Introduction to machine learning

RNN

Introducción al aprendizaje de la máquina

DeepMind

Self-driving cars

Cloud optimization

Healthcare
&
Medical images analysis

GAN

This is not a real cup of coffee

~~12/03/2020~~

~~30/04/2020~~

~~14/05/2020~~

May 20th 2020

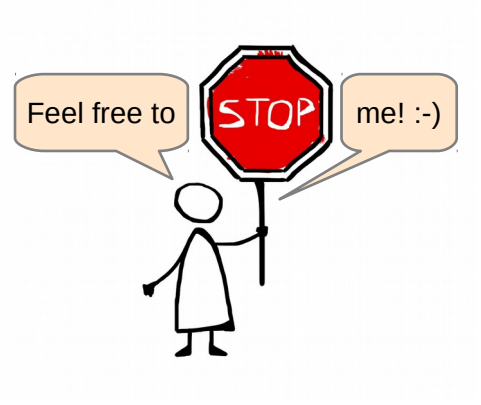
Luc Lesoil

Presentation

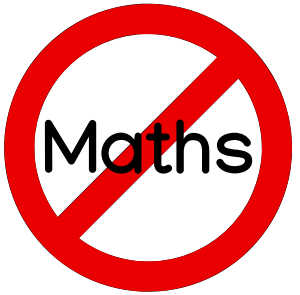
I. General
II. Supervised learning

Break – A video to introduce DiverSE

III. Unsupervised learning
IV. Reinforcement learning

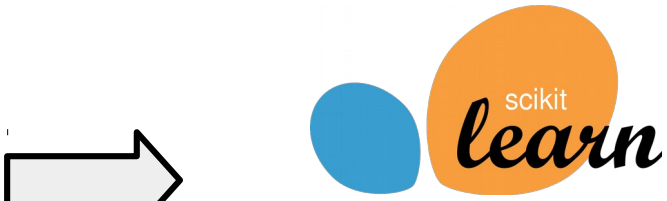


General introduction



- High-level presentation
- Non-exhaustive list
- Concrete cases

https://github.com/llesoil/ML_example



Supervised learning : basics

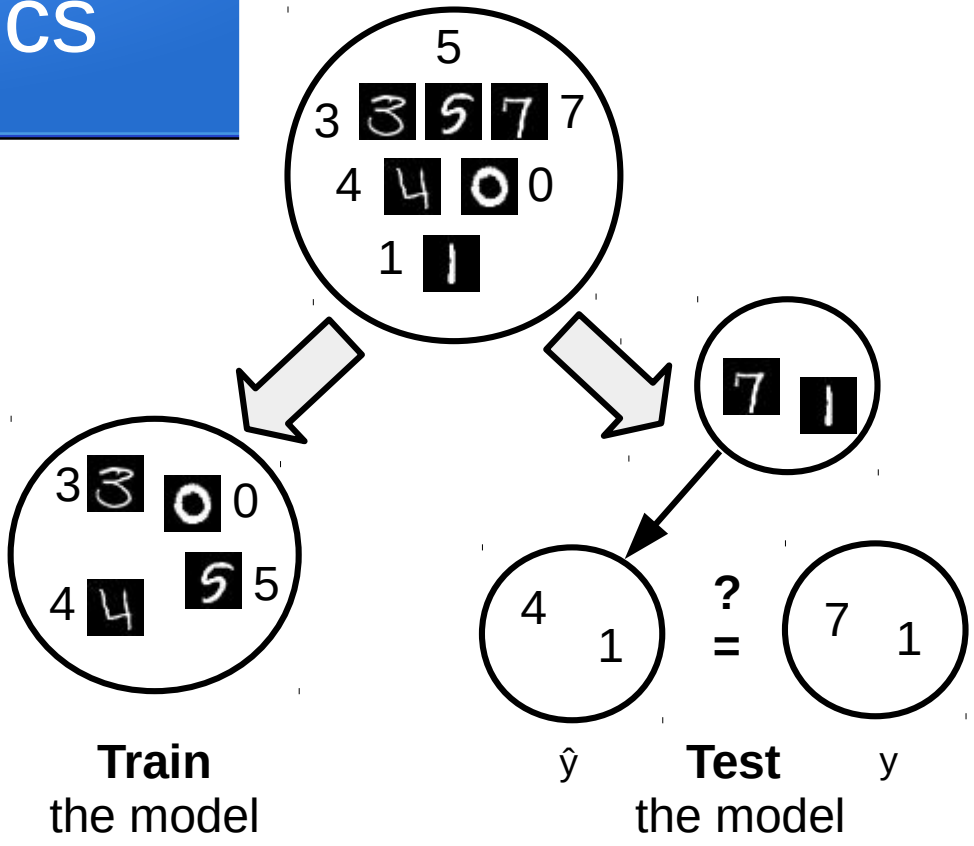
X : images of numbers

y : numbers

3 5 7
4 0
1

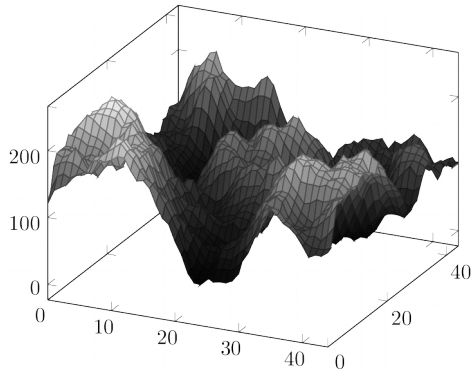
X : Explaining variables

y : Variable to predict, labels known



Supervised learning = Use X to predict y

Loss function



5

Today's Tux happiness : 3.2

2.5

Prediction 1 : 3

0

Prediction 2 : 0.5

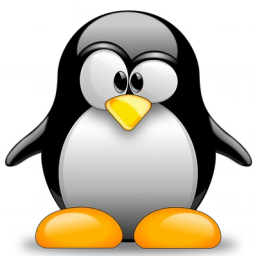
Compare quality of predictions?

Loss function

Minimize the loss function

=

better predictions



Tux

https://towardsdatascience.com/common-loss-functions-in-machine-learning-46af0ffc4d23

https://algorithmia.com/blog/introduction-to-loss-functions

Examples	
MAE	MAPE
Minkowski	Hinge
MSE	Cross-entropy

Type of

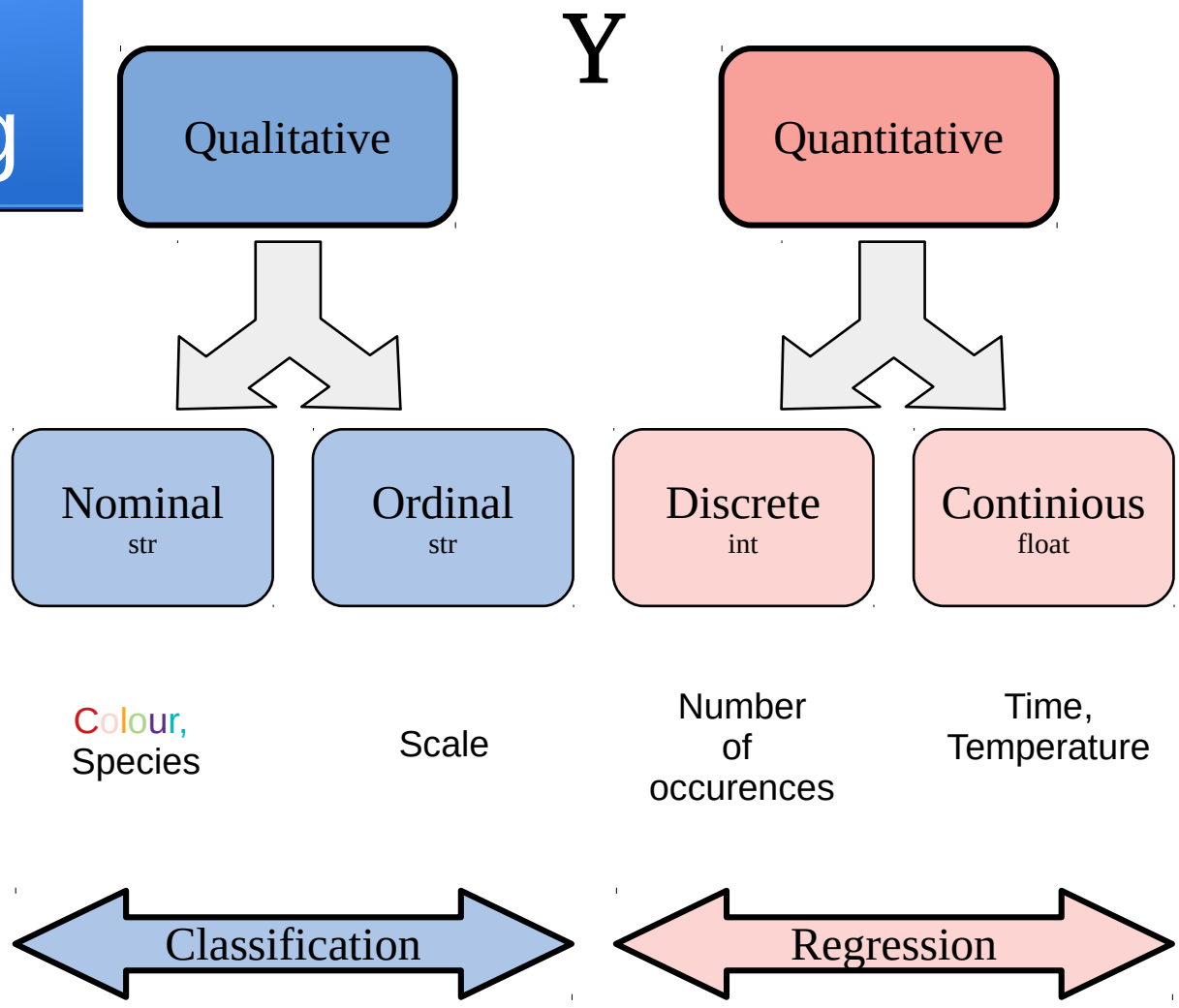
Supervised learning

Classification

→ Group or category

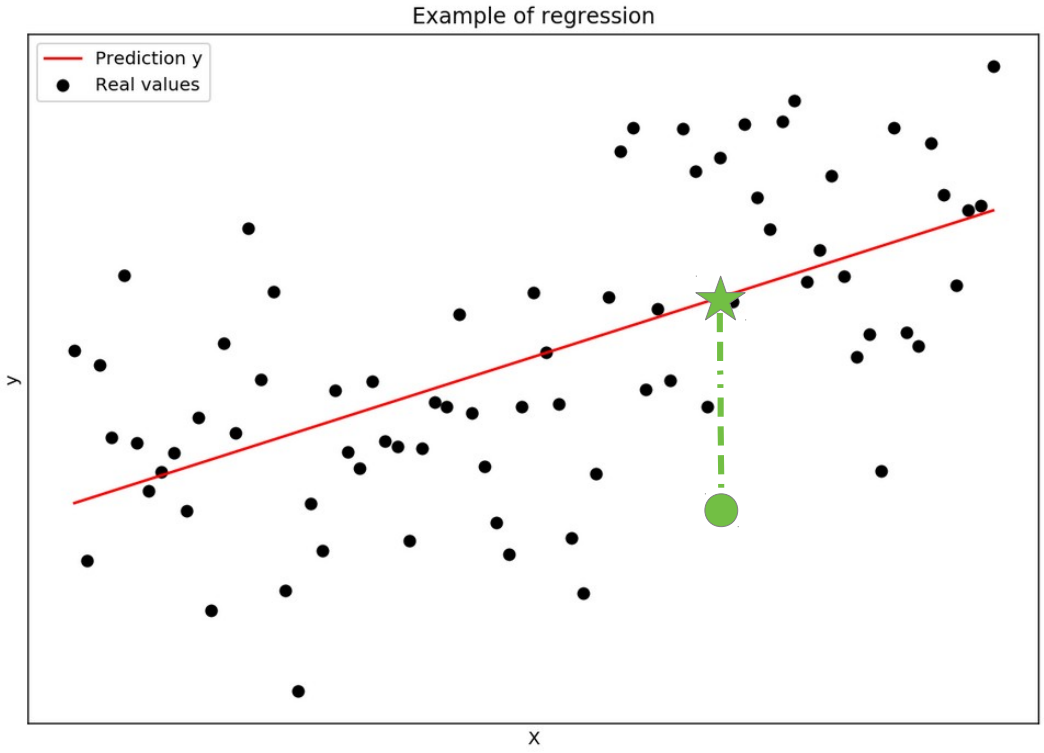
Regression

→ Value



Linear regression

- Simple
- ~~Complex dataset~~
- Linear relationship

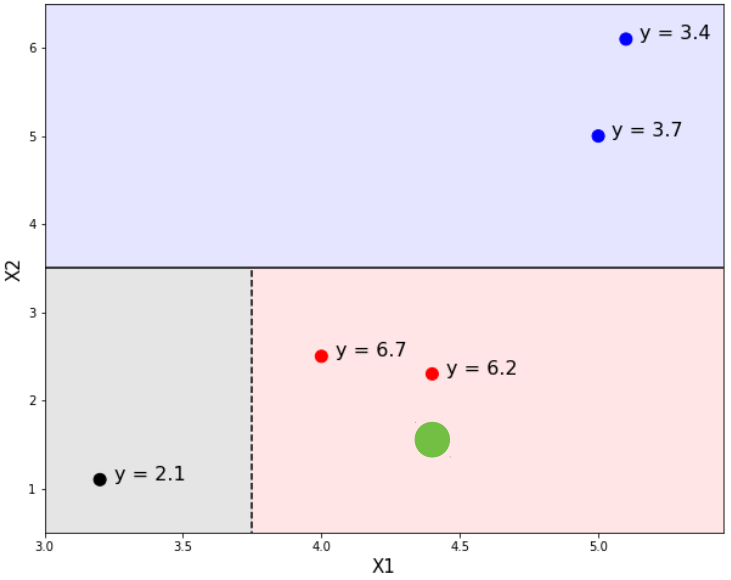


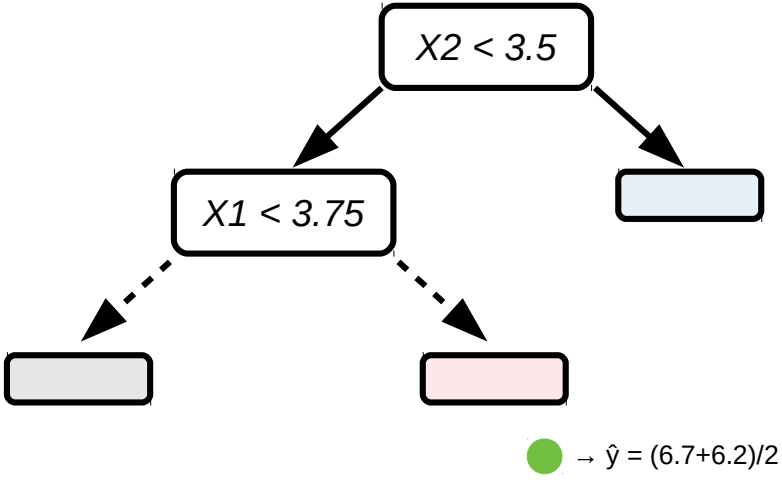
Fit the scatterplot with the red line

★ is the prediction of x = ●

Decision Tree (CART)

- Extract rules
- Simple to parameter
- Learning unit for many algorithms

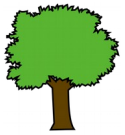




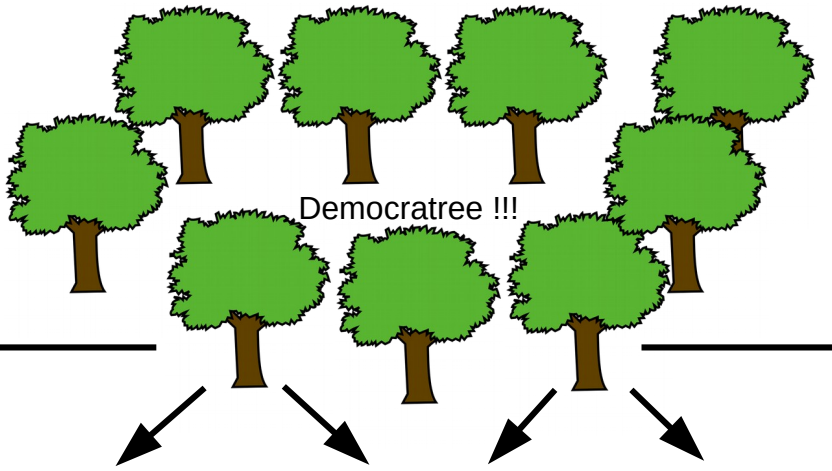
Classification Regression

Random Forest

- Bagging → robustness
- Metrics
- Good compromise



= 1 Decision Tree

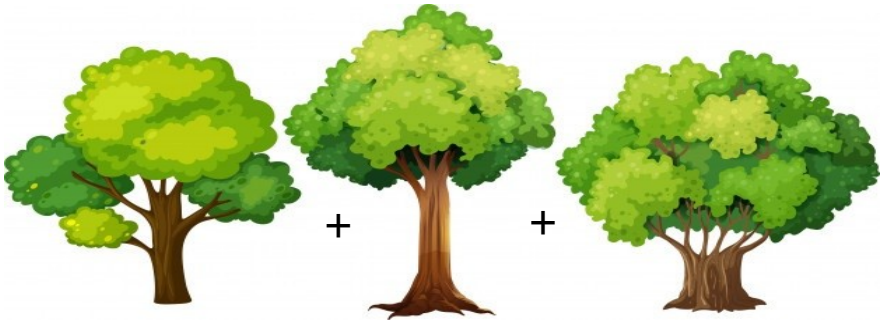


X_1	X_2	X_3
1	7	6
4	3	0
8	1	2
2	4	3

Classification Regression

Boosting tree

- Complex dataset
- Many hyperparameters
- XGBoost: the algorithm that wins every competition



Update the trees based on previous results

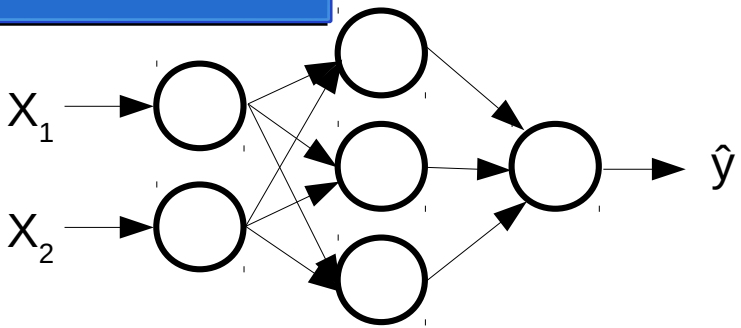
AdaBoost

XGBoost

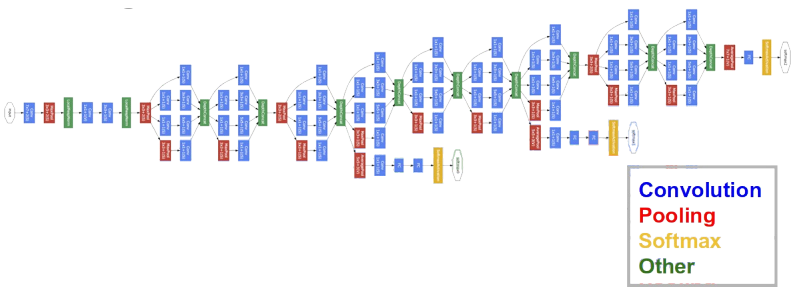
Classification Regression

Neural networks

- ~~Simple dataset~~
- Many hyperparameters
- Black box



Feedforward neural network

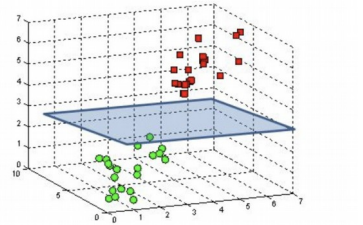


Classification Regression

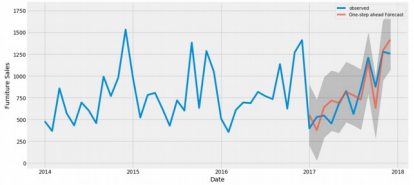
Others

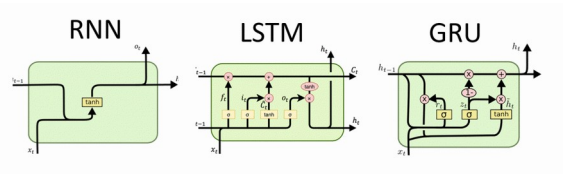
- Quantile/Polynomial/Piecewise regression, Ridge, ElasticNet, LASSO to select explaining variables

- Support Vector Machine : SVC or SVR



- Time series predictions : (S)AR(I)MA, RNN





Break - Video

- Demos
- Photos
- Some ideas?

Unsupervised learning

- Clustering
- Association
- Anomaly detection

Supervised Learning

- Step: 1

X → y

Teacher: Does picture 1 show a car or a flower?
Learner: A flower.
Teacher: No, it's a car.

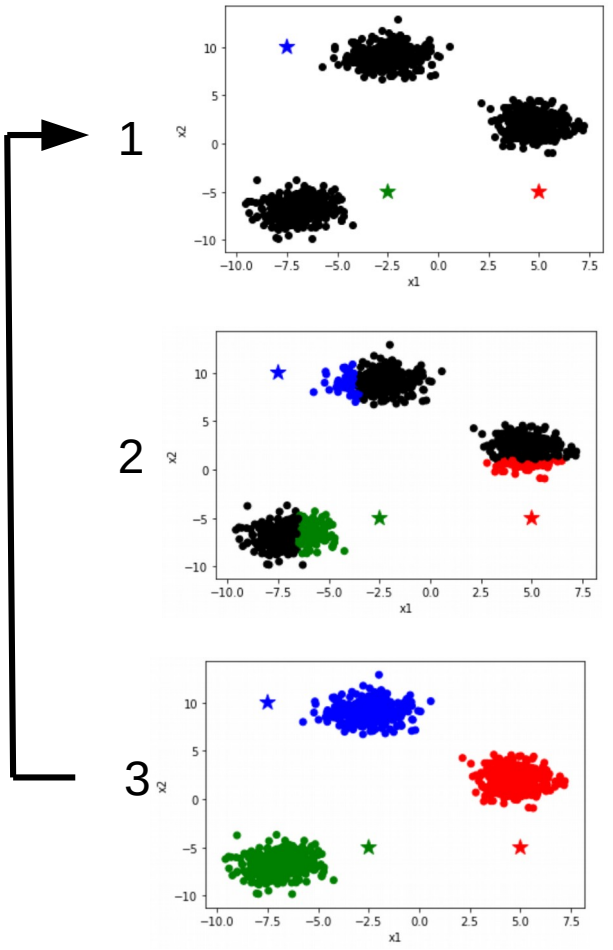
Step: 2
Teacher: Does picture 2 show a car or a flower?
Learner: A car.
Teacher: Yes, it's a car.

Step: 3 ....

	X	y
Supervised	✓	✓
Unsupervised	✓	✗

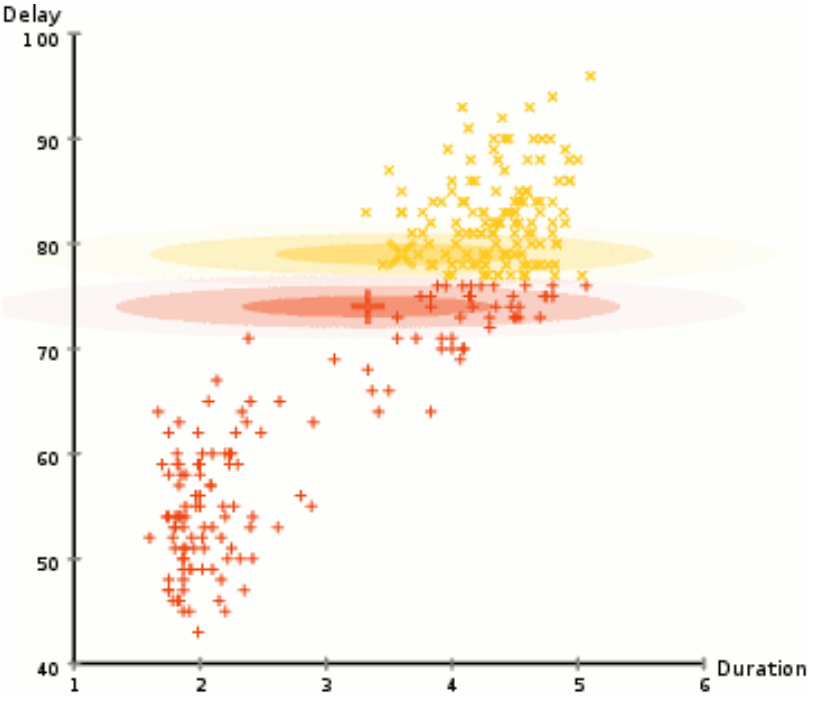
Kmeans

- Simple clustering
- Fast
- Few parameters



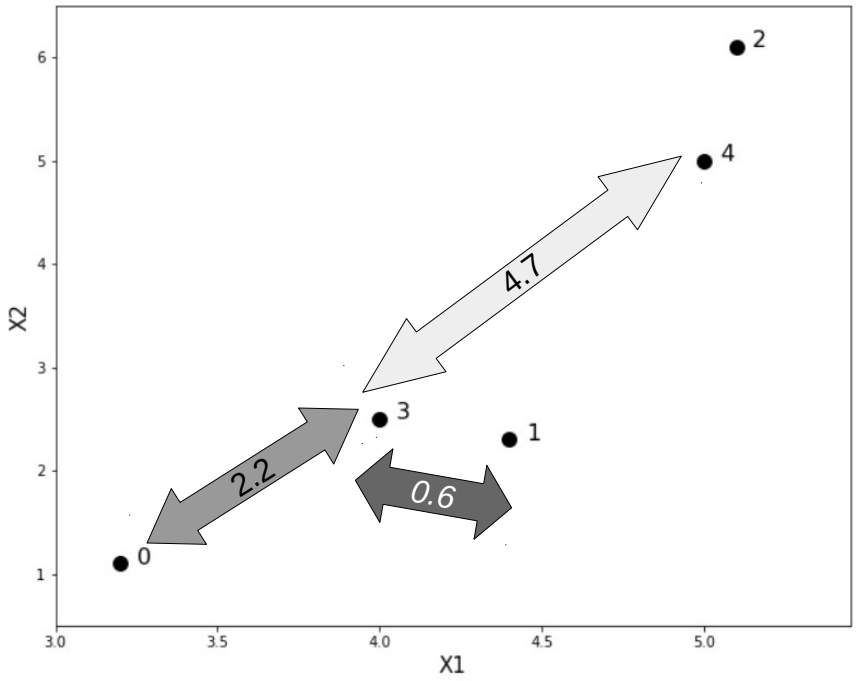
Gaussian Mixture Model

- Gaussian distribution
- Estimation of K
- Scale well - fast



K-Nearest neighbors

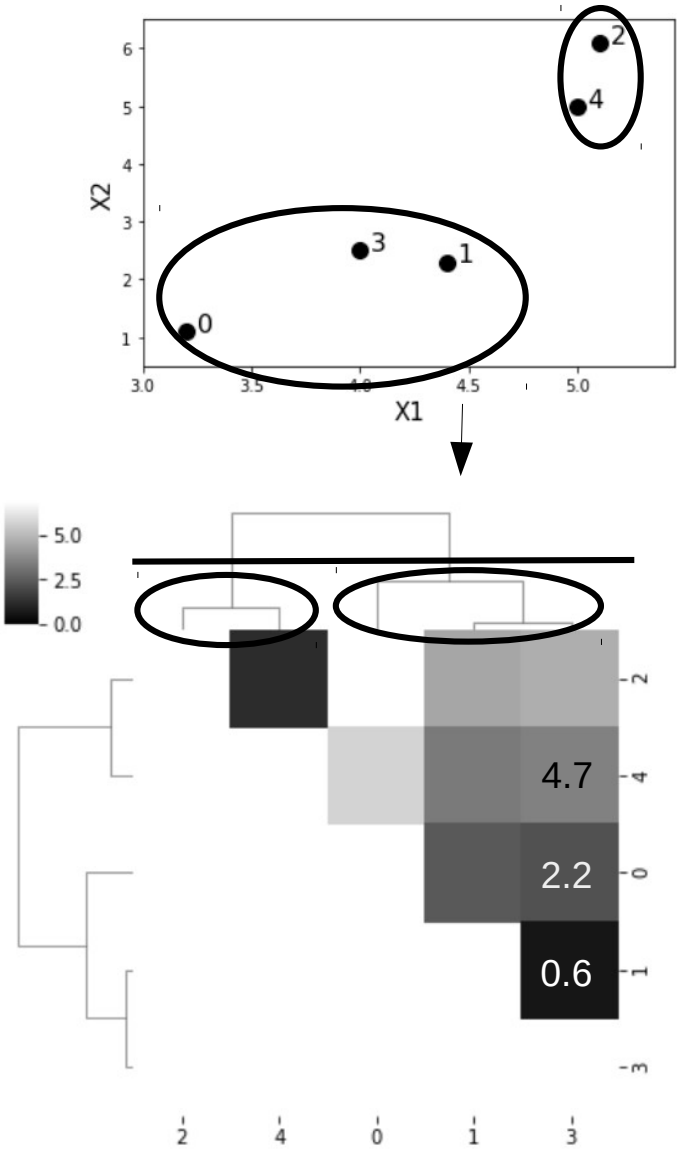
- Used in recommendation systems
- Supervised
- “You are the average of the five people you spend the most time with”



1 is the nearest neighbor of 3
0 is the second nearest

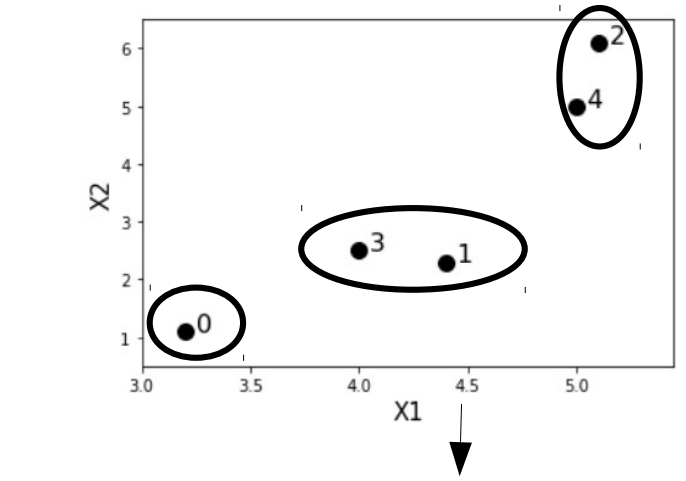
Hierarchical clustering

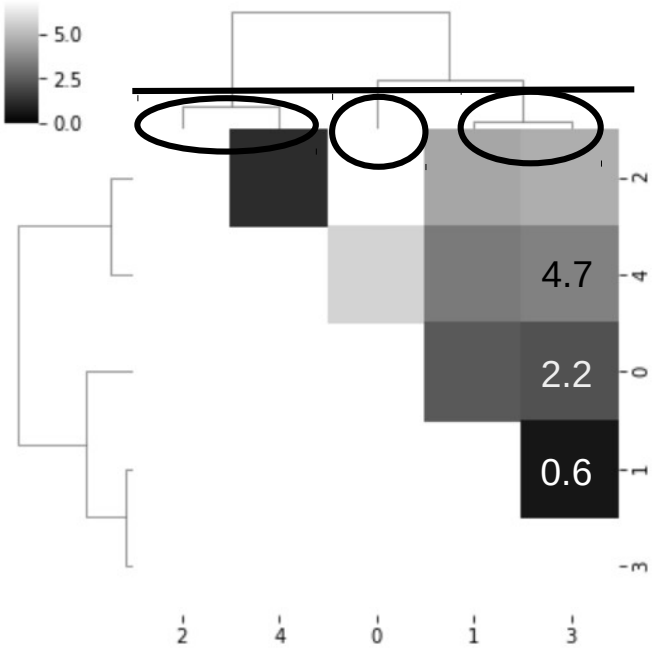
- Quadratic O(n²)
- Not designed for big dataset
- Full description of relationships



Hierarchical clustering (2)

- Quadratic O(n²)
- Not designed for big dataset
- Full description of relationships



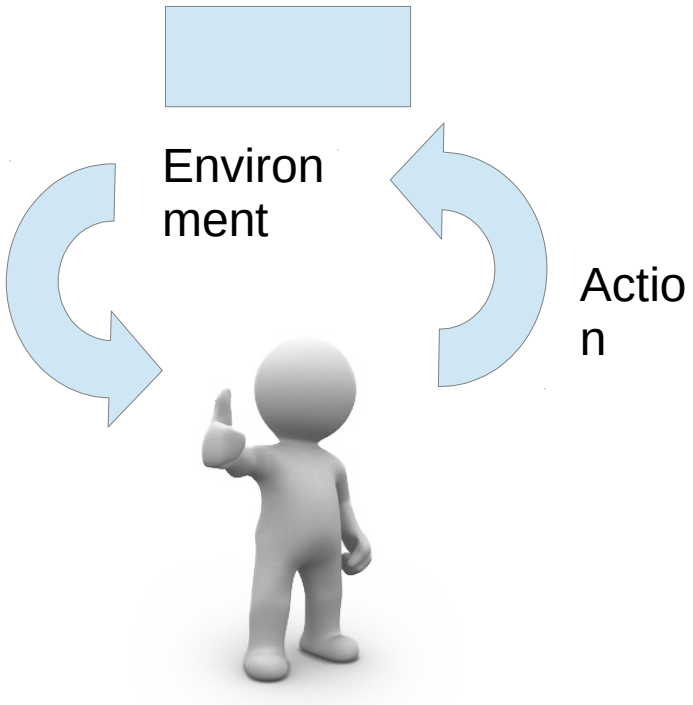


Others

- Hidden Markov Model, Bayesian classifier/networks
- Factorial analysis, other transforms (Fourier)
- Semi-supervised

Reinforcement Learning

Reinforcement learning
What's behind



Notions

- State S
- Action A
- Reward R
- Policy Pi
-
-

Reinforcement learning

Reinforcement learning

References

- Mnist dataset
-
-
- https://www.slideshare.net/cprakash2011/reinforcement-learning-40052403/5
- https://brilliant.org/wiki/gaussian-mixture-model/
- MARIQ : https://www.youtube.com/watch?v=CacRZmjDIr4