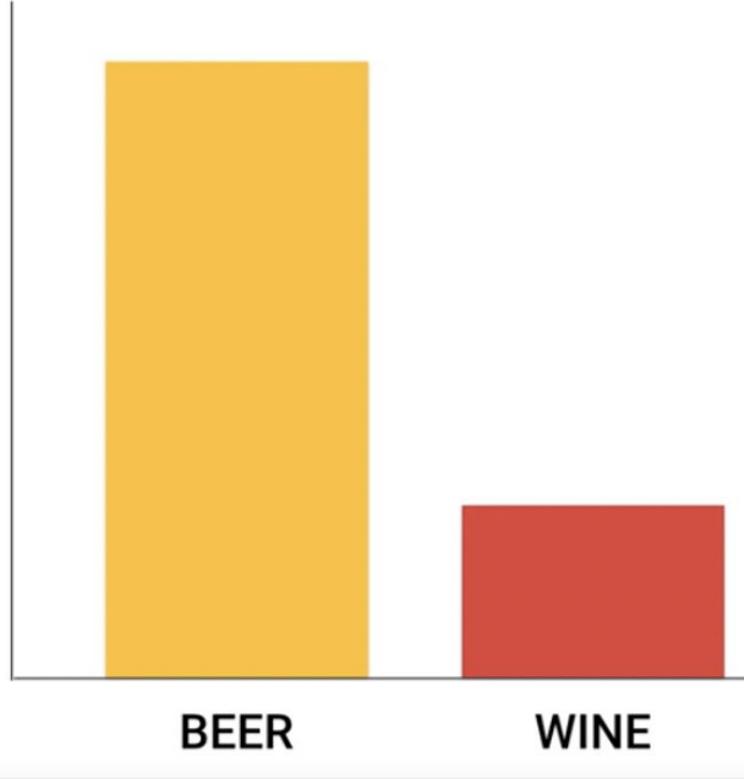
Step2: Training Data

Color (nm)	Alcohol %	Beer or Wine?
610	5	Beer
599	13	Wine
693	14	Wine



Color (nm)	Alcohol %	Beer or Wine?
693	14	Wine
610	5	Beer
599	13	Wine

Step2: Be careful with the data



Check on the data available

Is there a bias on the data?

Do I need different data ?

Normally 80% training 20% test



Training



Evaluation

Step3: Model to be used



It will depend on the data distribution



It will depend on what I want to achieve

It will depend on the type of data

Step3: Choose the model

Machine Learning



Supervised

- Classification
 - Naïve Bayes
 - SVM
 - Random Decision Forests
- Regression
 - Linear
 - Logistic



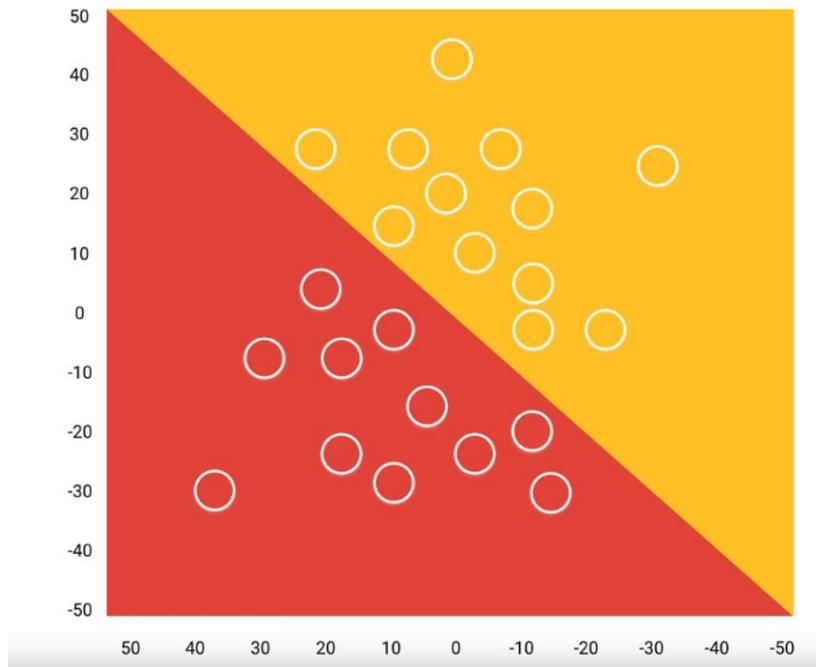
Label

Unsupervised

- Clustering
 - K-means
- Dimensionality reduction
 - Principal Component Analysis
 - SVD



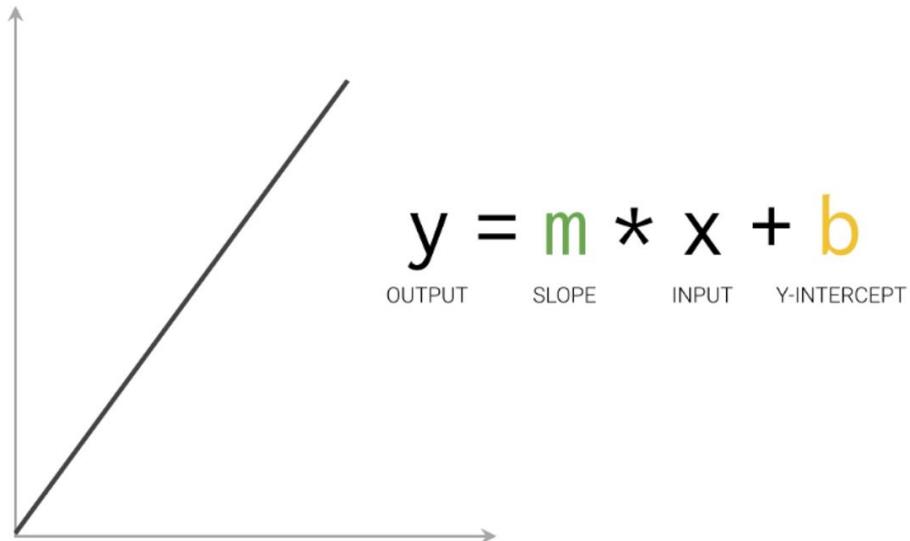
Step3: Model to be choosen



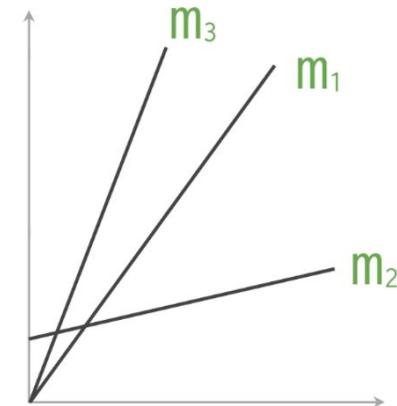
Two variables: colour
and alcohol

Linear model ?

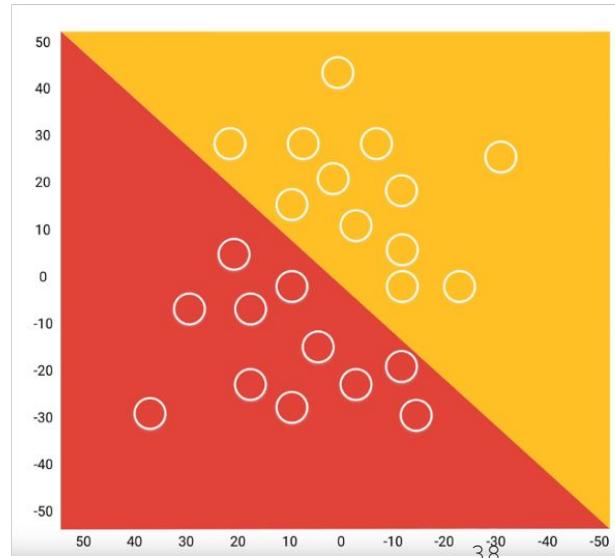
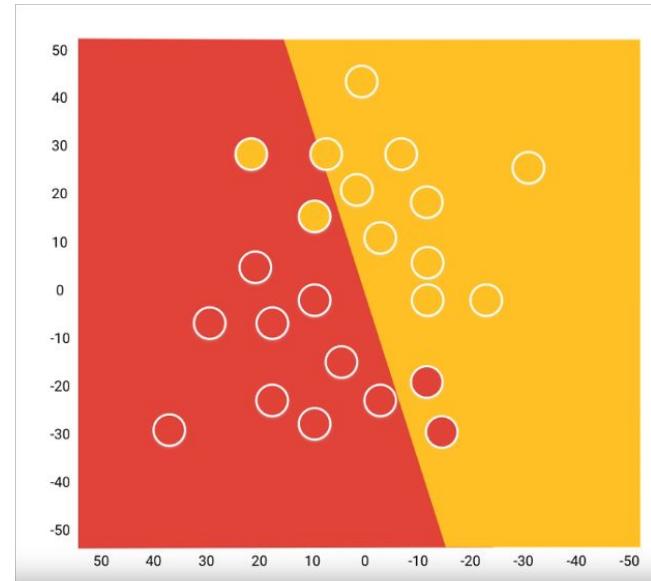
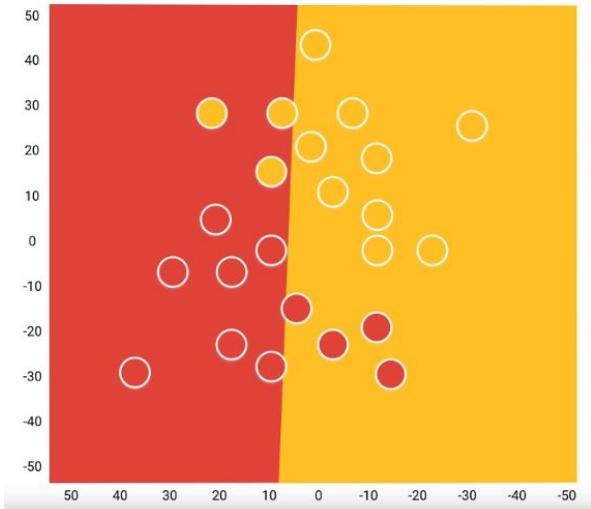
Step3: Linear model



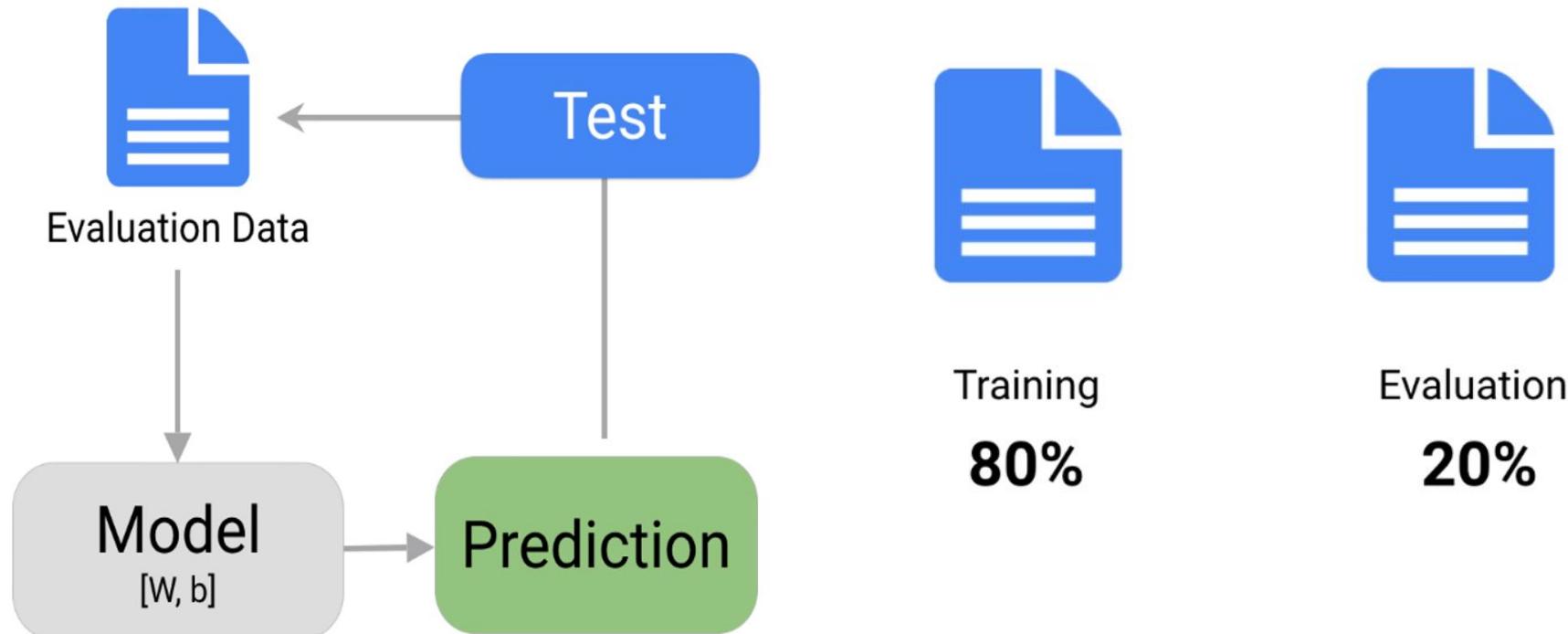
Two variables to train
the model m and b



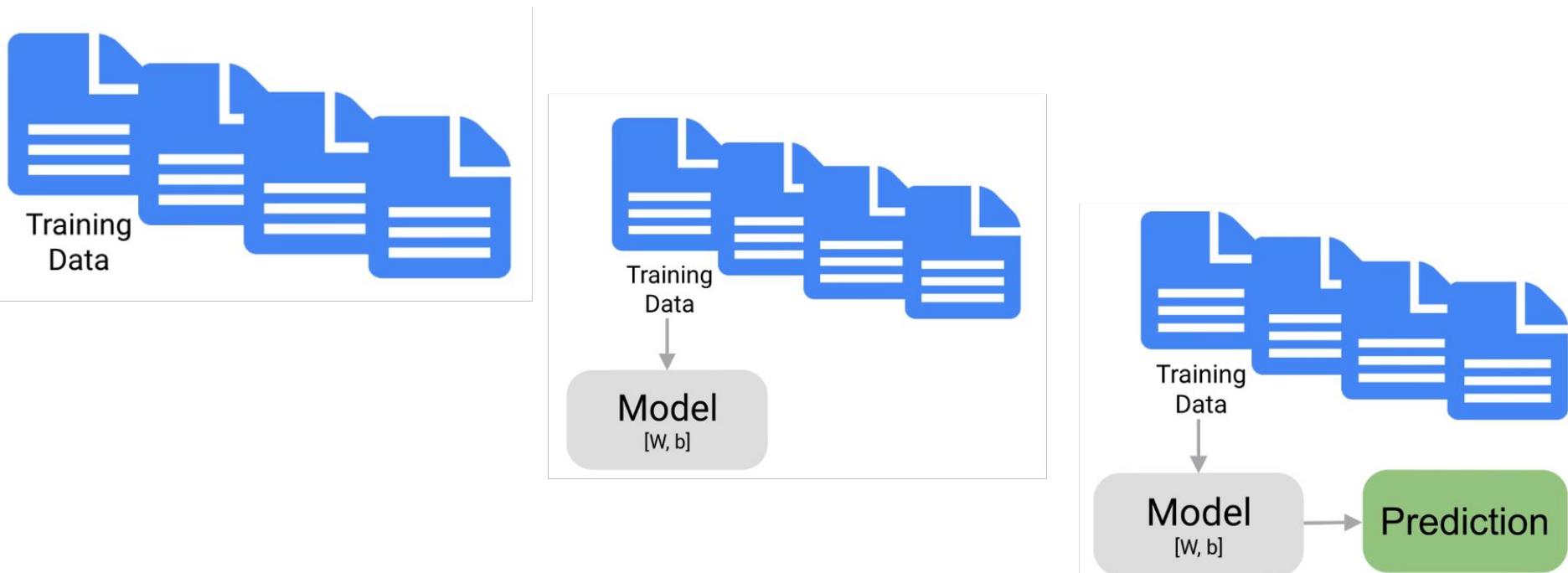
Step4: Training



Step5: Evaluate

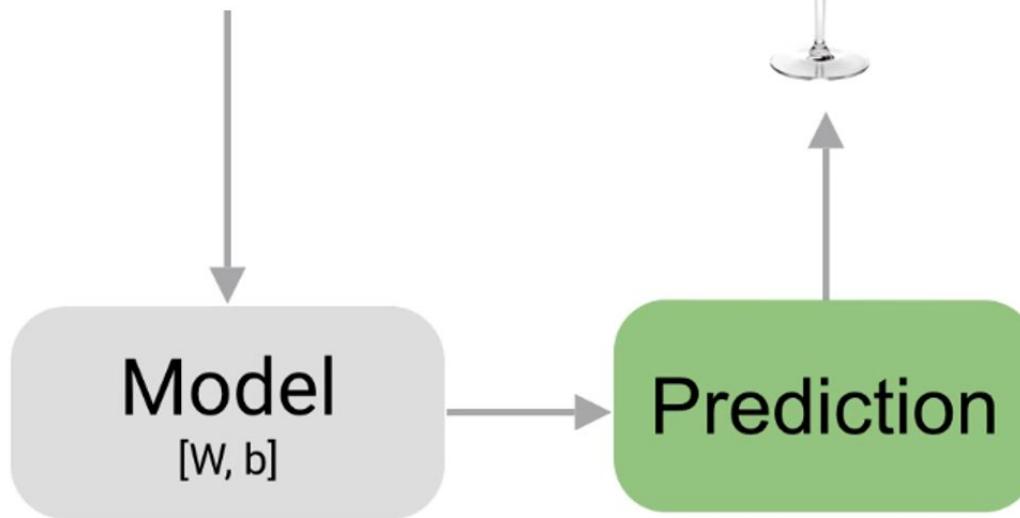


Step6: Refined parameters

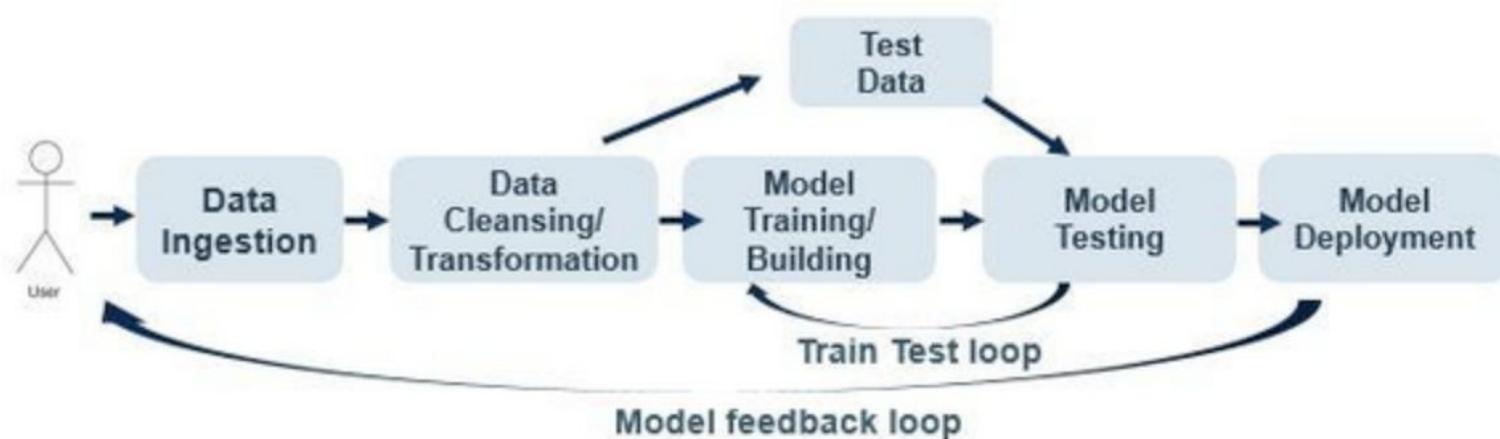


Step6: Prediction

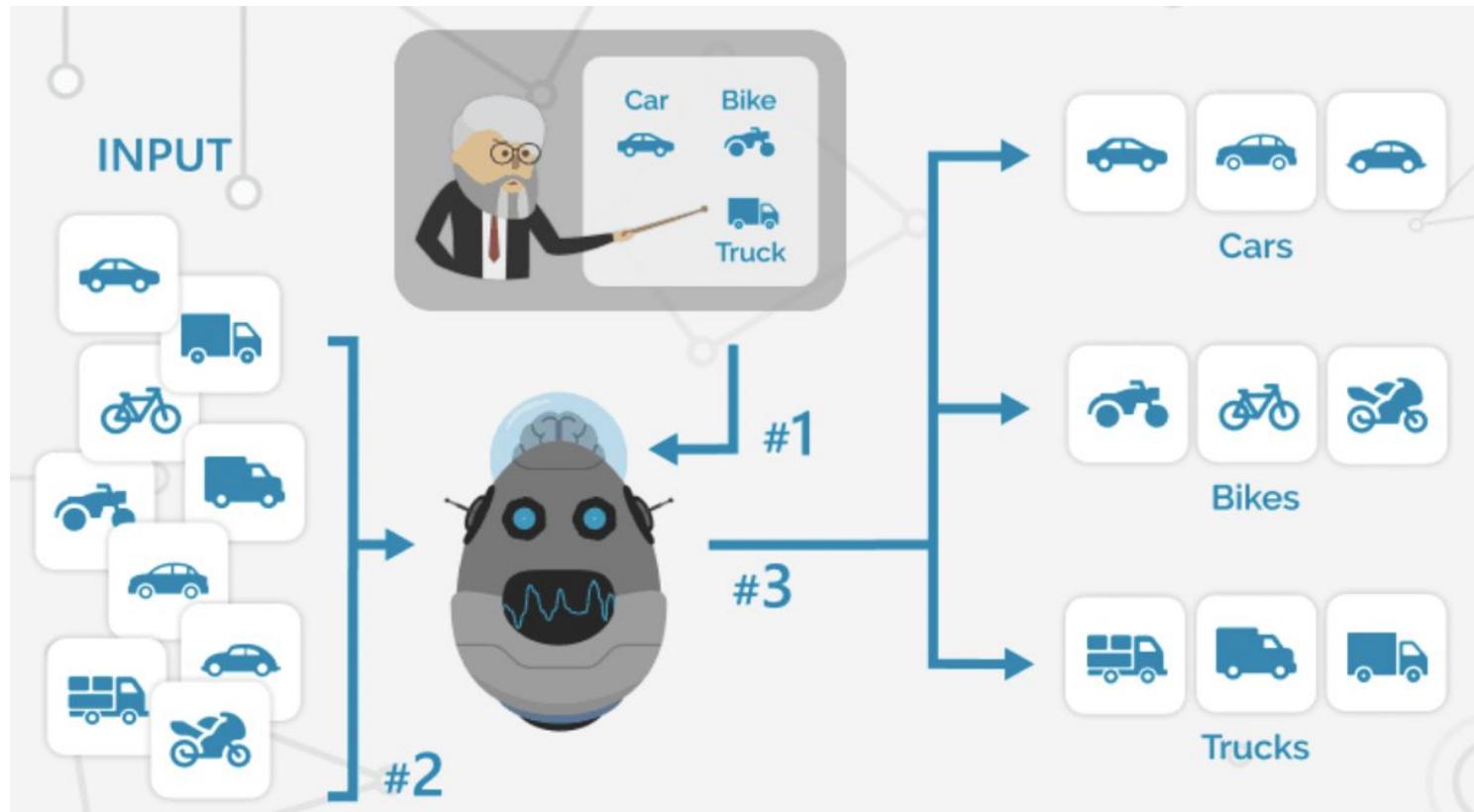
Color: 660nm
Alcohol: 12%



Typical machine learning cycle



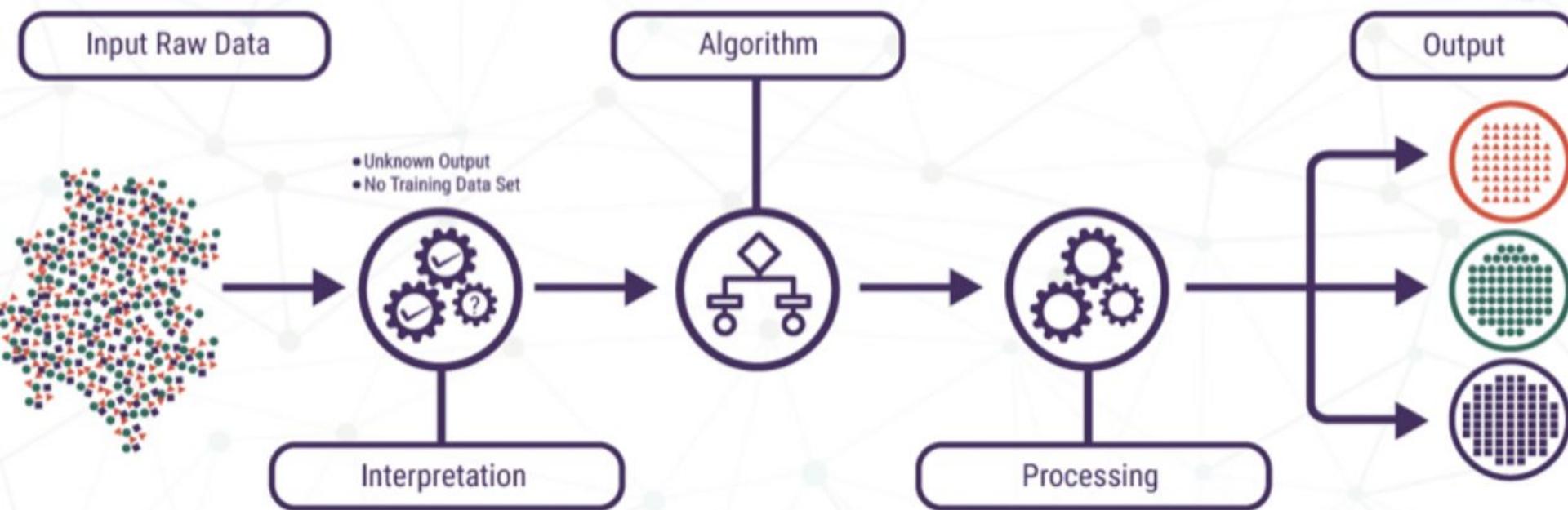
Supervised Learning



Supervised learning example

- Spam filtering - Classification
- Home prices - Regression
- Some disease recognition
- Facial recognition

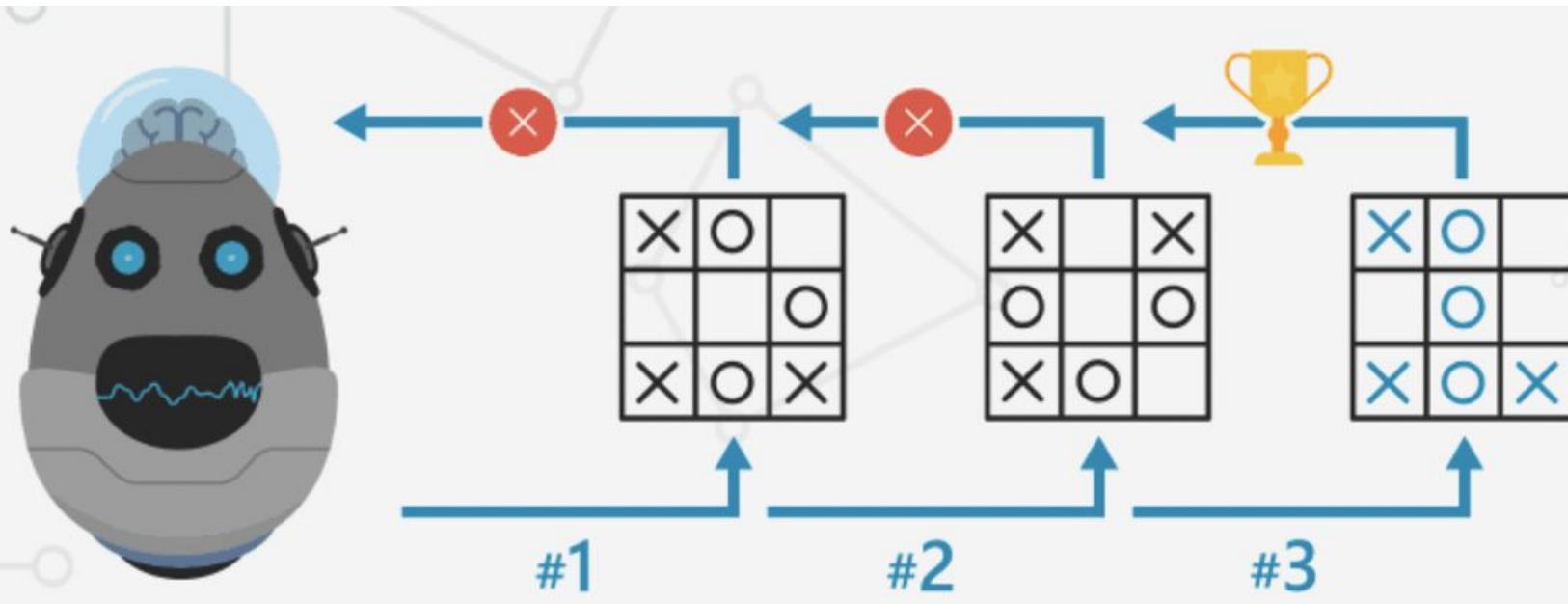
UnSupervised Learning



Unsupervised learning example

- Biology - for genetic and species grouping
- Medical imaging - for distinguishing between different kinds of tissues
- Market research - for differentiating groups of customers based on some attributes
- Recommender systems - giving you better Amazon purchase suggestions or Netflix movie matches.

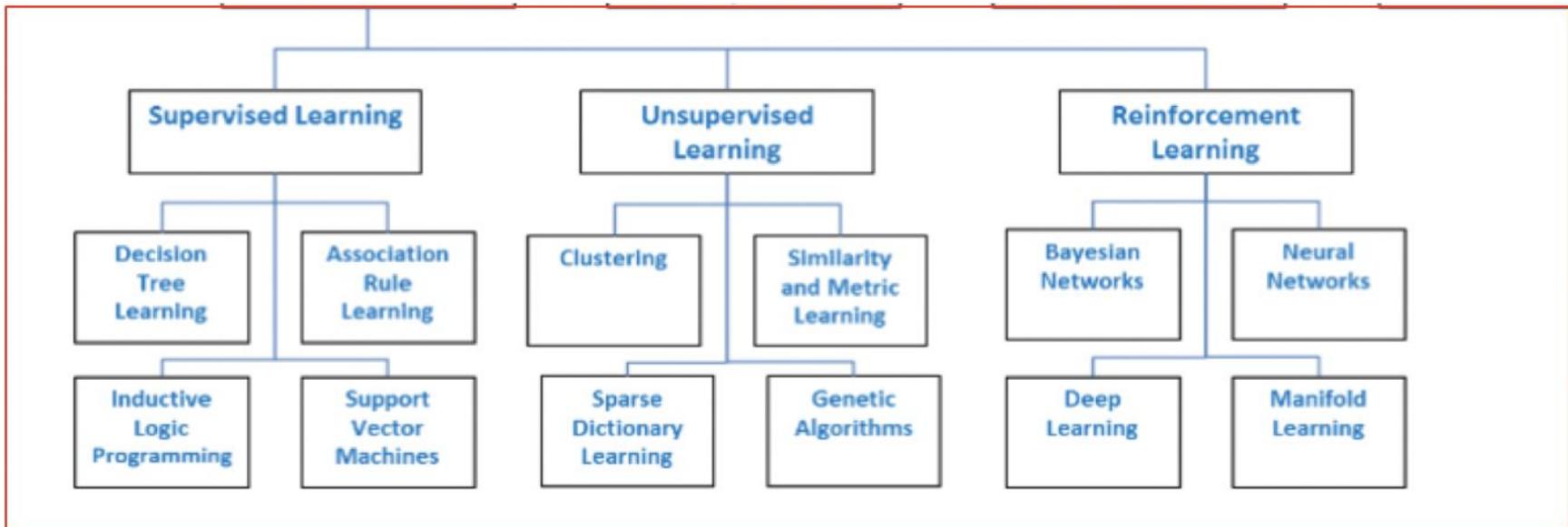
Reinforcement Learning



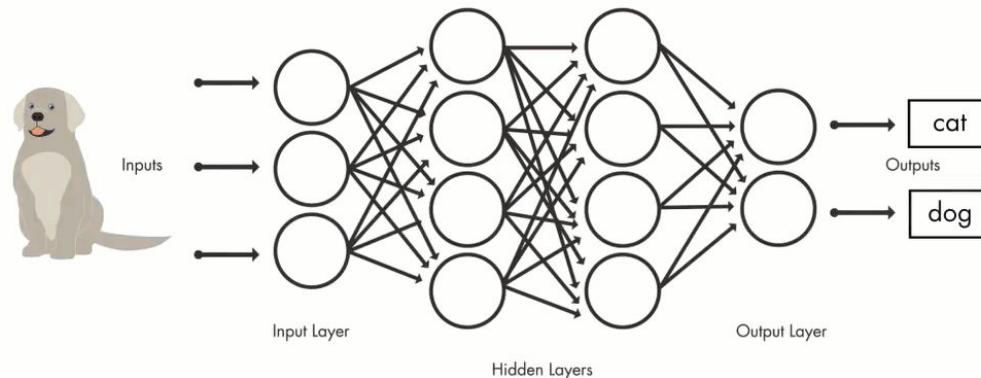
APRENDIZAJE POR REFUERZO

Reinforcement learning refers to goal-oriented algorithms, which learn how to attain a complex objective

Other Examples



What is deep learning?

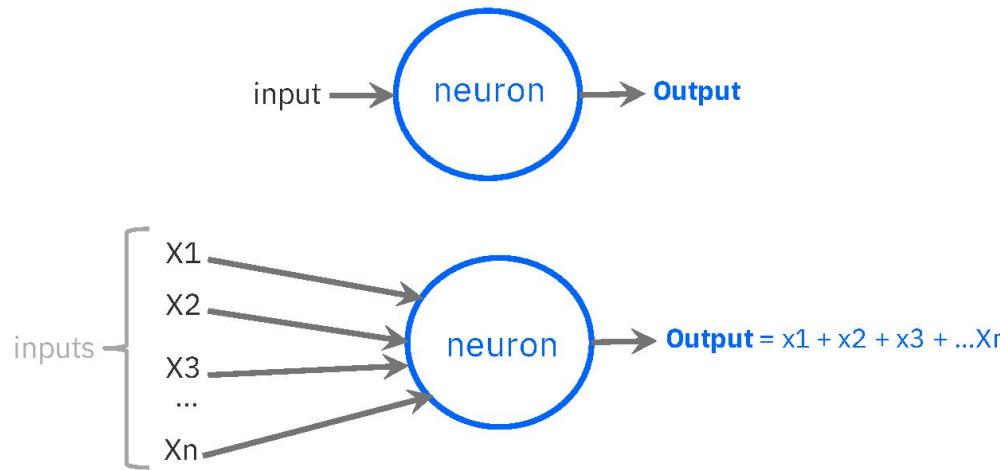


Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called **artificial neural networks**.

Neuron?

What is an artificial neuron?

Think of them as calculators



How it works ?

How does deep learning work?

- ① start with your data

data
data
data
data
data



- ② Define a neural network



- ③ Model learns to recognize patterns in historical data



trained model

- ④ Enter new data into your model



- ⑤ If patterns in the new data match the training data then the model makes accurate predictions

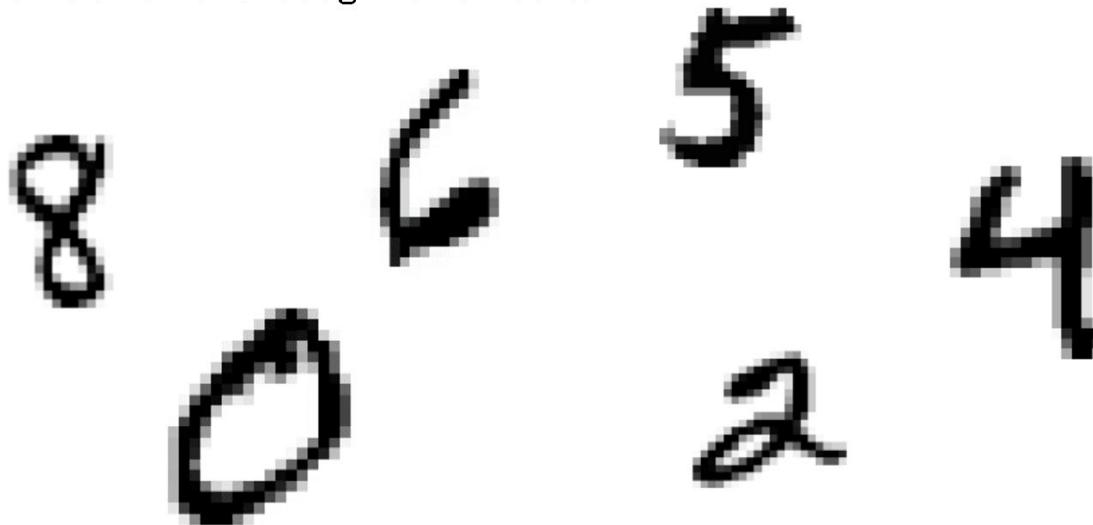


prediction

[TENSORFLOW PLAYGROUND](#)

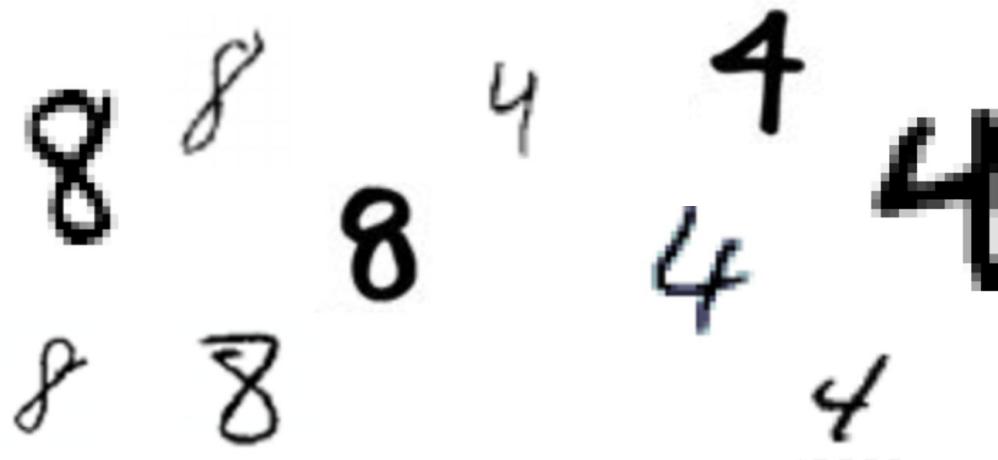
How do human recognize numbers?

How do humans recognize numbers?



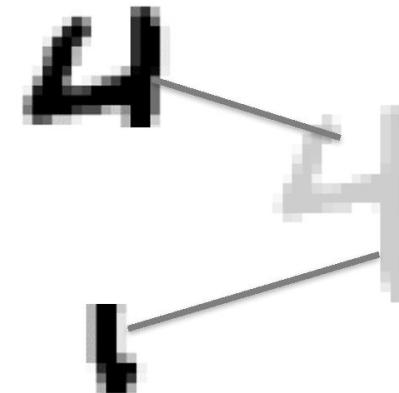
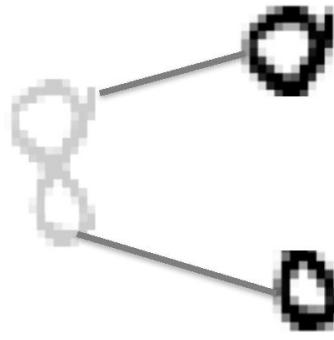
Human brains

Human brains detect patterns within variations



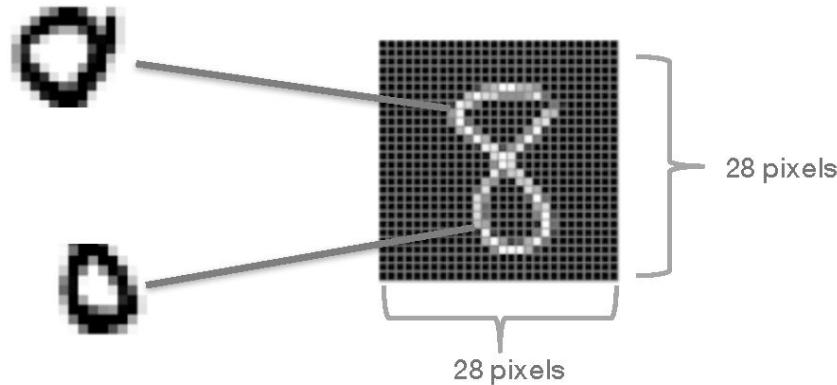
Decompose?

Perhaps by decomposing into sub-parts?



Pixels

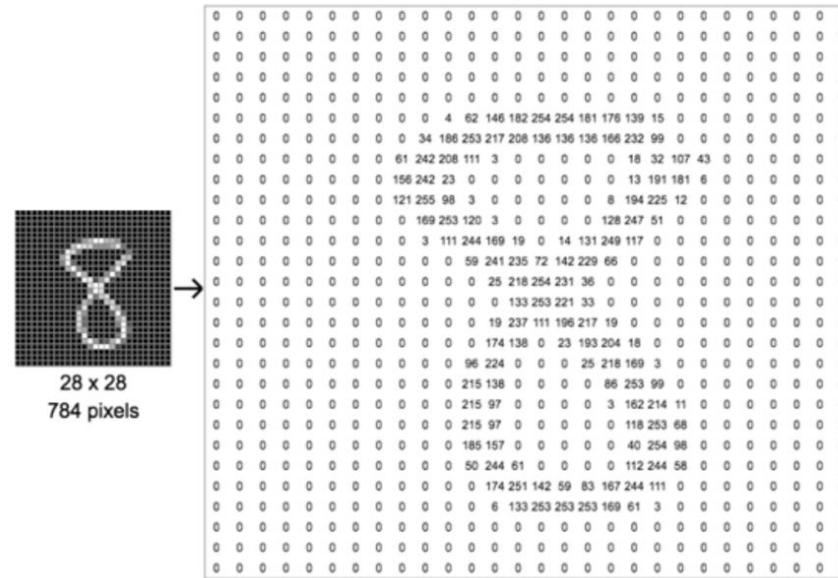
Pixels of an image capture variations in light and dark



Source: https://ml4a.github.io/ml4a/neural_networks/

How to teach a computer?

How do we teach a computer these 784 pixels are the number 8?

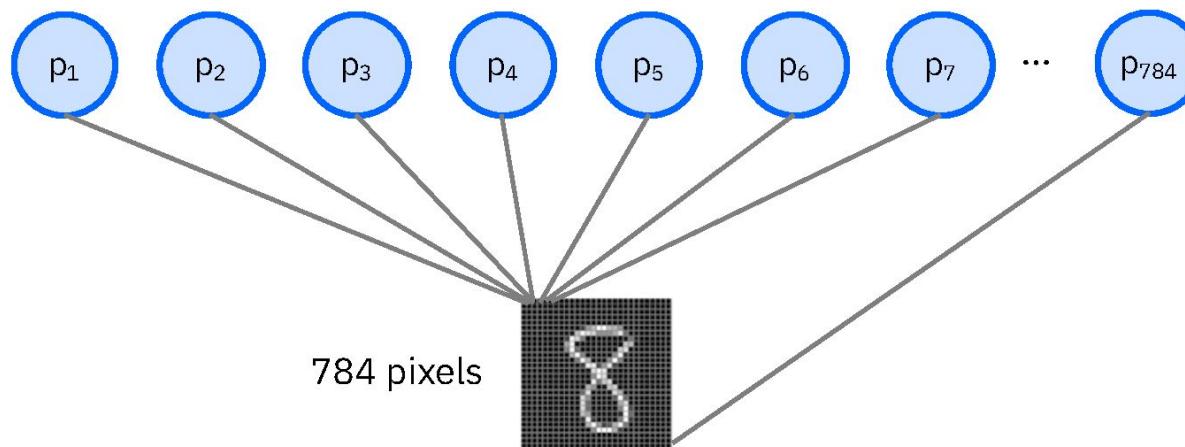


Source: https://ml4a.github.io/ml4a/neural_networks

Activation between 0 and 1

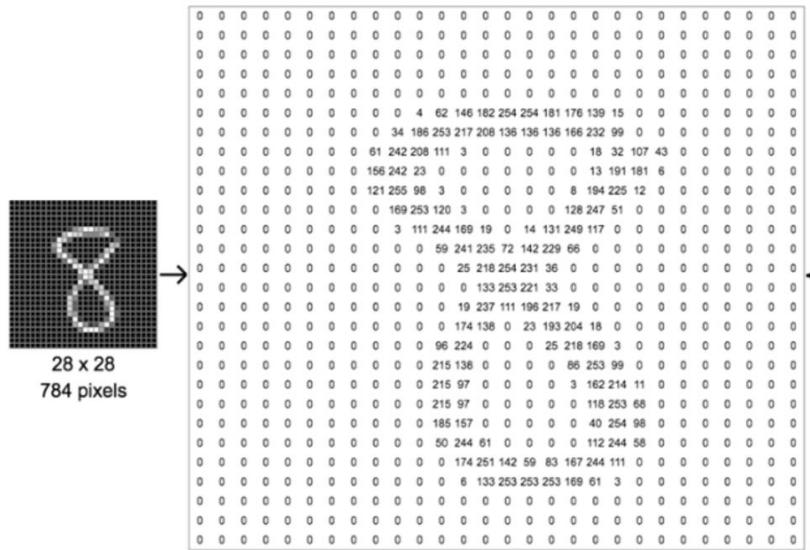
Our first layer of neurons

One neuron per pixel in our image example

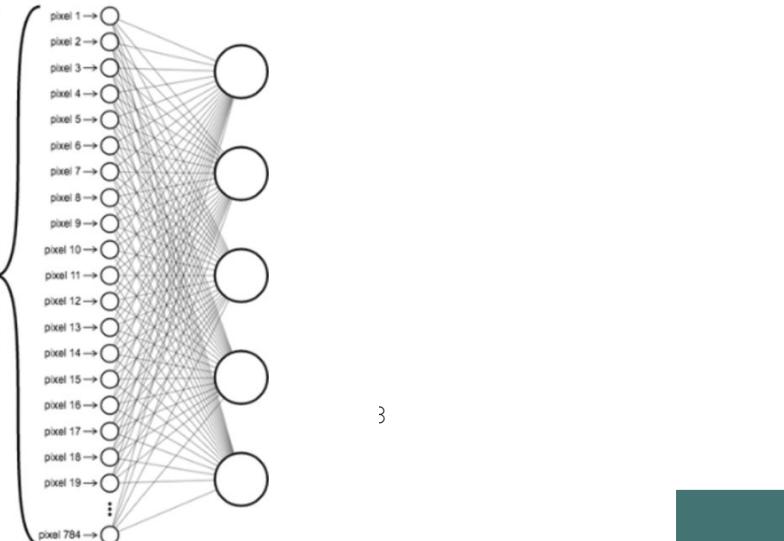


Identify the pixels

Putting it all together

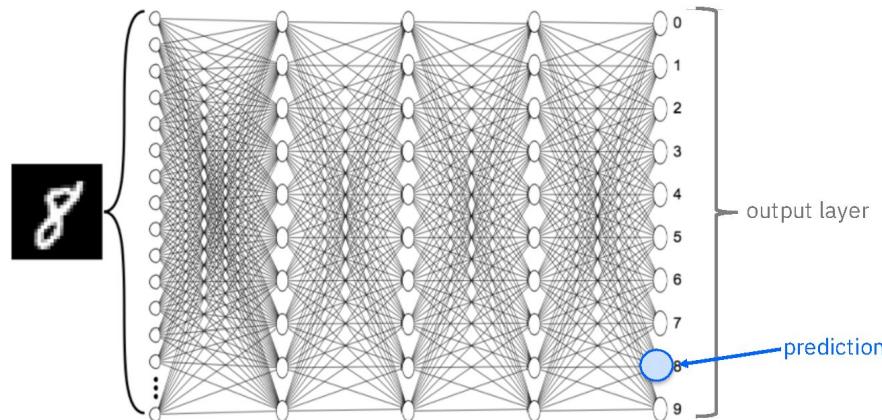


Source: https://ml4a.github.io/ml4a/neural_networks



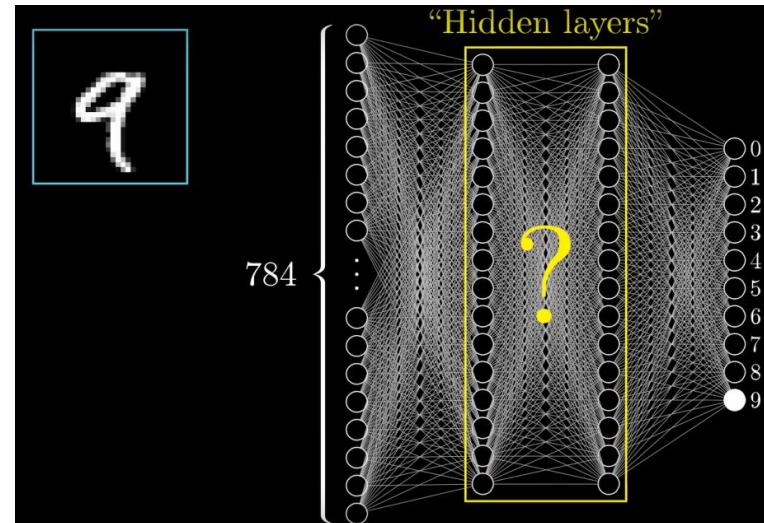
Each layer do additional calculation

Each layer transforms the input to match the desired output

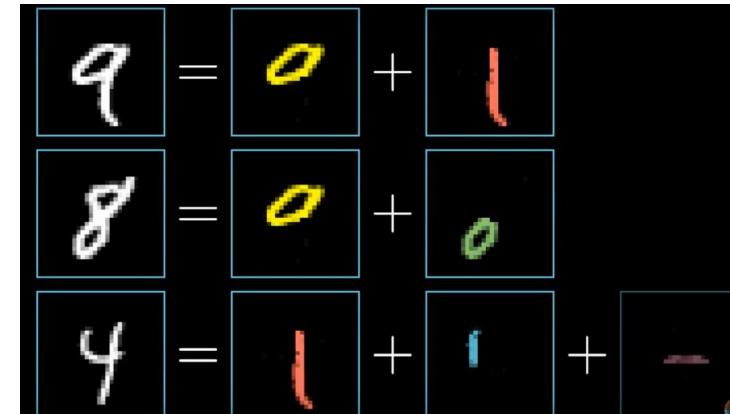
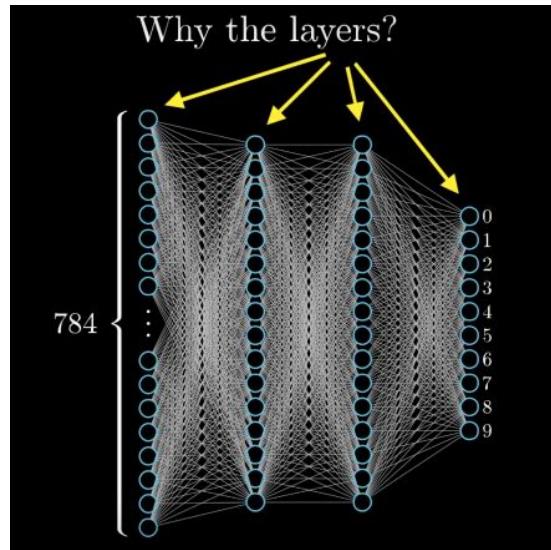


IBM Cloud / Watson and Cloud Platform / © 2018 IBM Corporation

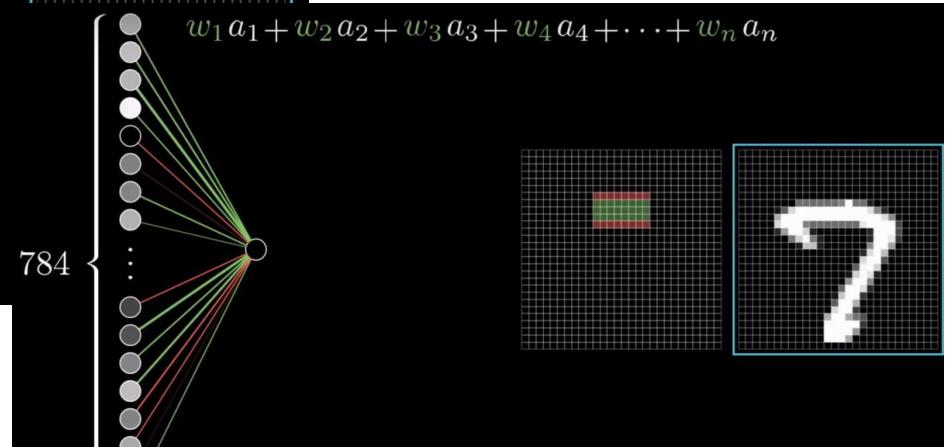
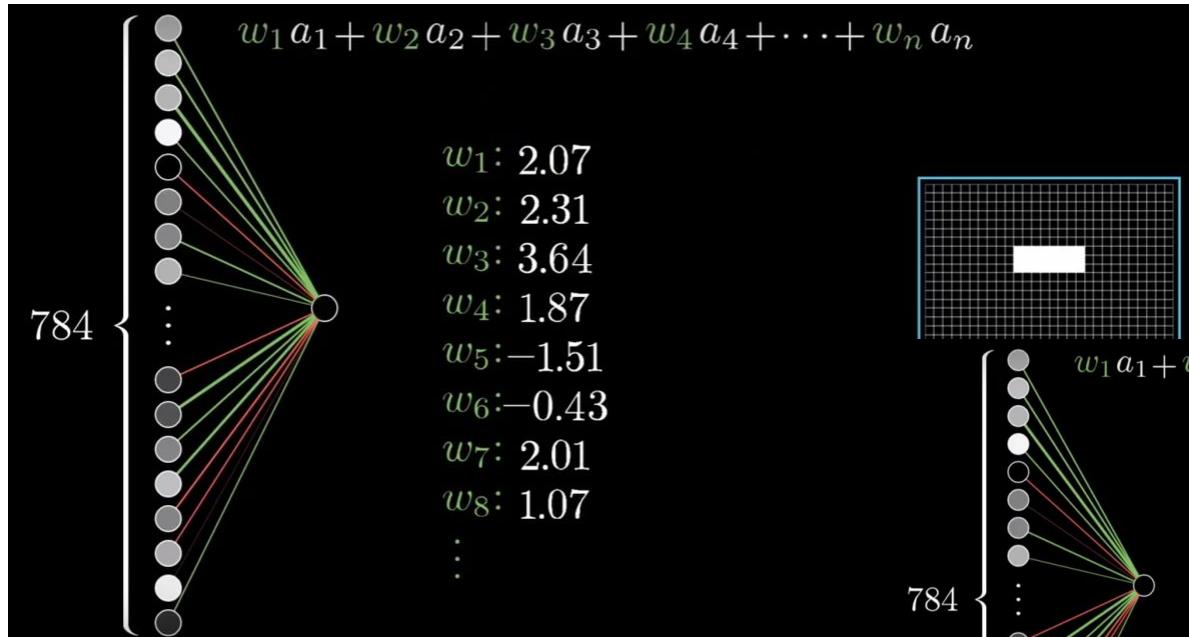
Source: https://mlda.github.io/mlda/neural_networks/



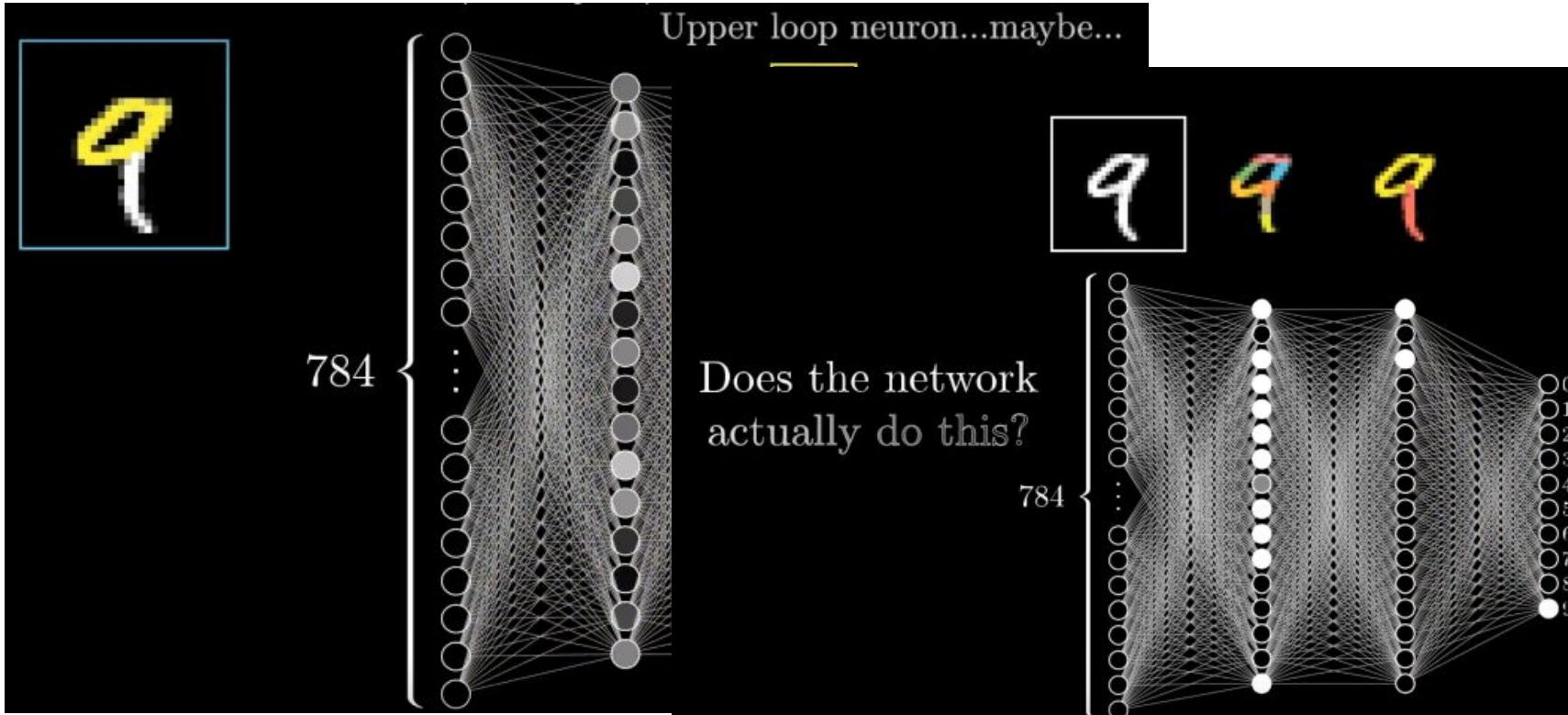
Why the layers?



We associate activations with “weight”

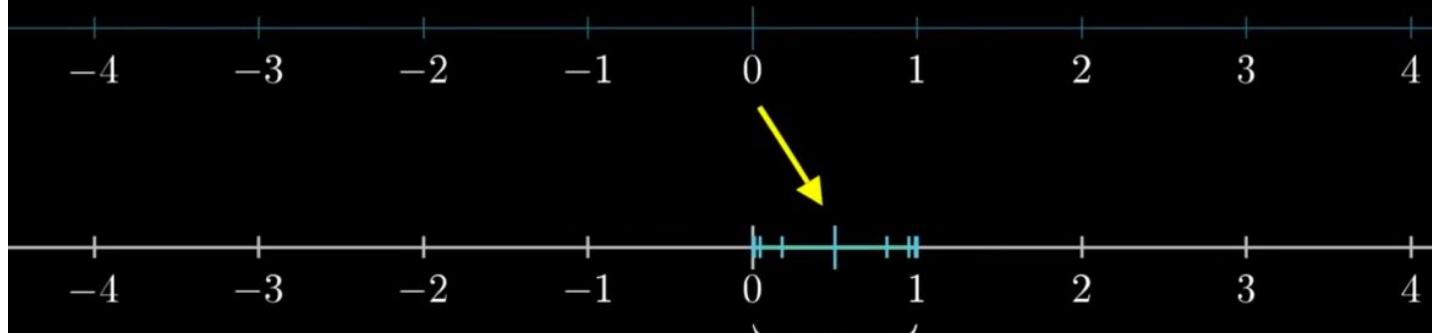


Different layers



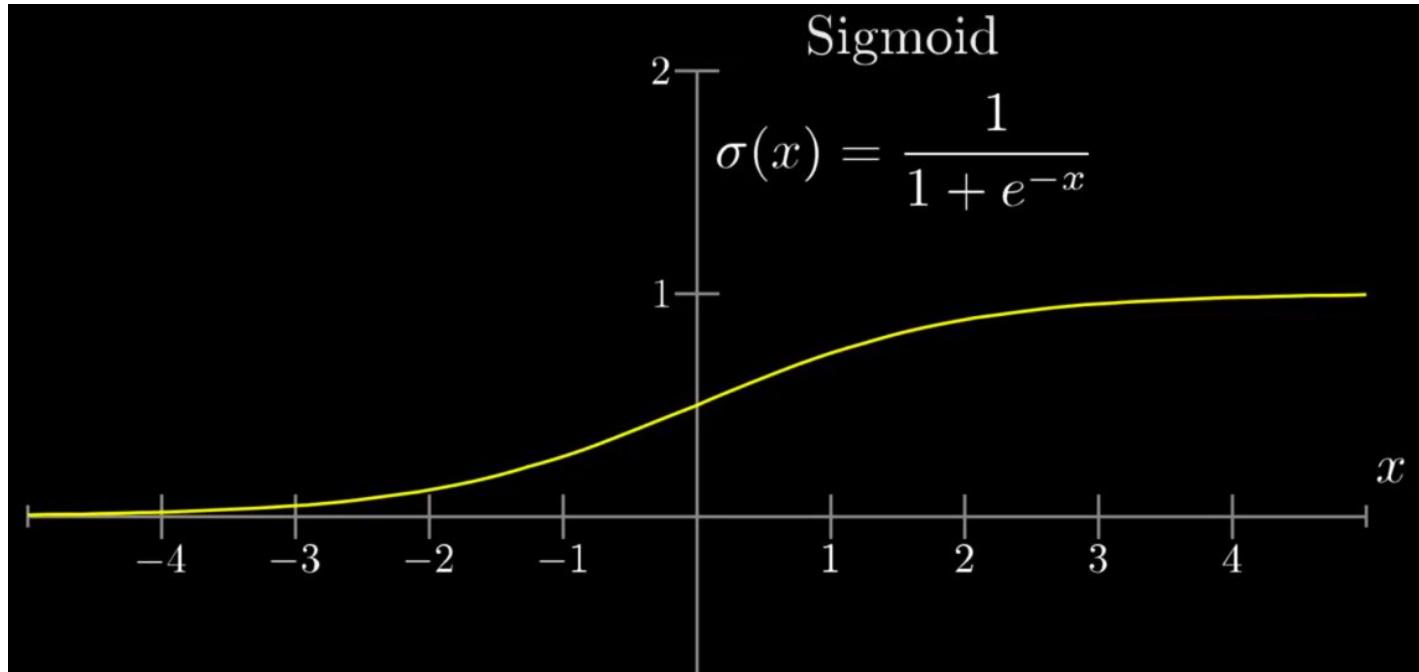
One formula?

$$w_1a_1 + w_2a_2 + w_3a_3 + w_4a_4 + \cdots + w_na_n$$

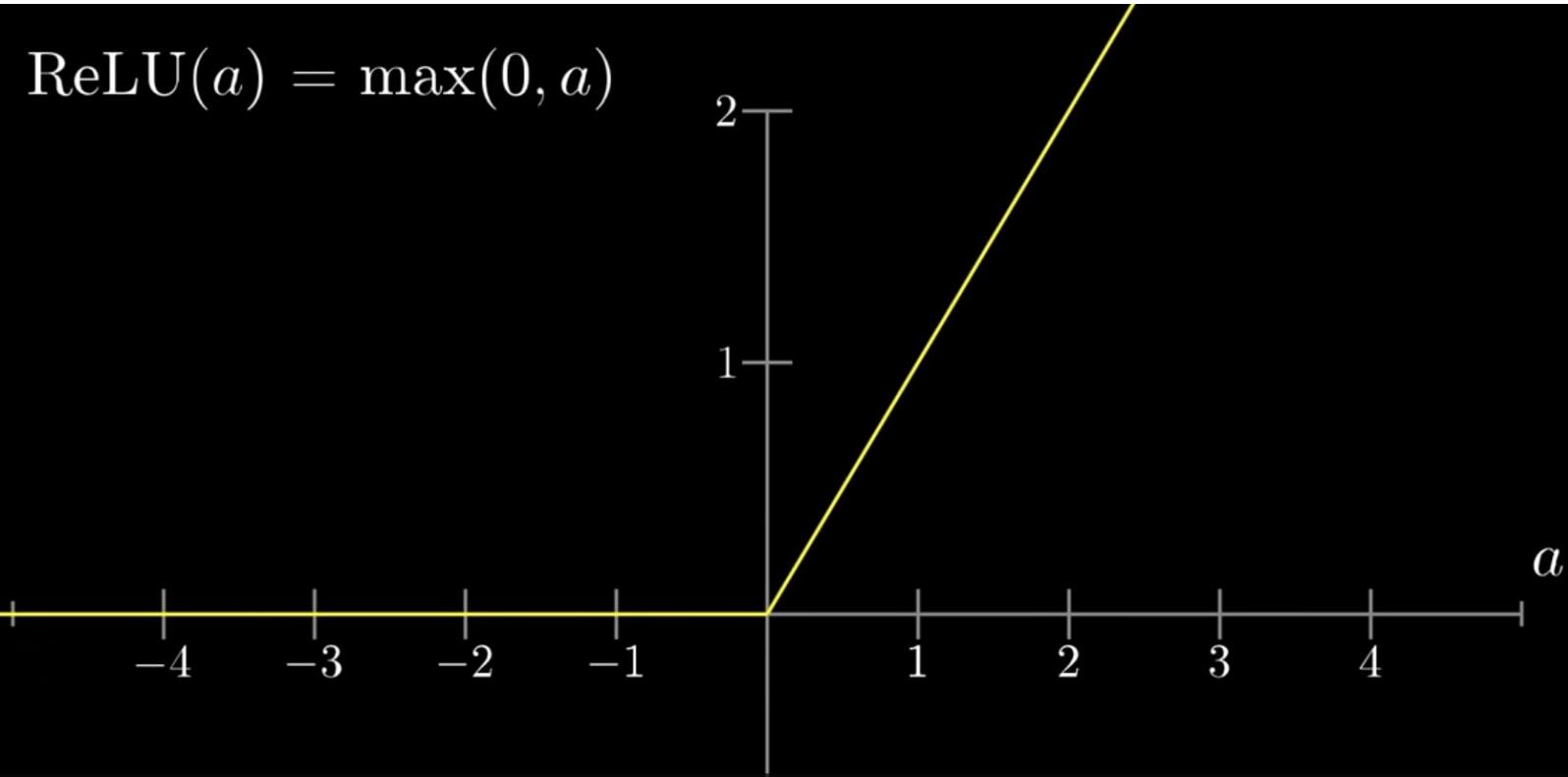


Activations should be in this range

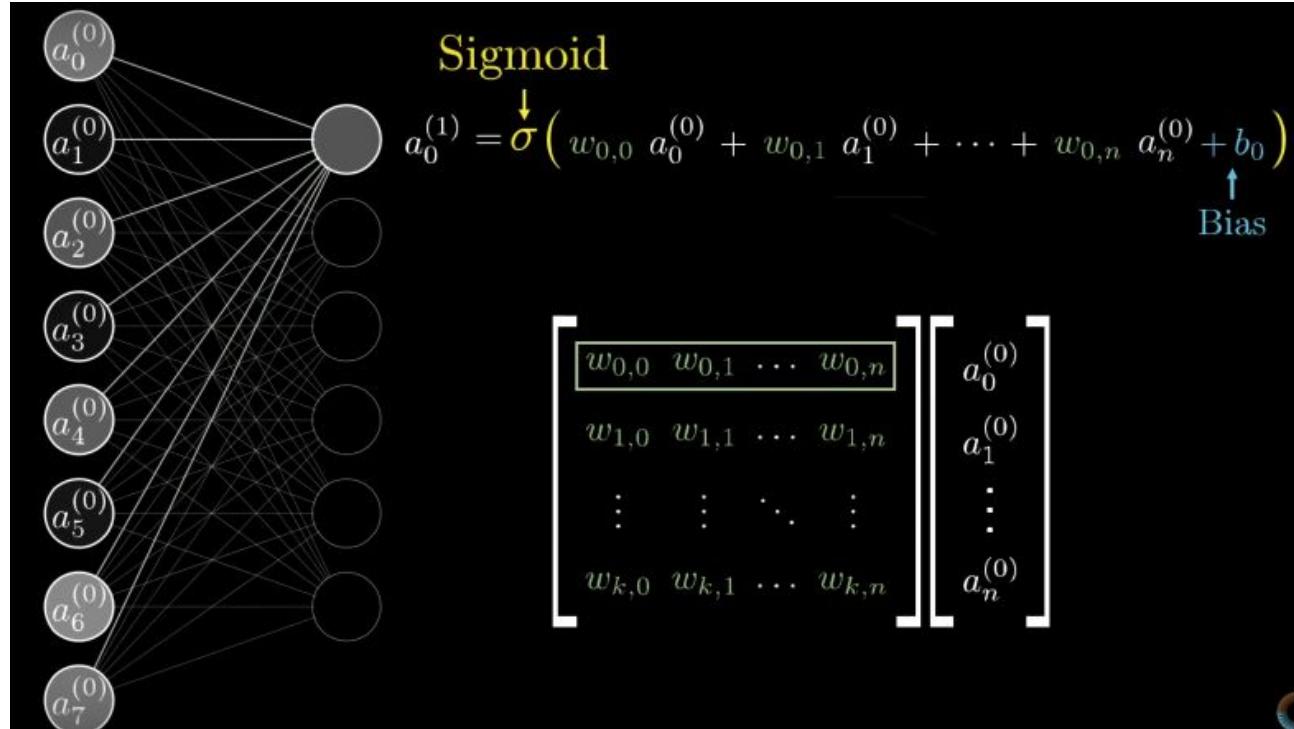
Sigmoid



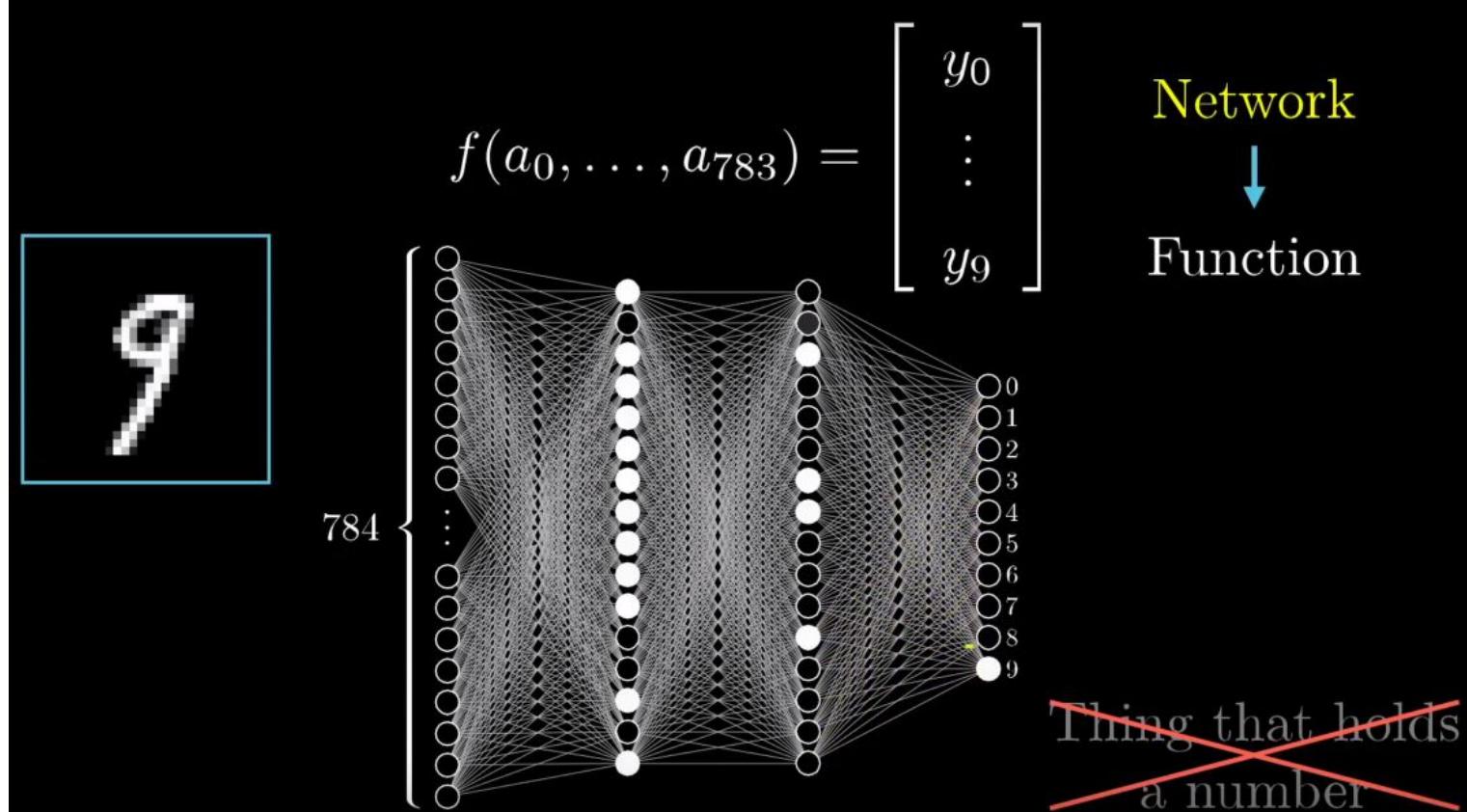
$$\text{ReLU}(a) = \max(0, a)$$



One mathematical formula: lineal algebra



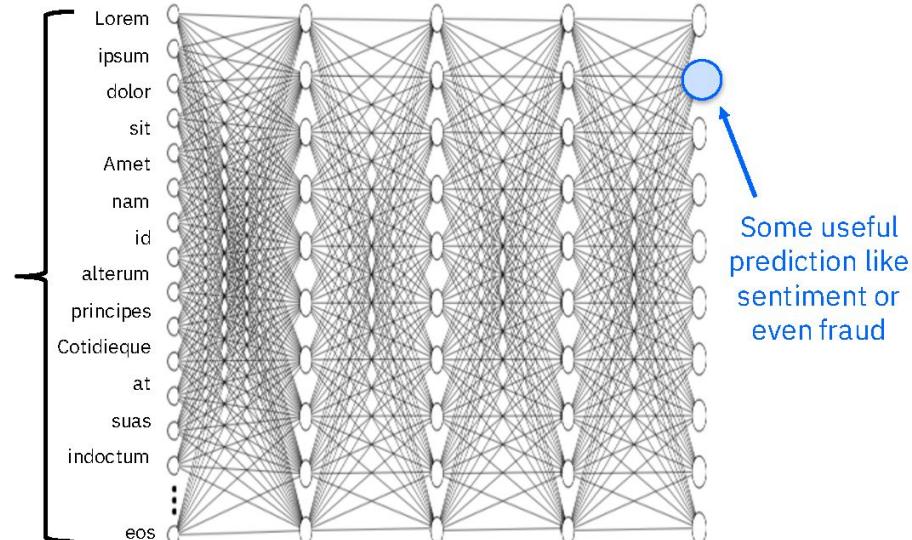
The network is a function



Not only graphics

Not just pixels. Text, sound and much more can be used as input!

 Lorem ipsum dolor sit amet, nam id
 alterum principes cotidieque, at
 suis indoctum his. No inani soleat
 sed, per illum quaestio id. No
 prompta luptatum sit. His alii
 alterum feugiat ne. Eu delenit
 expetendis duo, no possit utamur
 patrioquemei. Admodum
 appellantur at quo, albucius
 periculis adolescens an mel, veri
 quaerendum sea ut. Eam noluisse
 copiosae democritum ei, cu eos.



IBM Cloud / Watson and Cloud Platform / © 2018 IBM Corporation

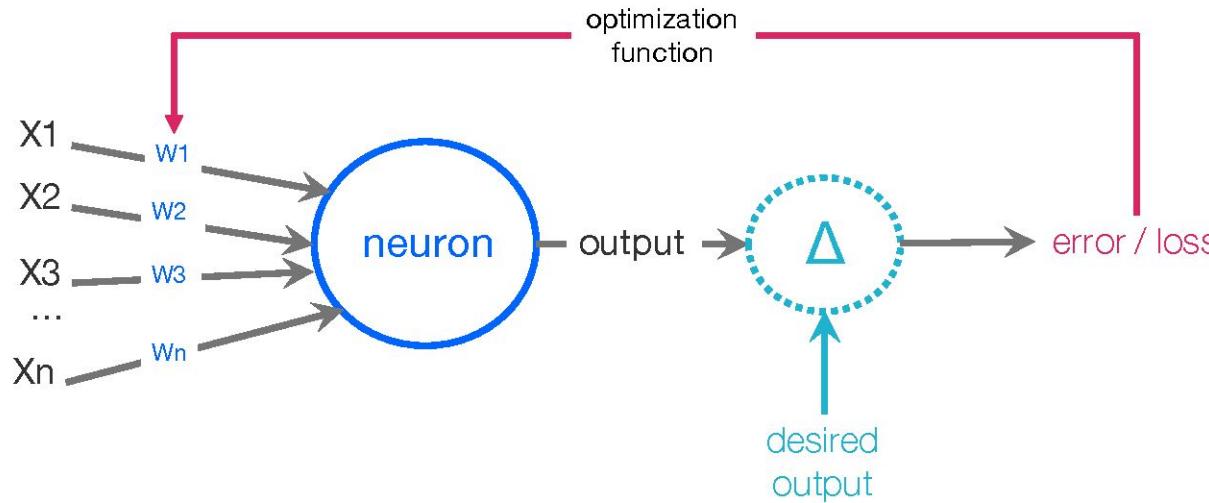
Source: https://ml4a.github.io/ml4a/neural_networks/

Video recomendado
<https://www.youtube.com/watch?v=aircAruvnKk>

Iterations

Backpropagation: Iteratively train a neuron

Adjust weights until the output matches expectation



How it works ?

How does deep learning work?

- ① start with your data

data
data
data
data
data



- ② Define a neural network



- ③ Model learns to recognize patterns in historical data



trained model

- ④ Enter new data into your model



- ⑤ If patterns in the new data match the training data then the model makes accurate predictions

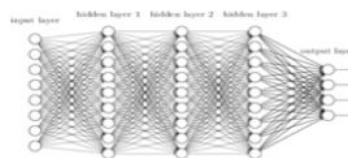
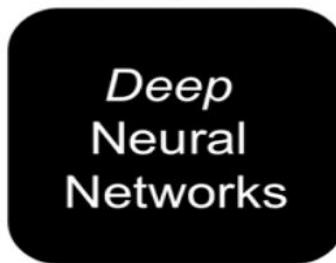


prediction

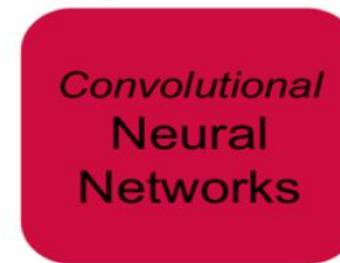
[TENSORFLOW PLAYGROUND](#)

Deep learning

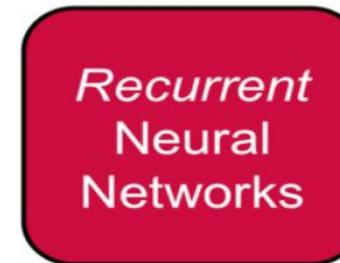
providing lift for classification and forecasting models



feature extraction and classification of images

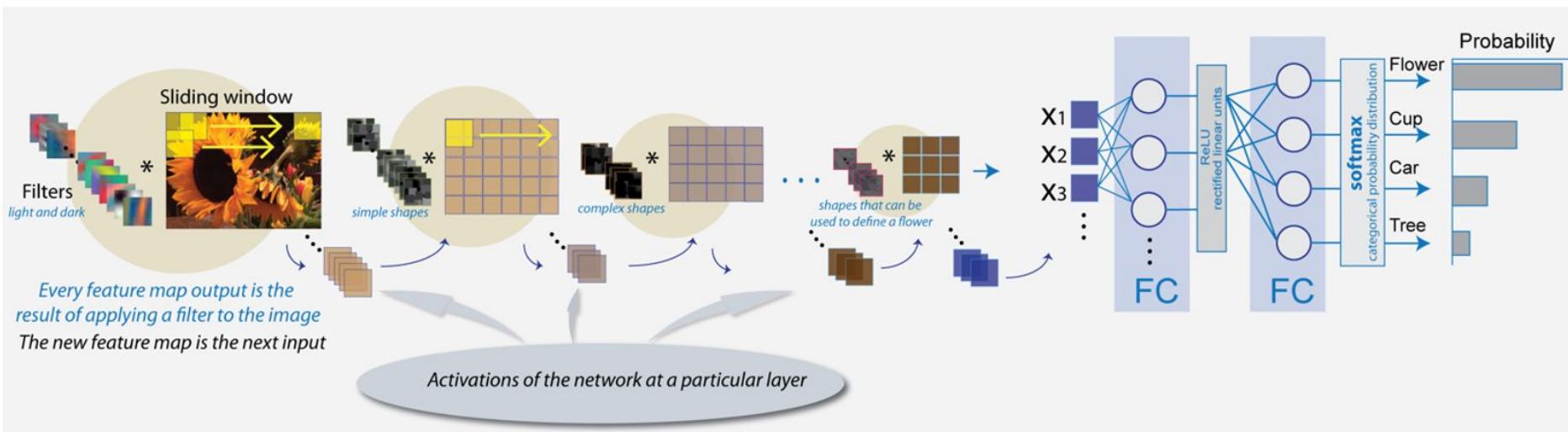
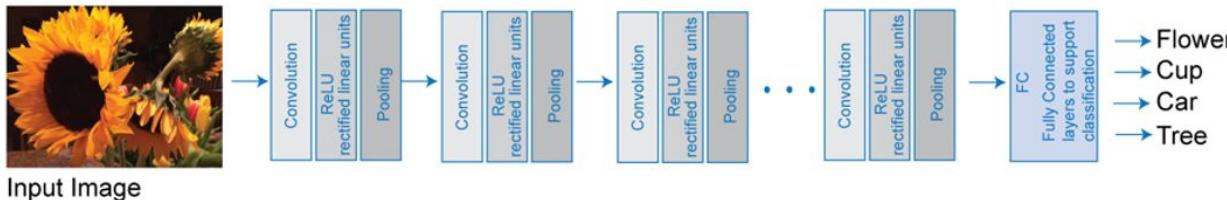


for sequence of events, language models, time series, etc.

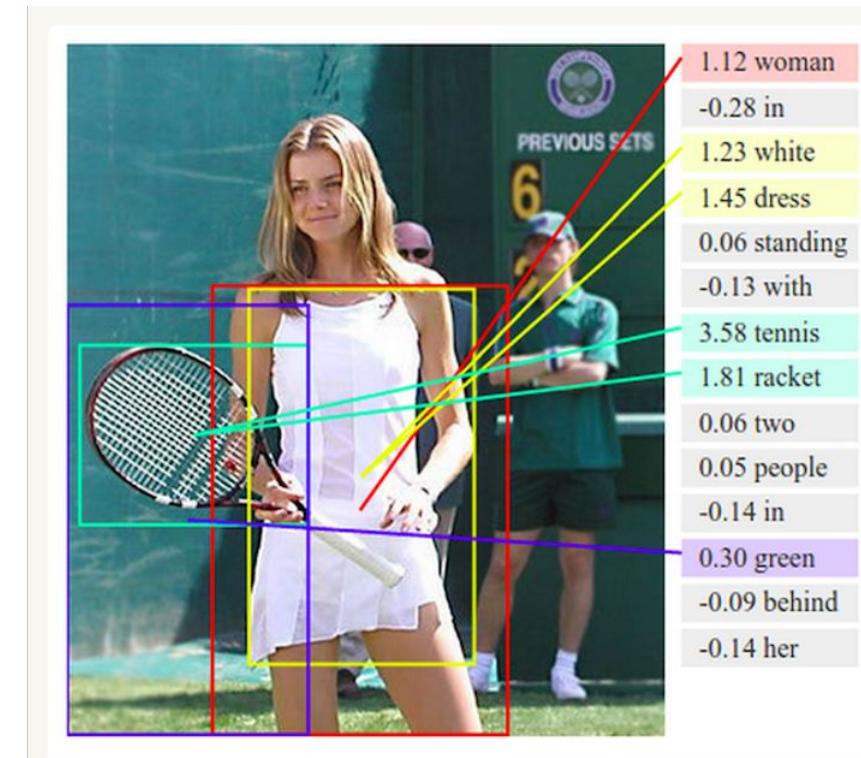
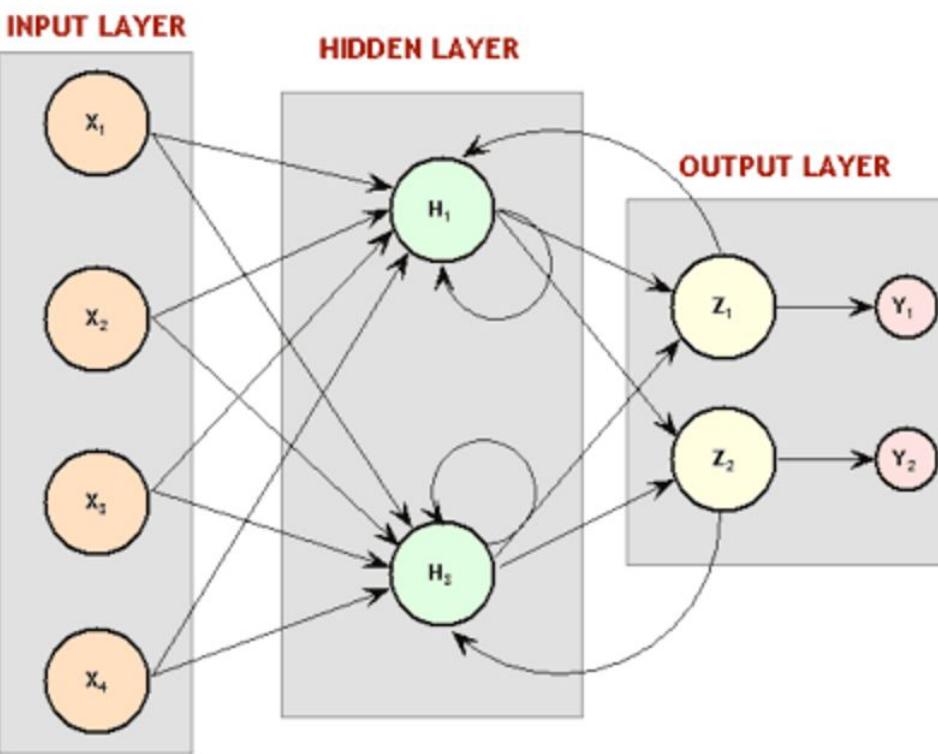


Convolutional neural network

Assume the data are images

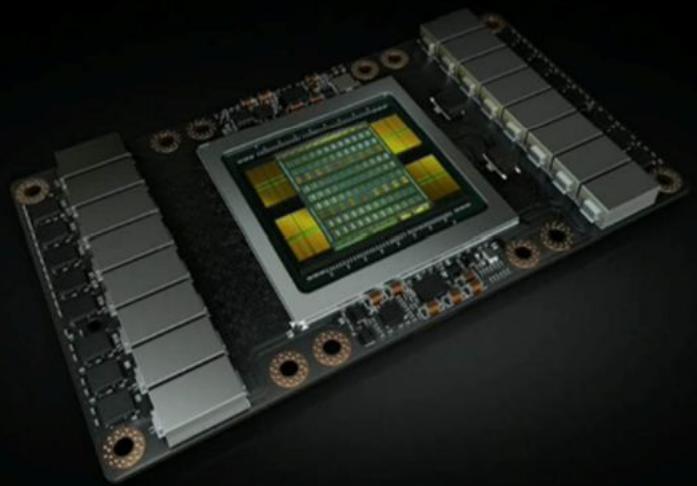


Recurrent neural network



New capacity processing

GPU = Graphics Processing Unit



TESLA V100

GIANT LEAP FOR AI & HPC
VOLTA WITH NEW TENSOR CORE

21B xtors | TSMC 12nm FFN | 815mm²

5,120 CUDA cores

7.5 FP64 TFLOPS | 15 FP32 TFLOPS

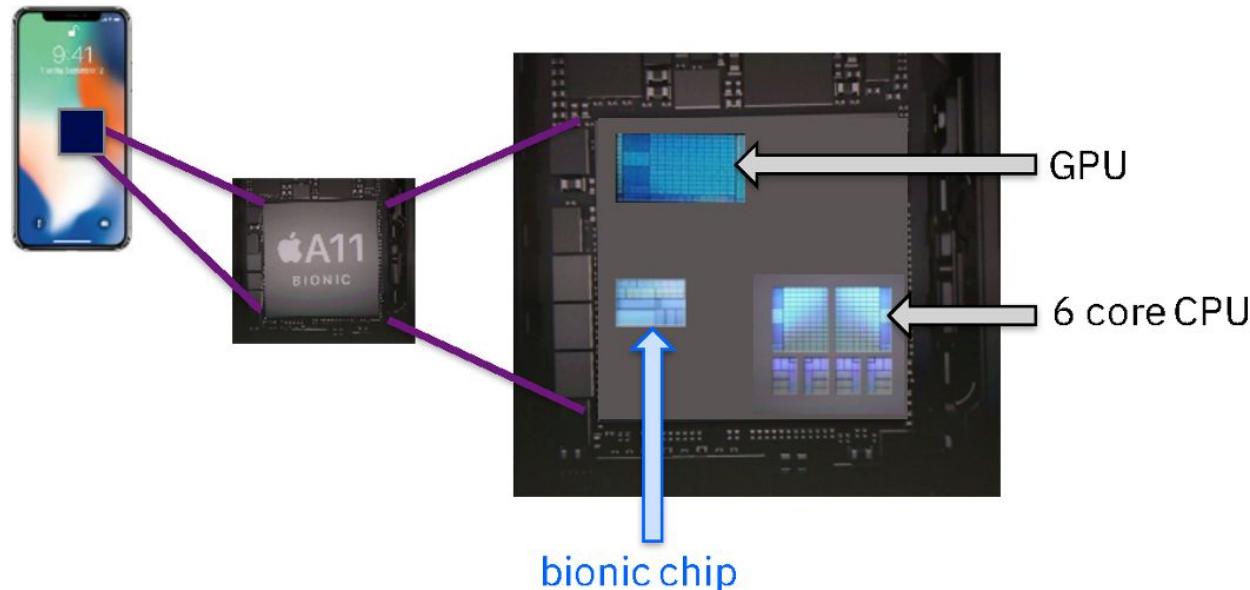
NEW 120 Tensor TFLOPS

20MB SM RF | 16MB Cache | 16GB HBM2 @ 900 GB/s

300 GB/s NVLink

New chips

iPhone X's Bionic chip: built in machine learning



Multiframework approach deep learning

Take a Multi-Framework Approach to Deep Learning



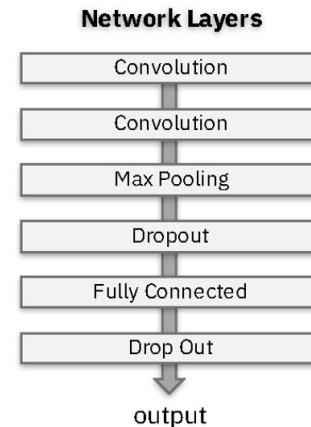
New frameworks emerging monthly.

Tensorflow was awesome yesterday but has static graphs so PyTorch's dynamic graphs are now popular

Keras as an example

A neural network in Keras

```
model = Sequential()  
model.add(Convolution2D(32, 3, 3, activation='relu',input_shape=(1,28,28)))  
  
model.add(Convolution2D(32, 3, 3, activation='relu'))  
  
model.add(MaxPooling2D(pool_size=(2,2)))  
  
model.add(Dropout(0.25)) model.add(Flatten())  
  
model.add(Dense(128, activation='relu'))  
  
model.add(Dropout(0.5))  
  
model.add(Dense(10, activation='softmax'))
```



Watson studio: Combines different framework

Watson Studio: Tools for building AI applications

The screenshot shows the Watson Studio interface. At the top, there's a navigation bar with links for IBM Watson, Projects, Tools, Catalog, Community, Services, Docs, Support, Manage, and a user icon. Below the navigation bar, the page title is "My Projects / Deep Learning Experiments". There are buttons for "Add to project", "New notebook", and other management options. The main content area is divided into sections: "Notebooks" and "Experiments".

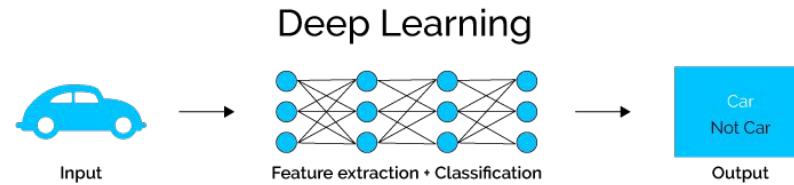
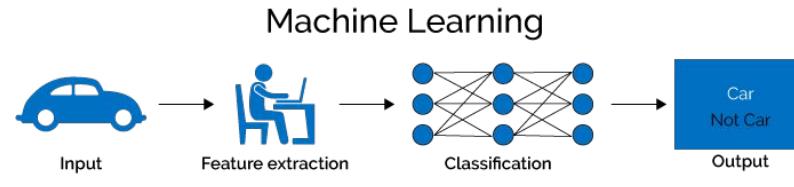
Notebooks

NAME	SHARED	SCHEDULED	STATUS	LANGUAGE	LAST EDITOR	LAST MODIFIED	ACTIONS
Getting Started with Keras				Python 3.5	Anthony Stevens	3 Apr 2018	
Keras mnist experiment with HPO				Python 3.5	Anthony Stevens	2 Apr 2018	
Style transfer experiments with Watson Machine Learning					Anthony Stevens	2 Apr 2018	

Experiments

NAME	CREATED BY	LAST MODIFIED	ACTIONS
CIFAR-p100-rbfopt=25		14 Apr 2018	
CIFAR-p100-random=25		14 Apr 2018	

ML vs DL



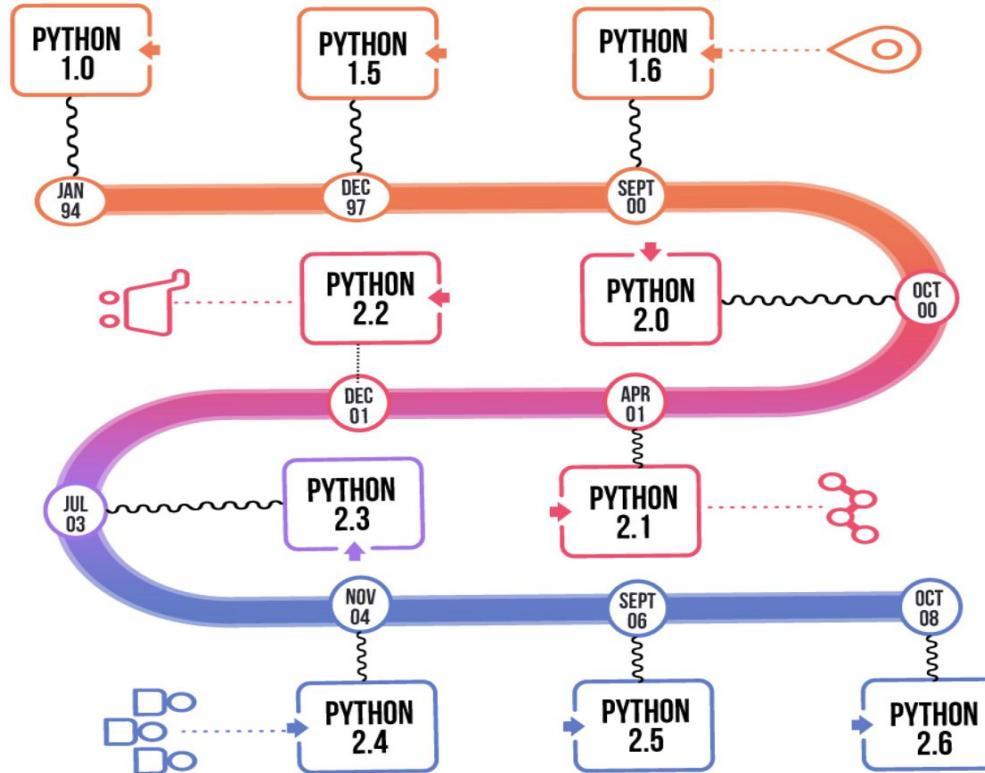
Machine Learning

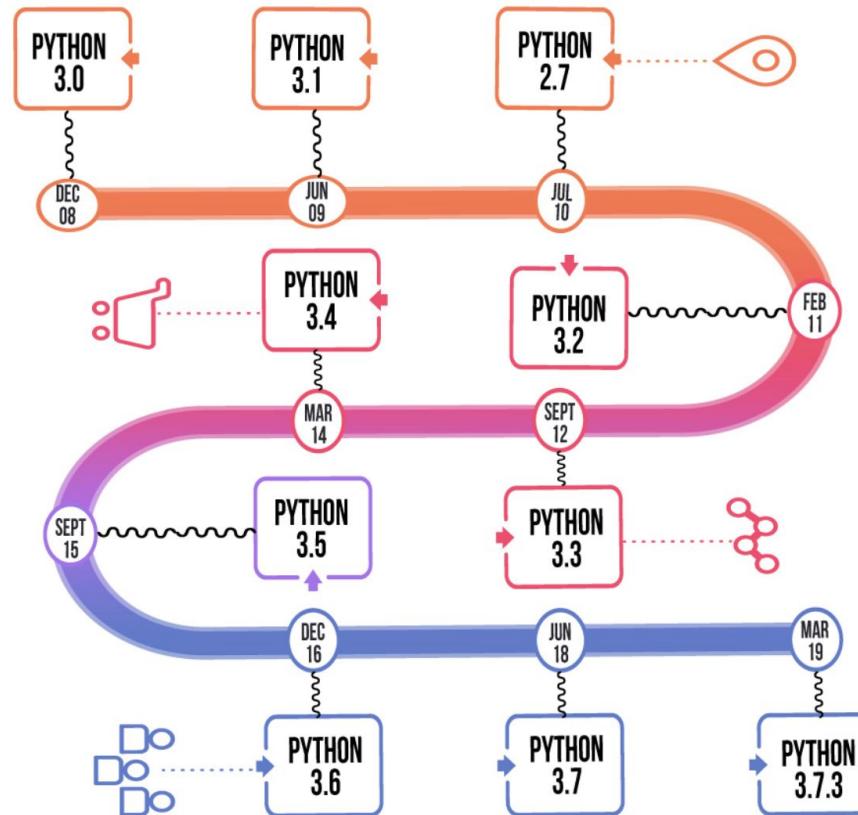
- + Small datasets
- + Normally training is fast
- Manual extraction of characteristics
- Limited accuracy

Deep Learning

- It needs bigger datasets
- Intensive computation for training
- + Automatic extraction
- + Better accuracy

why python?





Python

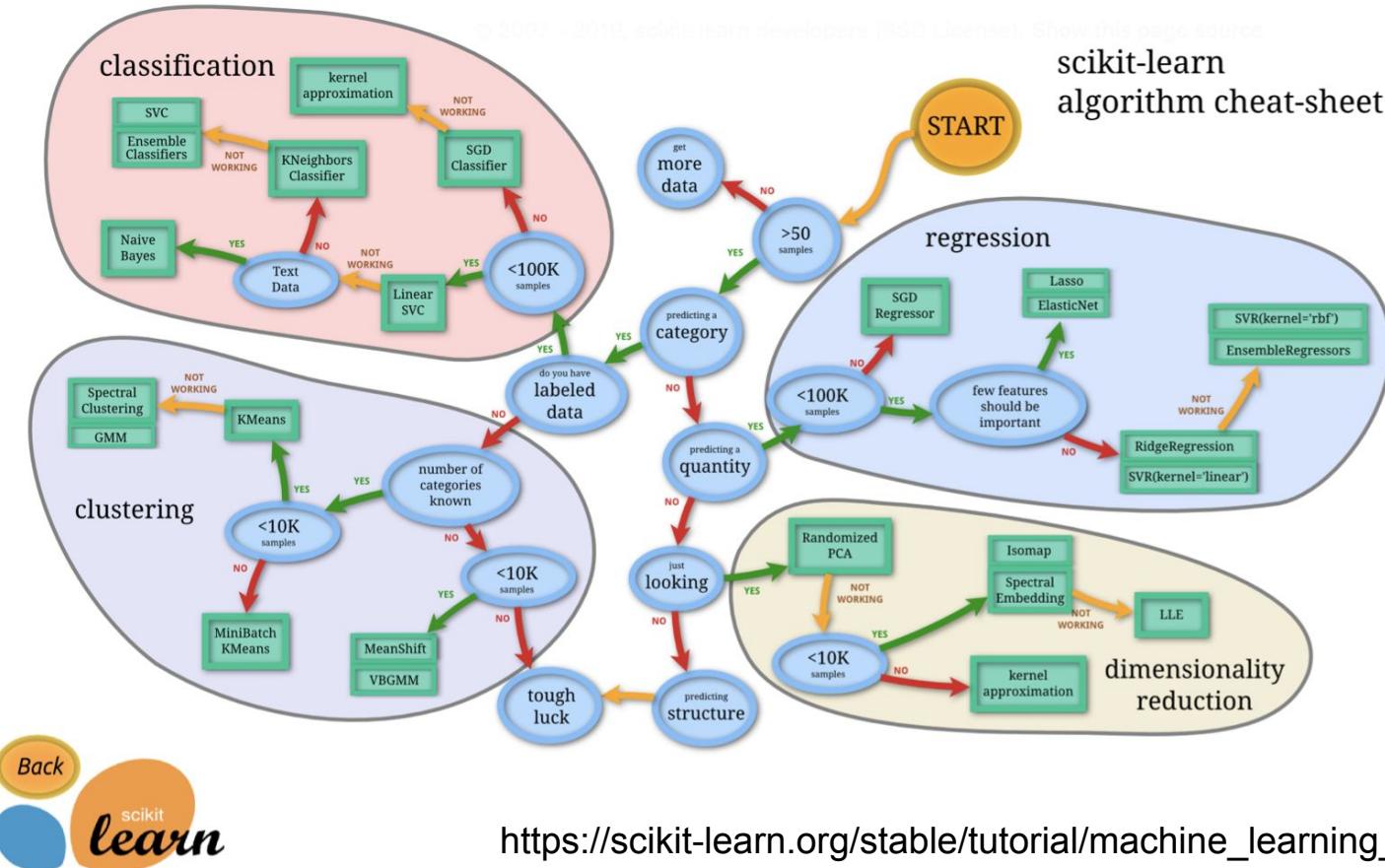
- Purpose general language
- Easy syntax compared to other languages
- Python community
- More than 70.000 libraries

Library Name	Type	Commits	Contributors	Releases	Watch	Star	Fork	Commits/ Contributors	Commits/ Releases	Star/ Contributors
 matplotlib	Visualization	25 747	725	70	498	7 292	398	36	368	10
 Bokeh	Visualization	16 983	294	58	363	7 615	2 000	58	293	26
 plotly	Visualization	2 906	48	8	198	3 444	850	61	363	72
 Seaborn	Visualization	2 044	83	13	205	4 856	752	25	157	59
 pydot	Visualization	169	12	12	17	193	80	14	14	16
 scikit-learn	Machine learning	22 753	1 084	86	2 114	28 098	14 005	21	265	26
XGBoost LightGBM CatBoost	Machine learning	3277 1083 1509	280 79 61	9 14 20	868 363 157	11 991 5 488 2 780	5 425 1 467 369	12 14 25	364 77 75	43 69 46
 elis	Machine learning	922	6	22	39	672	89	154	42	112
 SciPy	Data wrangling	19 150	608	99	301	4 447	2 318	31	193	7
 NumPy	Data wrangling	17 911	641	136	390	7 215	2 766	28	132	11

Library Name	Type	Commits	Contributors	Releases	Watch	Star	Fork	Commits/ Contributors	Commits/ Releases	Star/ Contributors
pandas	Data wrangling	17 144	1 165	93	858	14 294	5 788	15	184	12
SM StatsModels Statistics in Python	Statistics	10 067	153	21	234	2 868	1 240	66	479	19
TensorFlow	Deep learning	33 339	1 469	58	7 968	99 664	62 952	23	575	68
PYTORCH	Deep learning	11 306	635	16	816	15 512	3 483	18	707	24
Keras	Deep learning	4 539	671	41	1 673	29 444	10 964	7	1111	44
dist-keras elephas spark-deep-learning	Distributed deep learning	1125 170 67	5 13 11	7 5 3	41 97 116	431 913 920	106 189 206	225 13 6	161 34 22	86 70 84
Natural Language ToolKit	NLP	13 041	236	24	467	6 405	1 804	55	543	27
spaCy	NLP	8 623	215	56	425	9 258	1 446	40	154	43
gensim	NLP	3 603	273	52	415	6 995	2 689	13	69	26
Scrapy	Data scraping	6 625	281	81	1 723	27 277	6 469	24	82	97

Python usage and datascience

- Data science involves extrapolating useful information from massive stores of statistics, registers, and data.
- Python fills this need by being a **general-purpose** programming language.
- It allows you to **create CSV output** for easy data reading in a spreadsheet.
- Python can do this because it is lightweight and efficient at executing code, but it is also multi-functional.
- Python can support **object-oriented, structured and functional** programming



scikit

use cases

Example AI challenges

Media



Create highlights of sports events

Compliance



Is my organization compliant with latest regulatory documents

Industrial



Guide me through fixing malfunctioning components

Visual Inspection



Find rust on electric towers, using drones

Customer Care



Bot that can guide a user through buying the right insurance policy

Marketing / Business



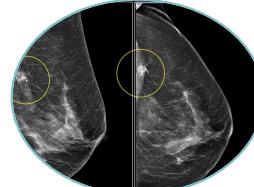
Summarize the strategic intent of a company based on recent news articles

IoT



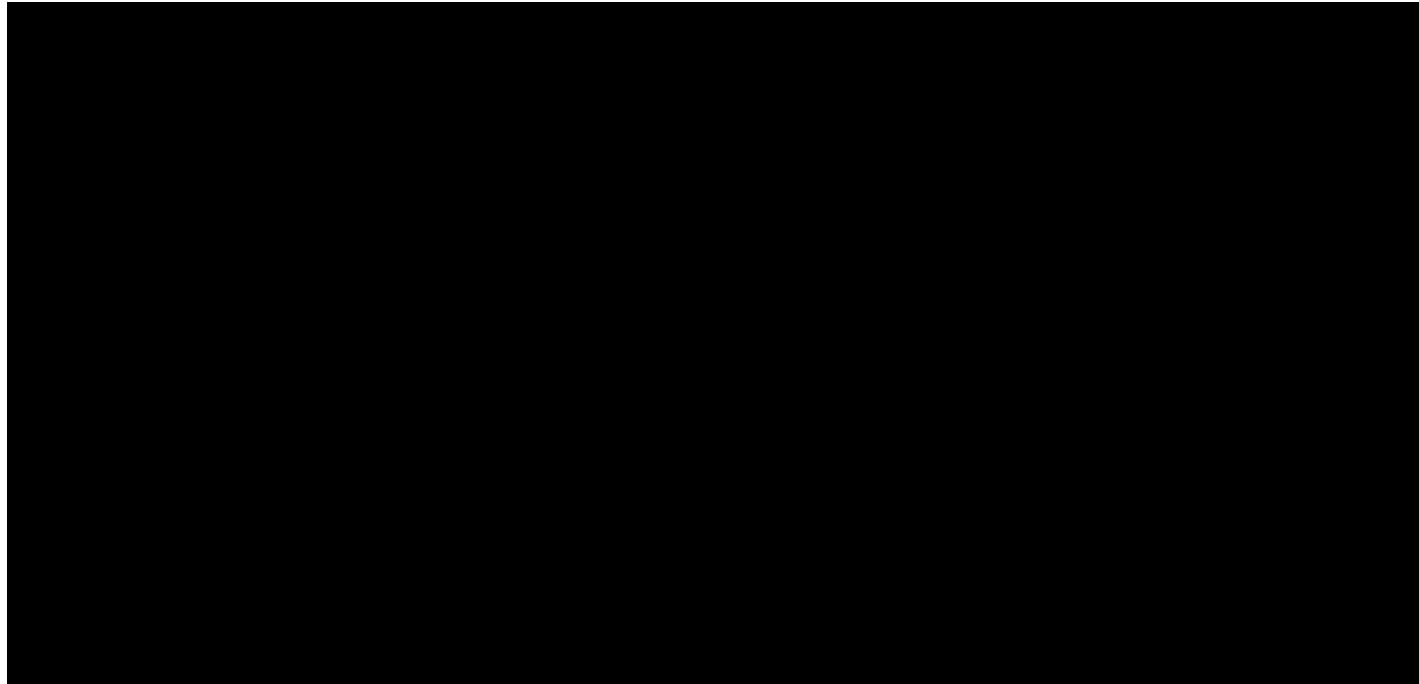
Predict yield of field based on images and sensor data

Healthcare



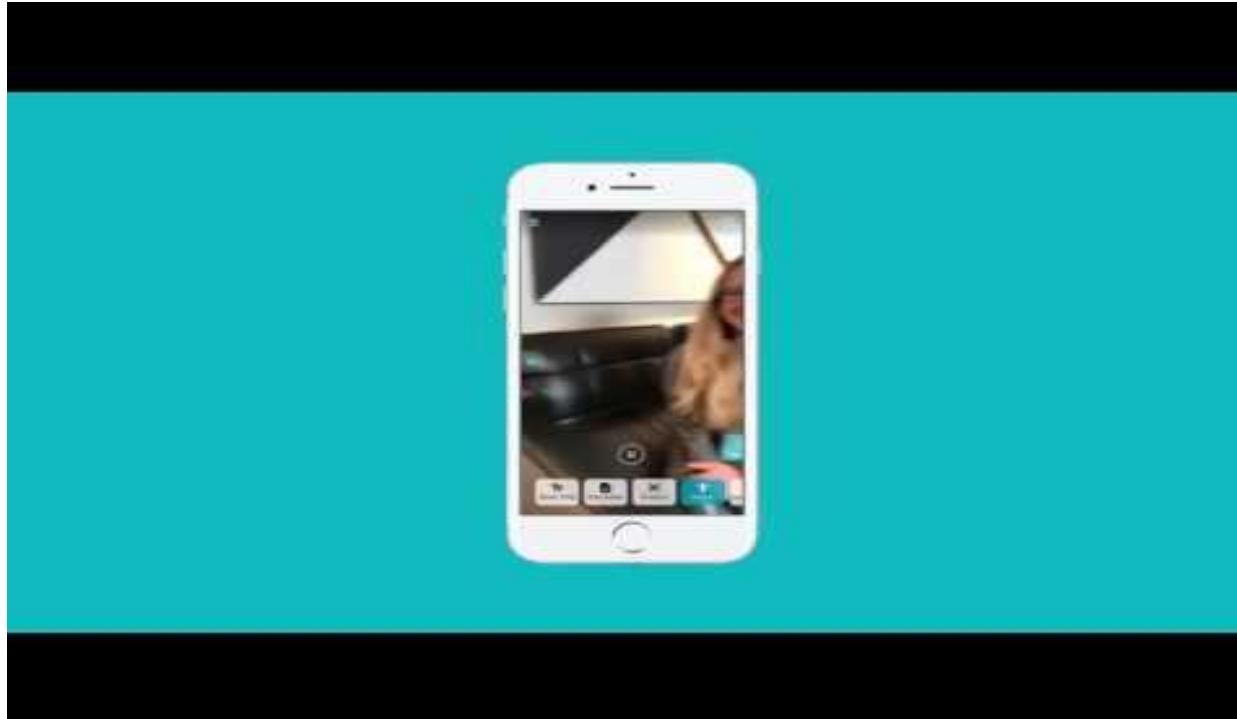
Improve the accuracy of breast cancer screening

Can we see with AI?



https://www.ted.com/talks/chieko_asakawa_how_new_technology_helps_blind_people_explore_the_world/discussion?refrer=playlist-557#t-4936

Seeing with AI



https://www.youtube.com/watch?v=bqeQByqf_f8

Microsoft: Seeing AI



Short Text
Speaks text as soon as it appears in front of the camera



Documents
Provides audio guidance to capture a printed page, and recognizes the text, along with its original formatting



Products
Gives audio beeps to help locate barcodes and then scans them to identify products



Person
Recognizes friends and describes people around you, including their emotions



Scene
An experimental feature to describe the scene around you



Currency
Identify currency bills when paying with cash



Light
Generate an audible tone corresponding to the brightness in your surroundings



Color
Describes the perceived color



Handwriting
Reads handwritten text

Capture new customer



The image shows a close-up of a man's face, partially blurred, smiling while holding a smartphone. The phone's screen displays the "imagin bank" logo in blue script. Overlaid on the image is promotional text in white and yellow.

**Cuenta imaginBank:
Sin comisiones.
Sin condiciones**

DESCÁRGATE LA APP

Will it be the future?



<https://www.youtube.com/watch?v=eHvyKWS5QzE>

Call center automatization



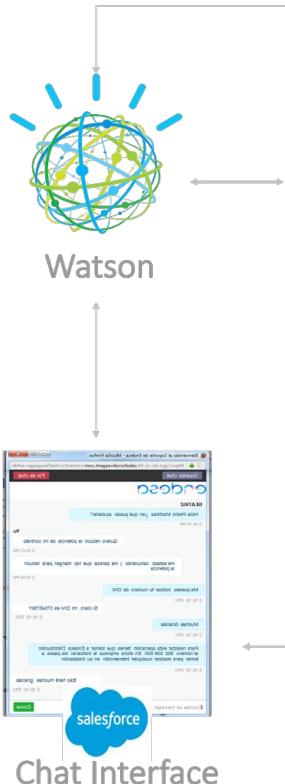
Better efficiency

LDA, Microsoft, Accenture

Watson & Salesforce



Chat interface



- Collect the start of all dialogues with customers, when it selected before starting the chat "Request for Information".
- Identify customer intent, through chat agent when it does not belong to the issues that have been trained.
- Customer maintains a contextualized dialogue.
- CRM information access through services available in this application, and interprets it in order to prepare the best response to the client.
- It provides customer response.
- Dialogue transferred to an agent when you do not find answers to sustain the conversation or new intentions appear.
- After the conversation with the customer, Watson Salesforce provide the necessary information so that it can proceed to closing.



CRM

Watson since 2015

CincoDías | lunes, 27 de julio de 2015

Inicio ▾ Mercados ▾ Empresas ▾ Economía ▾ Tecnología ▾ M...

ESTÁ PASANDO ▾ Crisis deuda europea | Grecia | Alexis Tsipras | IBEX 35 | Elecciones generales | BCE | P...

Mejorar el acceso al conocimiento por parte de sus empleados

CaixaBank trabaja con IBM para enseñar español al ordenador Watson

MARINAR IMÉNEZ | MADRID | 26-08-2014 13:09

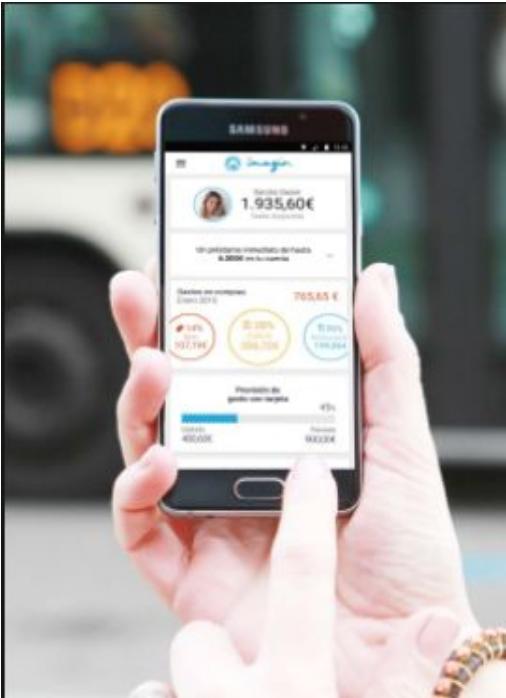
11 48 68 84 0

Temas relacionados: CaixaBank, Tecnología, Empresas, Ordenadores, Economía, Informática

Watson, el potente y mítico sistema informático de IBM que interacciona con los humanos de forma similar a como lo hacen las personas, está aprendiendo a hablar español. CaixaBank e...



ImaginBank: VIRTUAL ASSISTANT



Context & Solution

- Enhance the innovation with their customers
- Add simple questions about products
- Increase the digital Brand of Caixa
- Attract new customers – mainly millennials

Results

- Availability 24x7
- Immediate answers to questions
- Improve end user experience

<https://www.facebook.com/imaginBank/>

Watson support
taxpayers in meeting
their tax obligations

Goal : 60.000 tax
payers

What: Virtual
assistant

Benefits: Reduces
time and people
dedication



A screenshot of the official website of the Agencia Tributaria. The header includes the Spanish flag, the agency's name, and links for different user groups: Ciudadanos, Empresas y profesionales, Colaboradores, and Sede Electrónica. Below the header, there is a navigation menu for the 'Suministro Inmediato de Información del IVA' (SII). The main content area features sections for 'SERVICIOS DE AYUDA' (Identification, Calculations, Localizer, Virtual Assistant) and 'INFORMACIÓN' (Information about VAT registration through the electronic office). There are also links for direct access, feedback, and interest links.

ForceManager

Force Manager



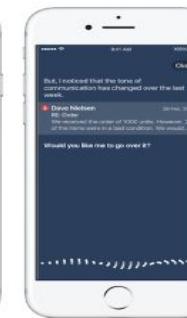
With
Watson™

ForceManager CRM

Easy, fast, reliable and smart.
Designed by salespeople, for salespeople



- +500 customers
- +30,000 users worldwide



4th Wave of CRM: i

The information rep
A competitive advai

Initial Bluemix subscription (ESA):
7,200€ / year

December 2018 Bluemix subscription:
141,000€ / year
(estimation covering half installed base)

CRM de IA

<https://vimeo.com/206389429/79d9771bee>

GRUPO SIRO



<https://www.youtube.com/watch?v=BXS65J9j0lg>

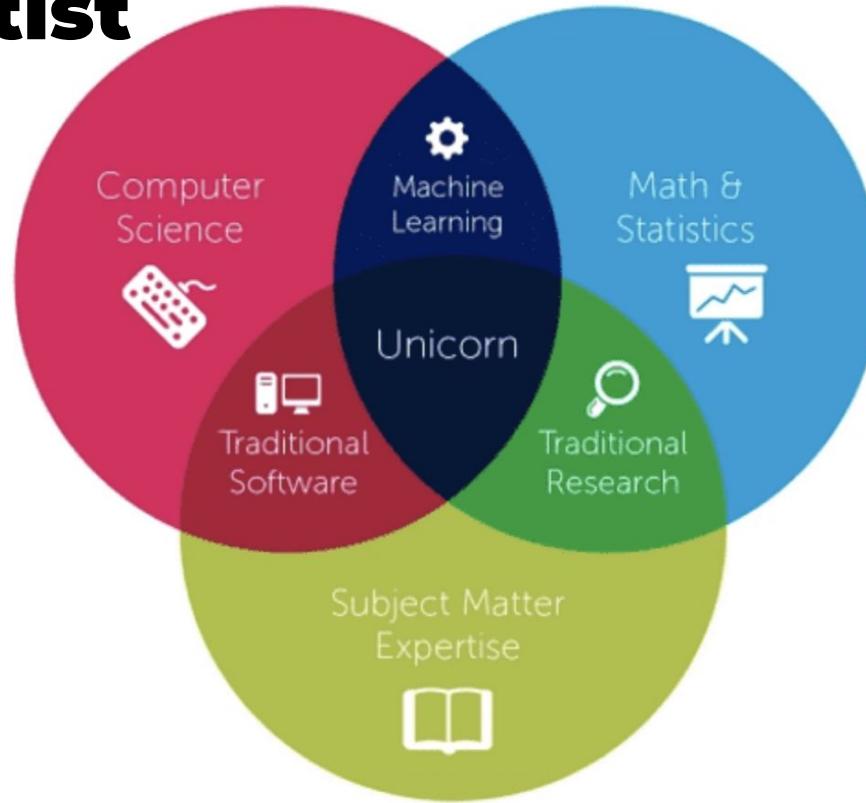
The role of Datascience

The role of datascience

Harvard Business Review called data science “the sexiest job of the 21st century,” and businesses from start-ups to well-established Fortune 500s are scrambling to fill these roles with the best and the brightest.

- **Identifying** the data-analytics problems that offer the greatest opportunities to the organization
- **Determining the correct** data sets and variables
- Collecting large sets of structured and unstructured data from disparate sources
- **Cleaning and validating** the data to ensure accuracy, completeness, and uniformity
- Devising and **applying models** and algorithms to mine the stores of big data
- Analyzing the **data to identify patterns** and trends
- Interpreting the **data to discover** solutions and opportunities (**business!**)
- **Communicating** findings to stakeholders using visualization and other means

Datascientist



Let's play!