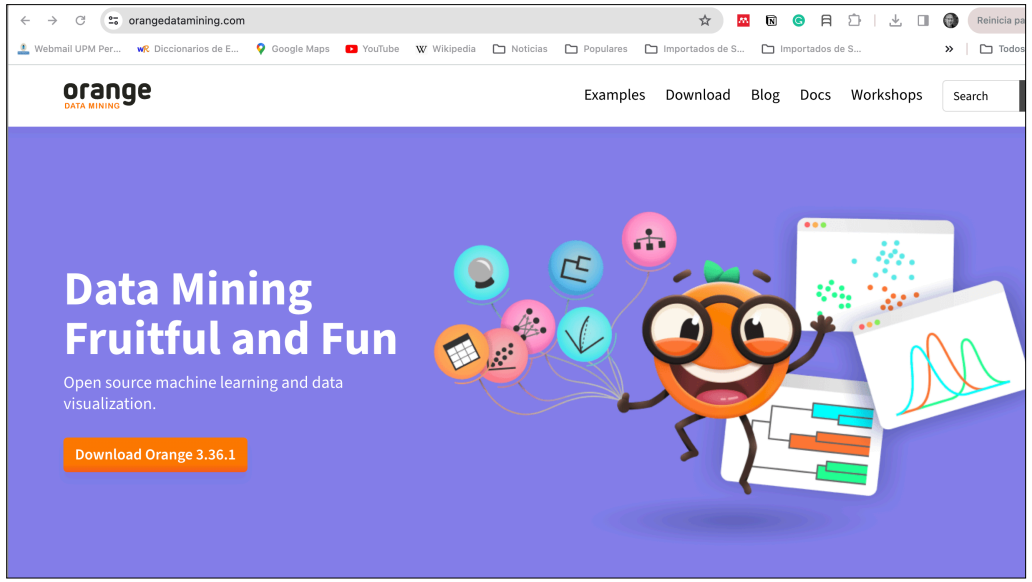
JAVIER GALEANO, UNIVERSIDAD EUROPEA

Machine Learning in Sports

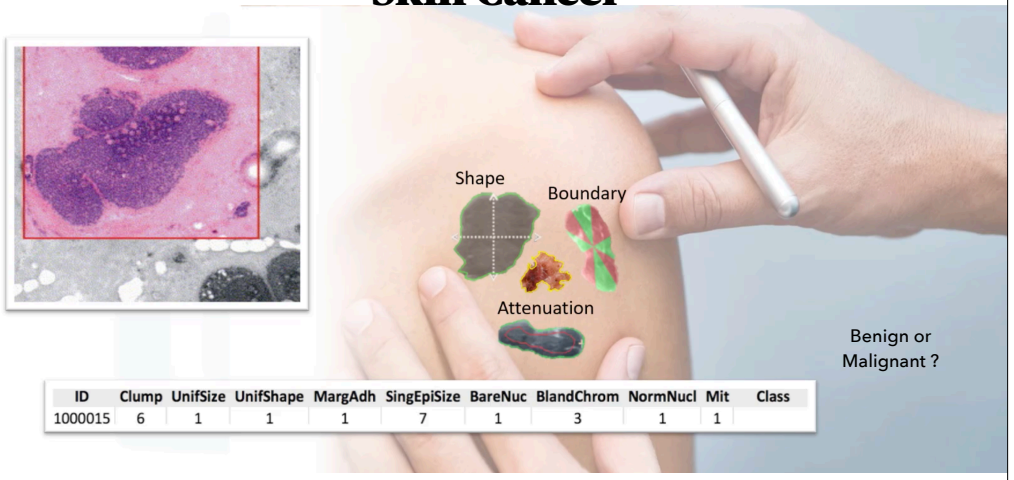
using Orange3

Escuela Universitaria
Real Madrid
Universidad Europea

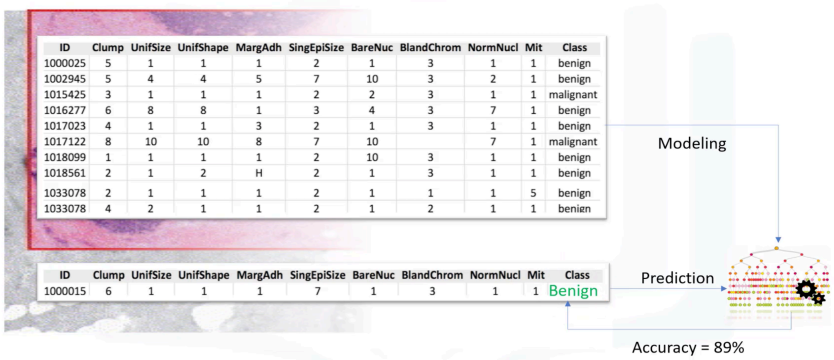


First example in Machine Learning

Skin Cancer



ID	Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
1000015	6	1	1	1	7	1	3	1	1	



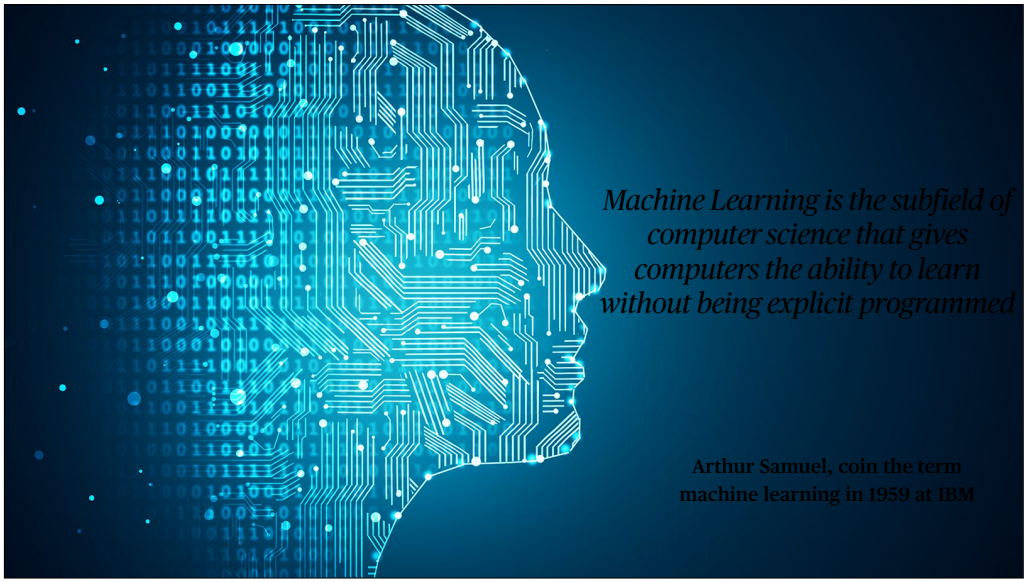
Examples in our daily life.

"Our lives are completely run by algorithms"

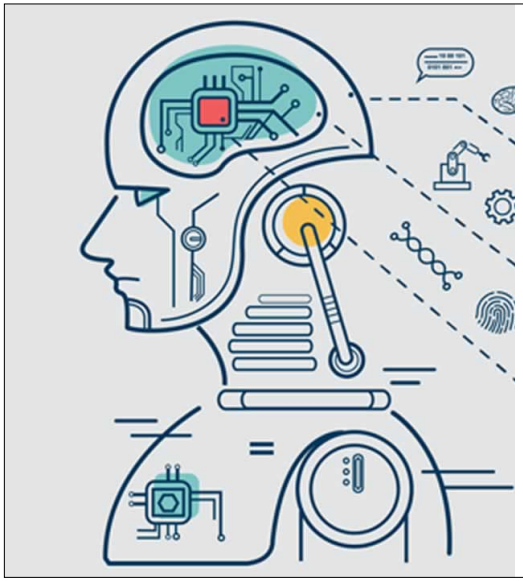
Marcus du Sautoy in The Creativity code

Examples of ML today

First definition of ML

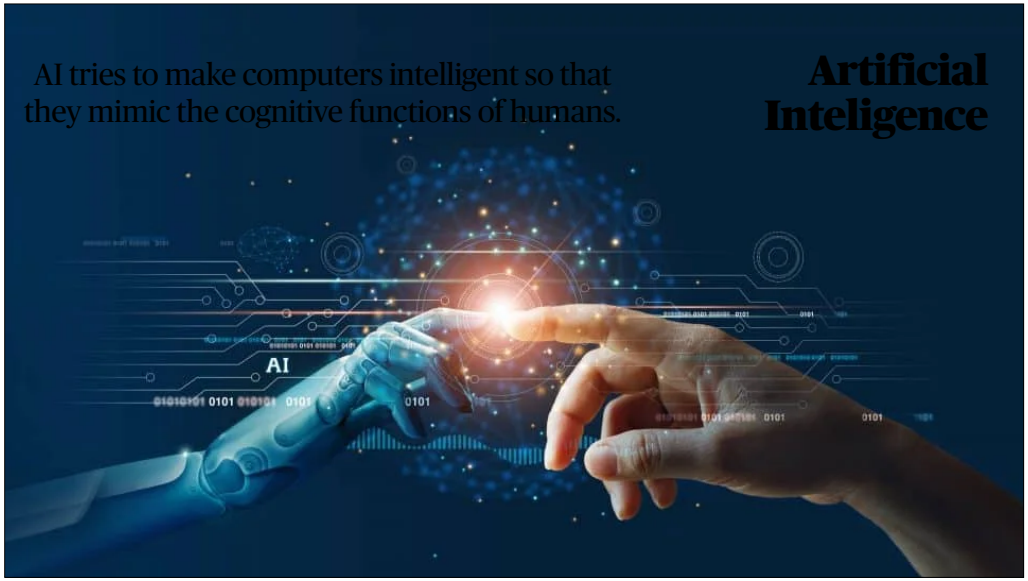


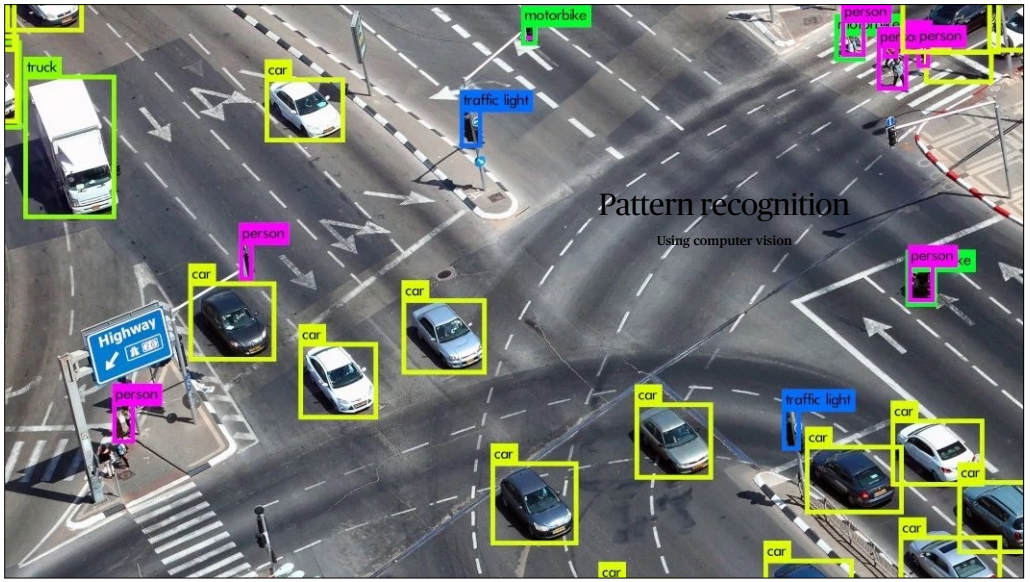
Some ideas

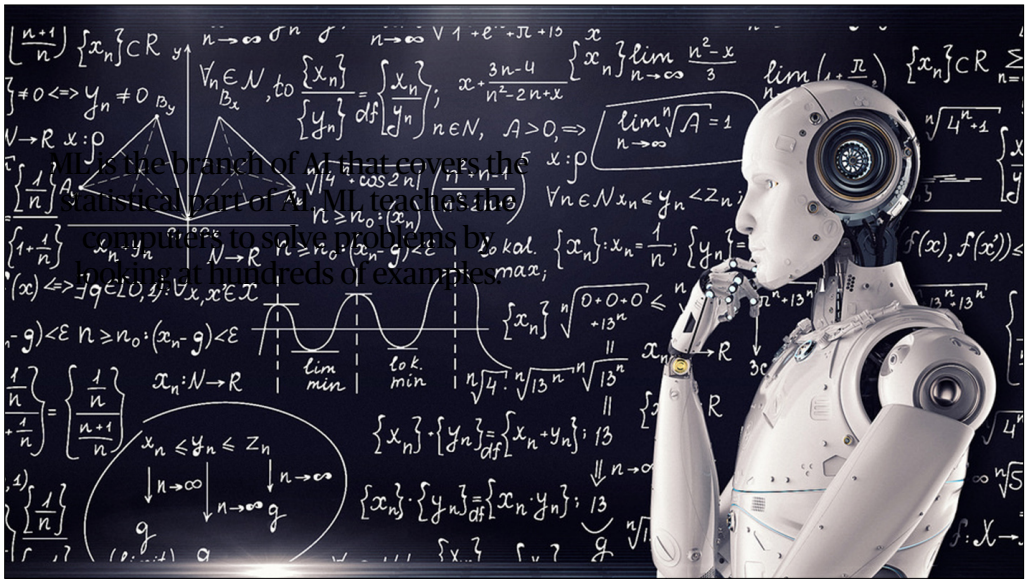


Differences between

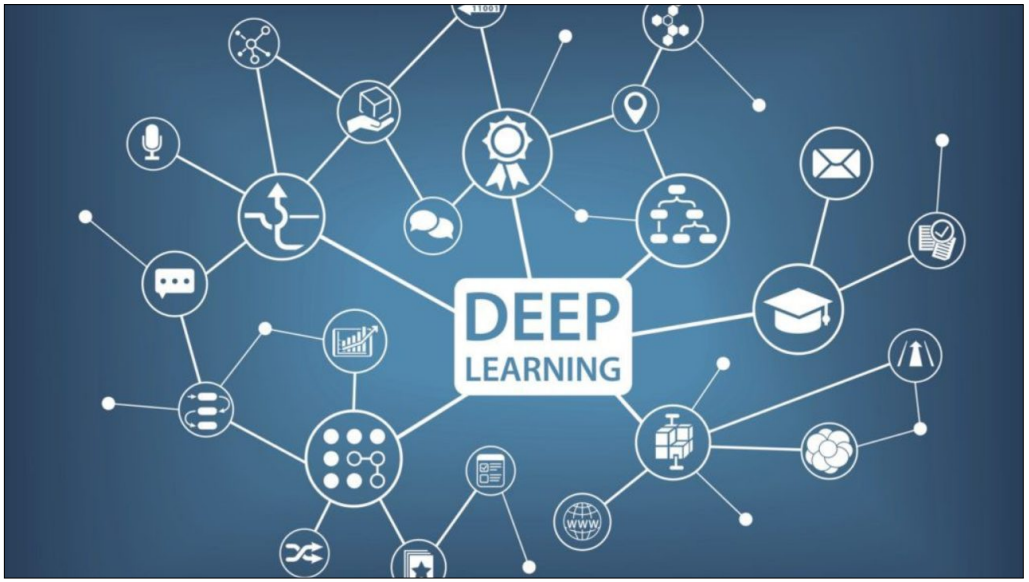
- Artificial Intelligence
- Machine Learning
- Deep Learning

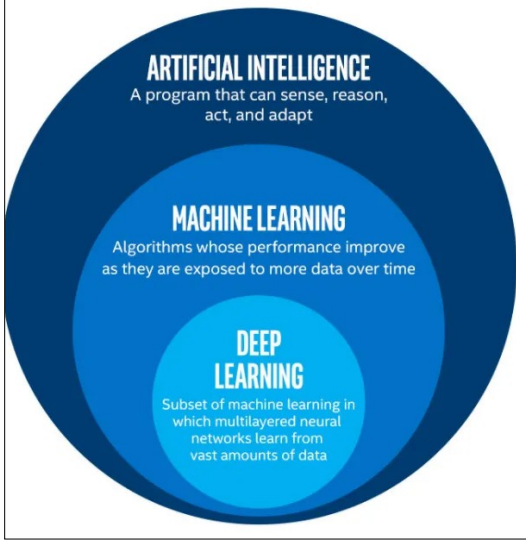






ML is the branch of AI that covers the statistical part of AI. ML teaches the computers to solve problems by looking at hundreds of examples.

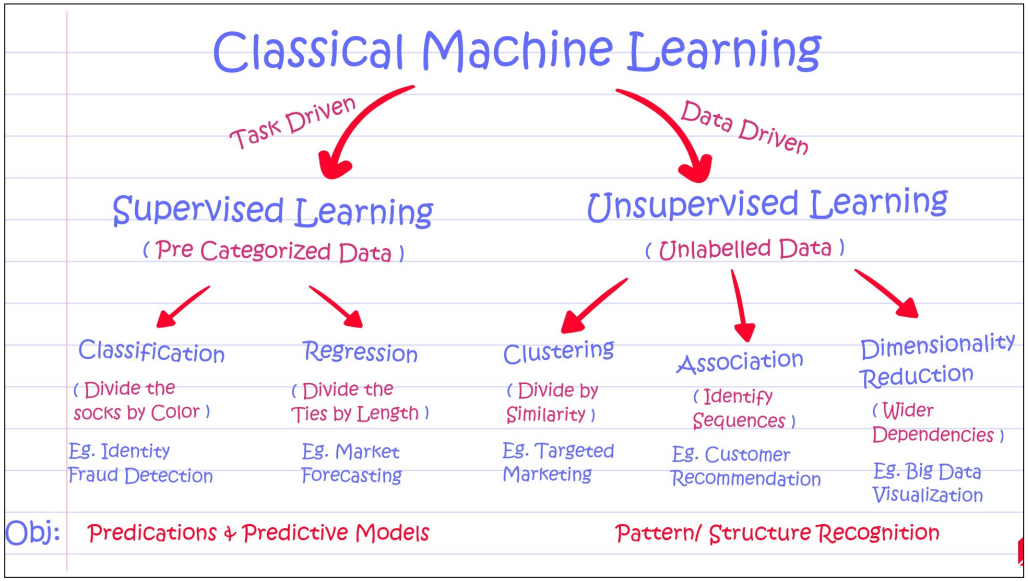




Mind map

- Mind map of:
  - artificial intelligence,
  - machine learning,
  - deep learning

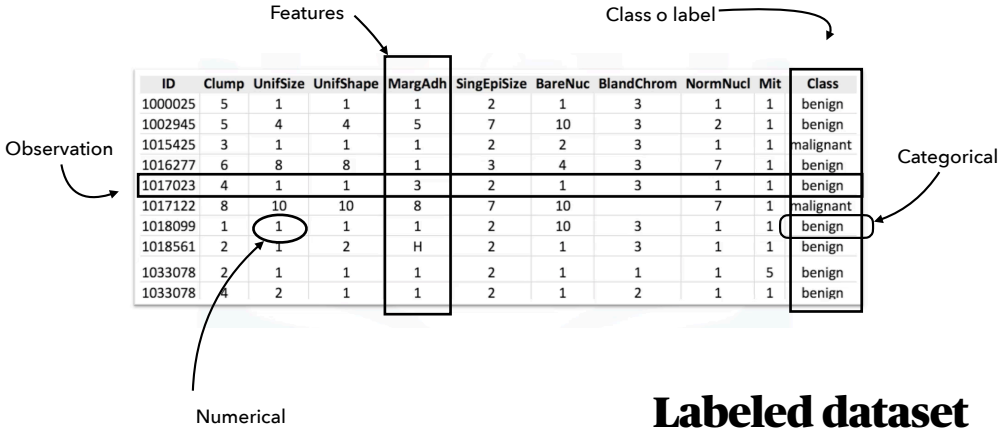
What problem in ML do I have?



ID	Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
1000025	5	1	1	1	2	1	3	1	1	benign
1002945	5	4	4	5	7	10	3	2	1	benign
1015425	3	1	1	1	2	2	3	1	1	malignant
1016277	6	8	8	1	3	4	3	7	1	benign
1017023	4	1	1	3	2	1	3	1	1	benign
1017122	8	10	10	8	7	10		7	1	malignant
1018099	1	1	1	1	2	10	3	1	1	benign
1018561	2	1	2	H	2	1	3	1	1	benign
1033078	2	1	1	1	2	1	1	1	5	benign
1033078	4	2	1	1	2	1	2	1	1	benign

Supervised model

In supervised models, we teach the model by training it with some data from labeled dataset



ID	Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
1000025	5	1	1	1	2	1	3	1	1	benign
1002945	5	4	4	5	7	10	3	2	1	benign
1015425	3	1	1	1	2	2	3	1	1	malignant
1016277	6	8	8	1	3	4	3	7	1	benign
1017023	4	1	1	3	2	1	3	1	1	benign
1017122	8	10	10	8	7	10		7	1	malignant
1018099	1	1	1	1	2	10	3	1	1	benign
1018561	2	1	2	H	2	1	3	1	1	benign
1033078	2	1	1	1	2	1	1	1	5	benign
1033078	4	2	1	1	2	1	2	1	1	benign

Labeled dataset

I am going to teach you

1 –> 2

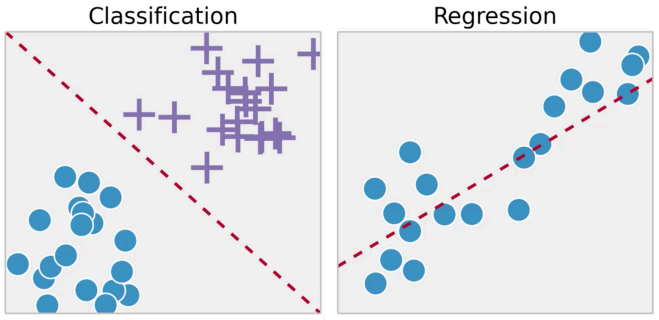
2 –> 4

5 –> 10

6 –> 12

10 –> ?

Types of supervised techniques

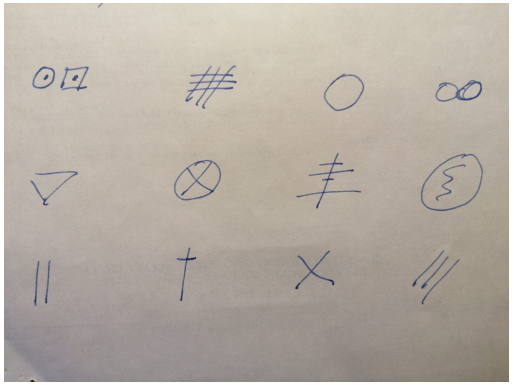


Customer Id	Age	Edu	Years Employed	Income	Card Debt	Other Debt	Address	DebtIncomeRatio
1	41	2	6	19	0.124	1.073	NBA001	6.3
2	47	1	26	100	4.582	8.218	NBA021	12.8
3	33	2	10	57	6.111	5.802	NBA013	20.9
4	29	2	4	19	0.681	0.516	NBA009	6.3
5	47	1	31	253	9.308	8.908	NBA008	7.2
6	40	1	23	81	0.998	7.831	NBA016	10.9
7	38	2	4	56	0.442	0.454	NBA013	1.6
8	42	3	0	64	0.279	3.945	NBA009	6.6
9	26	1	5	18	0.575	2.215	NBA006	15.5
10	47	3	23	115	0.653	3.947	NBA011	4
11	44	3	8	88	0.285	5.083	NBA010	6.1
12	34	2	9	40	0.374	0.266	NBA003	1.6

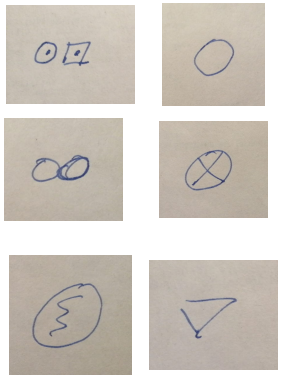
Unsupervised model

We do not supervise the model, but we let the model work on its own to discover information that may not be visible to the human eye.

I am going to teach you how to learn about an unsupervised dataset

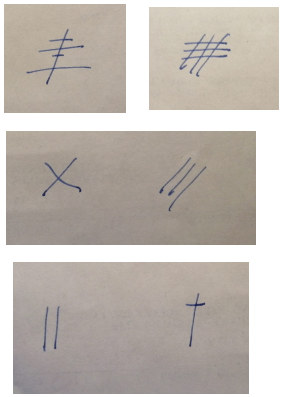


Polygonal language?

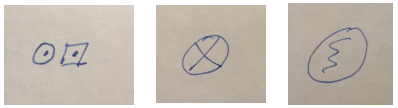


Two clusters

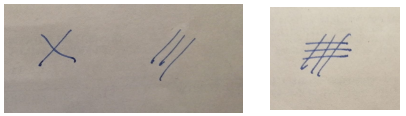
Linear language?



Fulled polygonal language?

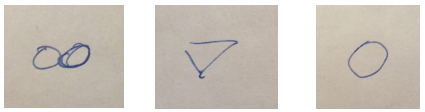


Diagonal Linear language?

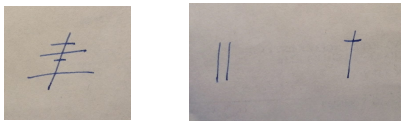


Four clusters?

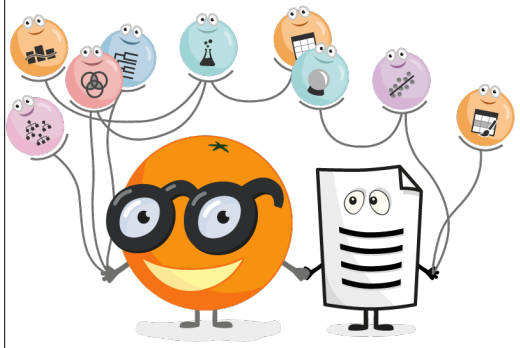
Empty polygonal language?



Straight Linear language?



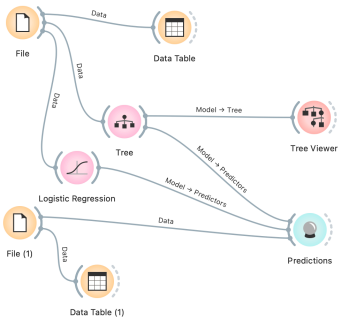
First steps using Orange3



ESCUELA DEL REAL MADRID

First program

- File –> Iris
- Data Table
- Scatter plot
- Selected data table



ESCUELA DEL REAL MADRID

First Predictions

- Download NBA Players dataset
- Tree –> Tree viewer
- Test file
- Prediction
- Logistic Regression

Classification problem in ML

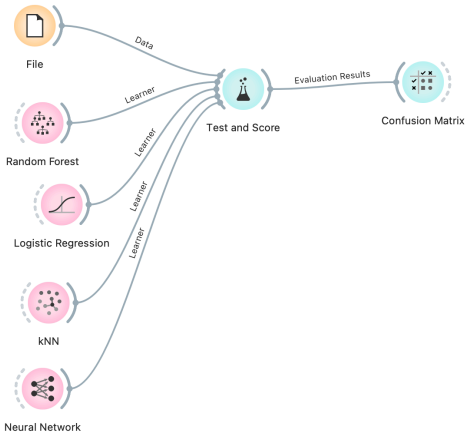
ID	Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
1000025	5	1	1	1	2	1	3	1	1	benign
1002945	5	4	4	5	7	10	3	2	1	benign
1015425	3	1	1	1	2	2	3	1	1	malignant
1016277	6	8	8	1	3	4	3	7	1	benign
1017023	4	1	1	3	2	1	3	1	1	benign
1017122	8	10	10	8	7	10		7	1	malignant
1018099	1	1	1	1	2	10	3	1	1	benign
1018561	2	1	2	H	2	1	3	1	1	benign
1033078	2	1	1	1	2	1	1	1	5	benign
1033078	4	2	1	1	2	1	2	1	1	benign

ESCUELA DEL REAL MADRID

Classification

In classification, the goal is to predict a *class label*, which is a choice from a predefined list of possibilities.

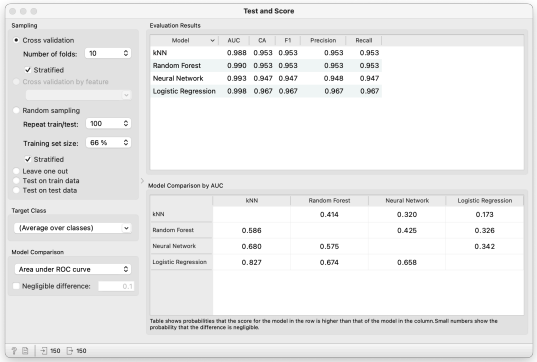
- Binary Classification
- Multiclass Classification



ESCUELA DEL REAL MADRID

Scoring

- K-cross validation
- Test & Score
- Confusion Matrix



Evaluation metrics in classification

- We are talking about some metrics:
- Jaccard index,
- Confusion matrix

Confusion Matrix

TP = True Positives
FN = False Negatives
FP = False Positives
TN = True Negatives

Precision $= \frac{TP}{TP + FP}$

Recall $= \frac{TP}{TP + FN}$

F1-score [0,1] $= \frac{2 * Precision * Recall}{Precision + Recall}$

Confusion Matrix

		Predicted			
		Iris-setosa	Iris-versicolor	Iris-virginica	Σ
Actual	Iris-setosa	50	0	0	50
	Iris-versicolor	0	47	3	50
	Iris-virginica	0	2	48	50
	Σ	50	49	51	150

Classification Accuracy is the proportion of correctly classified examples

$$CA = \frac{TP + TN}{Total}$$