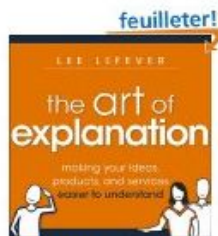


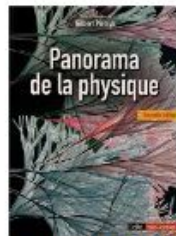
# Machine Learning Express



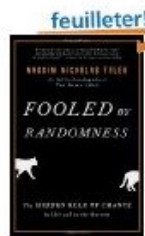
**The Art of Explanation:  
Making Your...  
Lee LeFever**  
Broché  
**EUR 21,83**



**Cro-Magnon toi-même ! :  
Petit guide...  
Michel Raymond**  
Poche  
★★★★☆ (10)  
**EUR 7,70**



**Panorama de la physique**  
Gilbert Pietryk  
Broché  
**EUR 27,08**



**Fooled by Randomness:  
The Hidden Role...  
Nassim Nicholas Taleb**  
Relié  
★★★★☆ (7)  
**EUR 17,72**



**A person riding a  
motorcycle on a dirt road.**



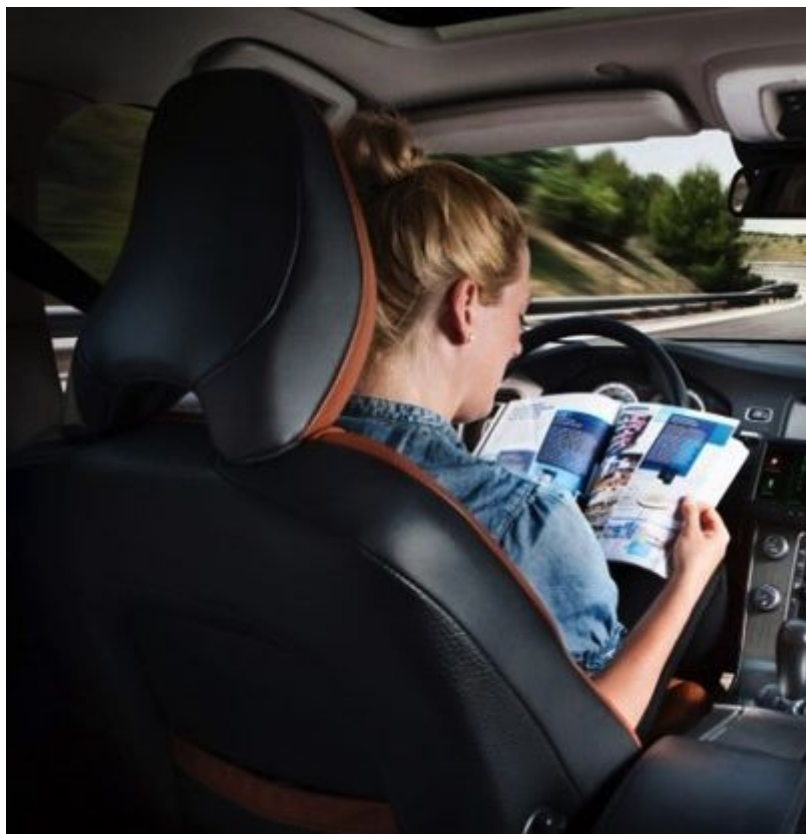
**A group of young people  
playing a game of frisbee.**

**Presented clip**



**Clip reconstructed  
from brain activity**





# Machine Learning

---

- ❑ Matérialisation concrète des travaux en Intelligence Artificielle
- ❑ Apprendre aux machines à réaliser des tâches sans les programmer explicitement
- ❑ Apprendre à partir des données
- ❑ Généraliser à partir de cas particuliers

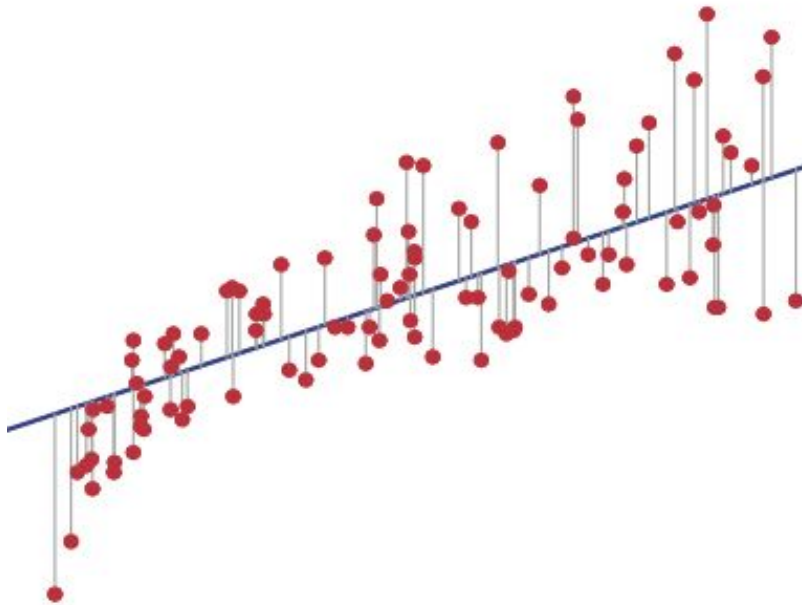
# Statistiques VS Machine Learning (version courte)

---

- Statistiques cherchent à expliquer comment ont été générées les données, ainsi que la dynamique sous-jacente
- ML s'intéresse plus au pouvoir prédictif et à l'exploitation des données

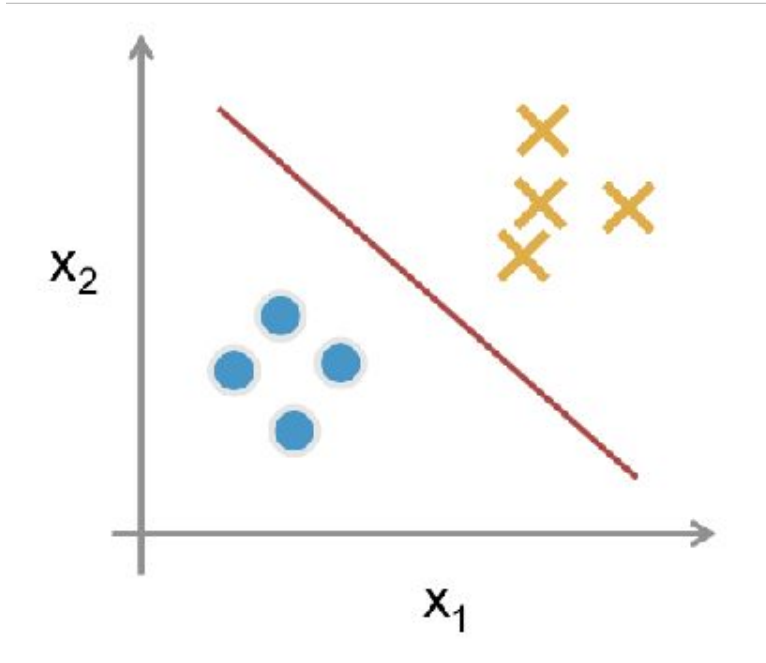
# Exemple régression linéaire

---

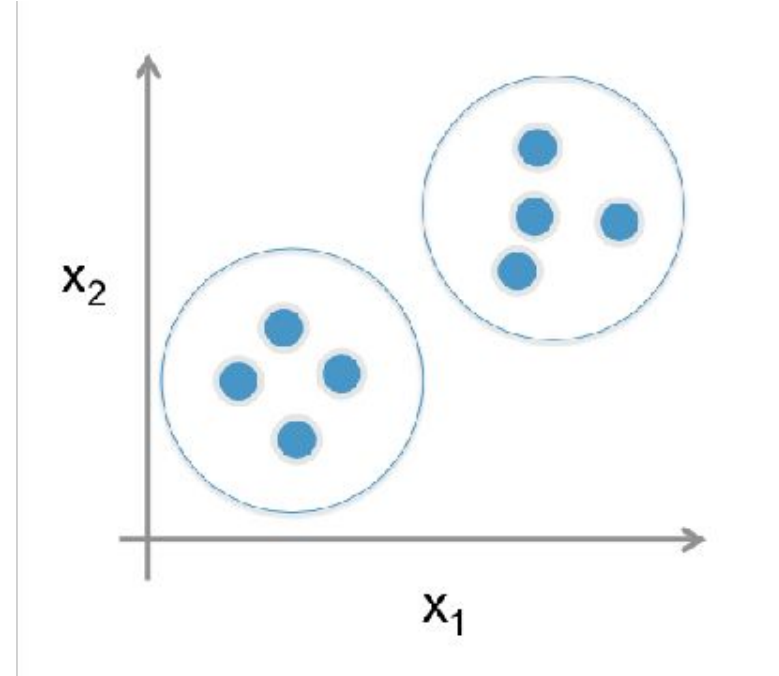


- Statistiques cherchent à expliquer la tendance générale des données
- ML souhaite, connaissant  $x$  d'une nouvelle variable, prédire le  $y$  correspondant

# Apprentissage supervisé / non supervisé

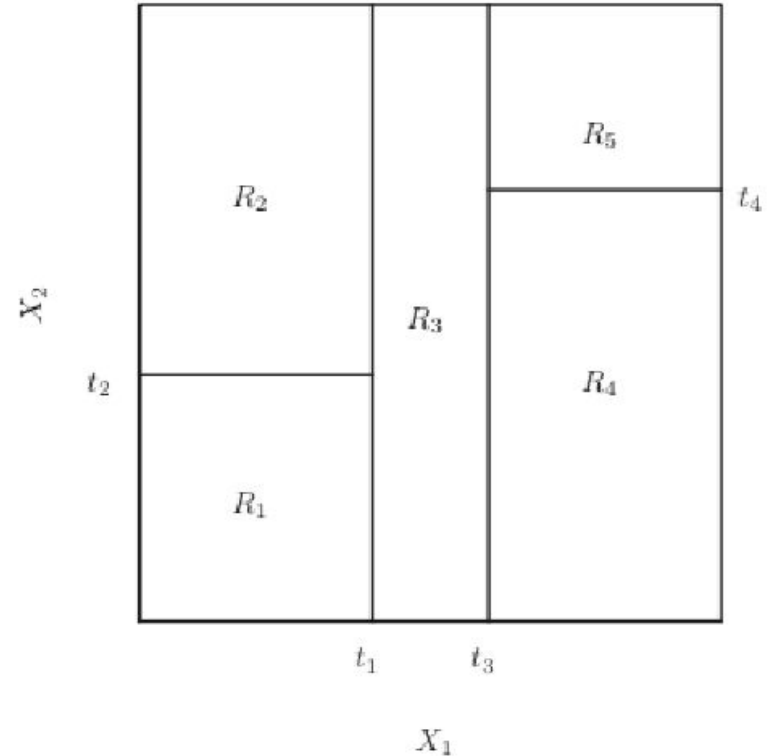
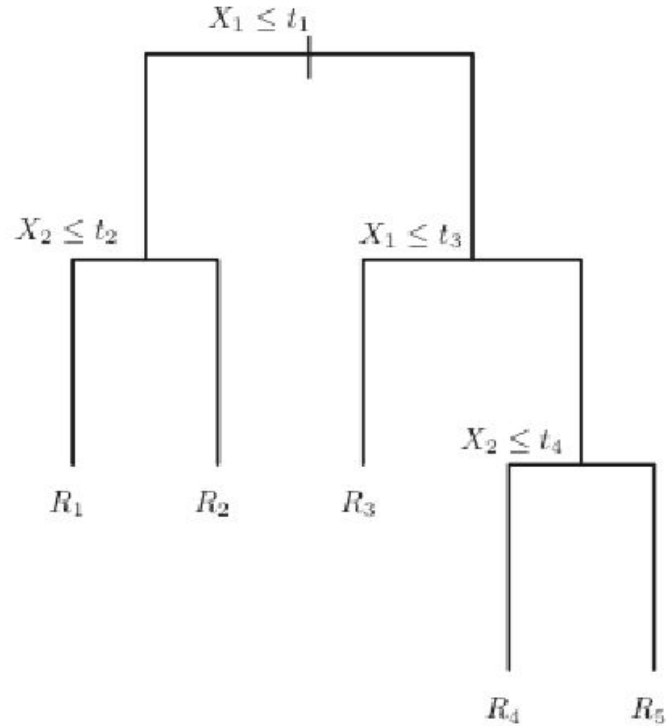


Régression / Classification



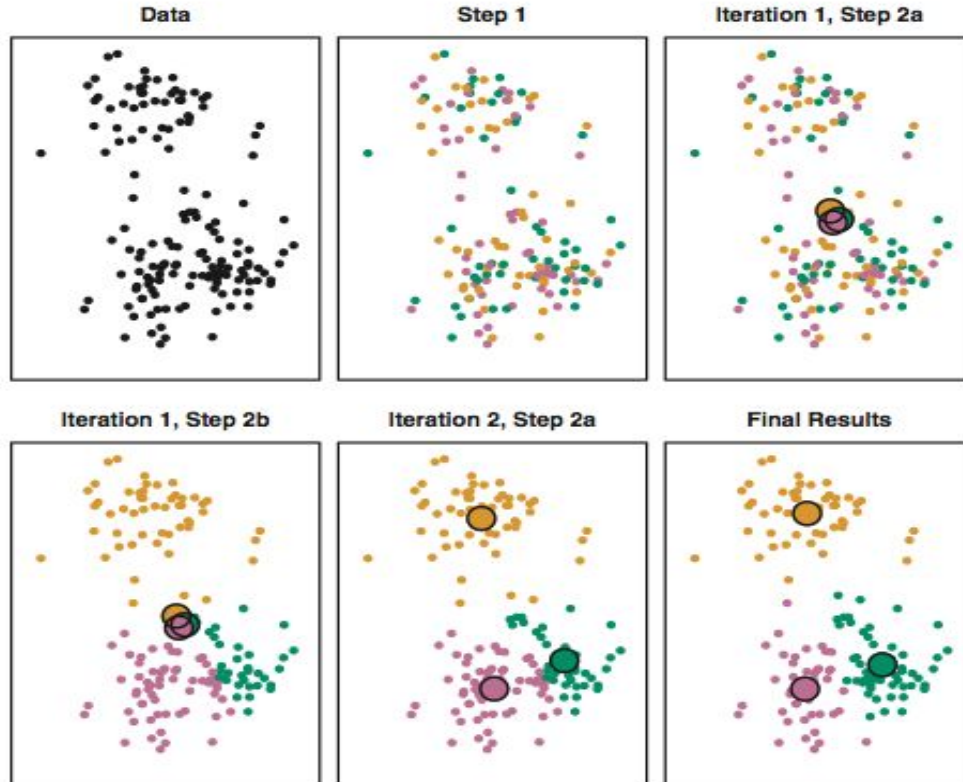
Clusterisation

# Classification : Arbres de décision (supervisé)

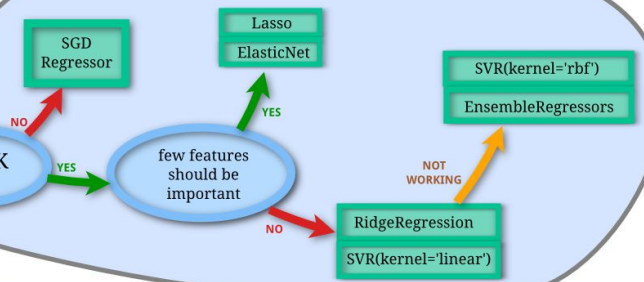




# Clusterisation (non-supervisé)

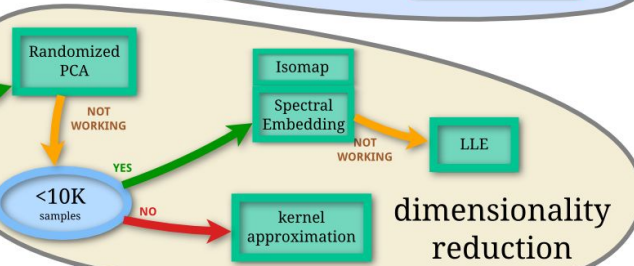


## classification

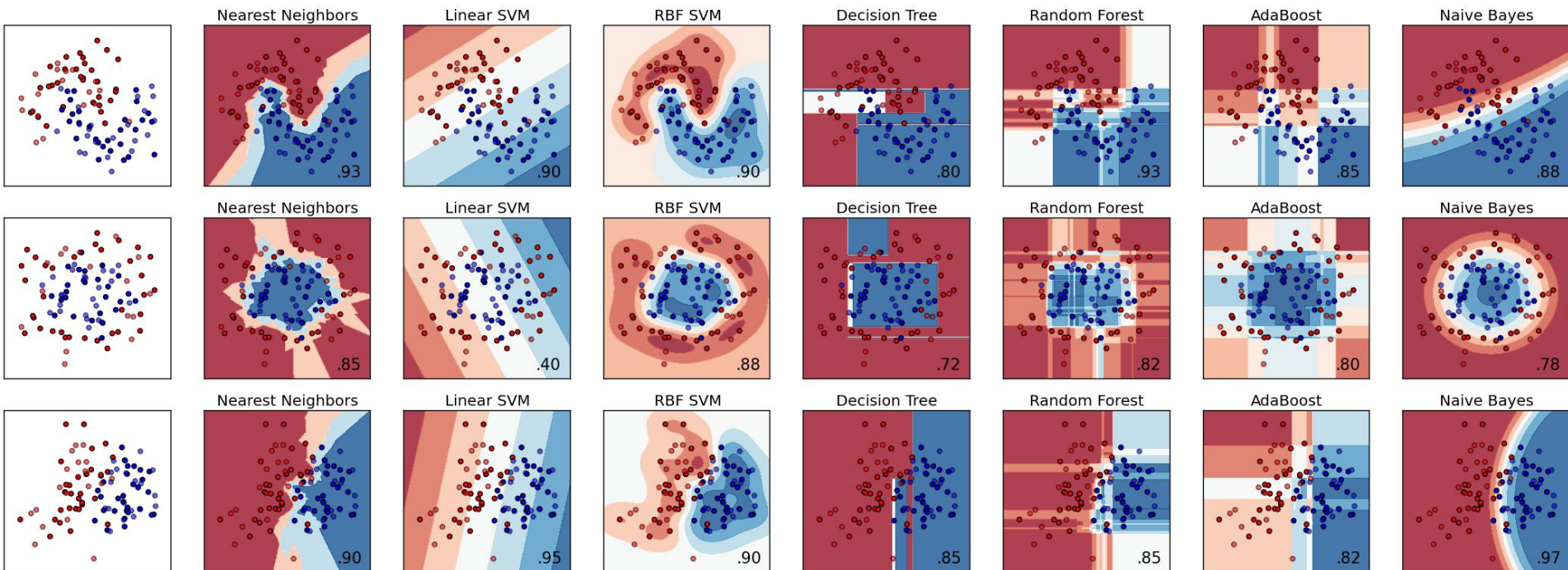


```

graph TD
    Root([do you have labeled data]) -- NO --> Spectral[Spectral Clustering]
    Root -- NO --> GMM[GMM]
    Root -- YES --> KMeans[KMeans]
    KMeans -- YES --> C1([number of categories known])
    C1 -- YES --> S1([<10K samples])
    C1 -- NO --> MBK[MiniBatch KMeans]
    S1 -- YES --> C2([number of categories known])
    C2 -- YES --> MS[MeanShift]
    C2 -- YES --> VBG[VBGMM]
    C2 -- NO --> Tough([tough])
    S1 -- NO --> Spectral
    
```

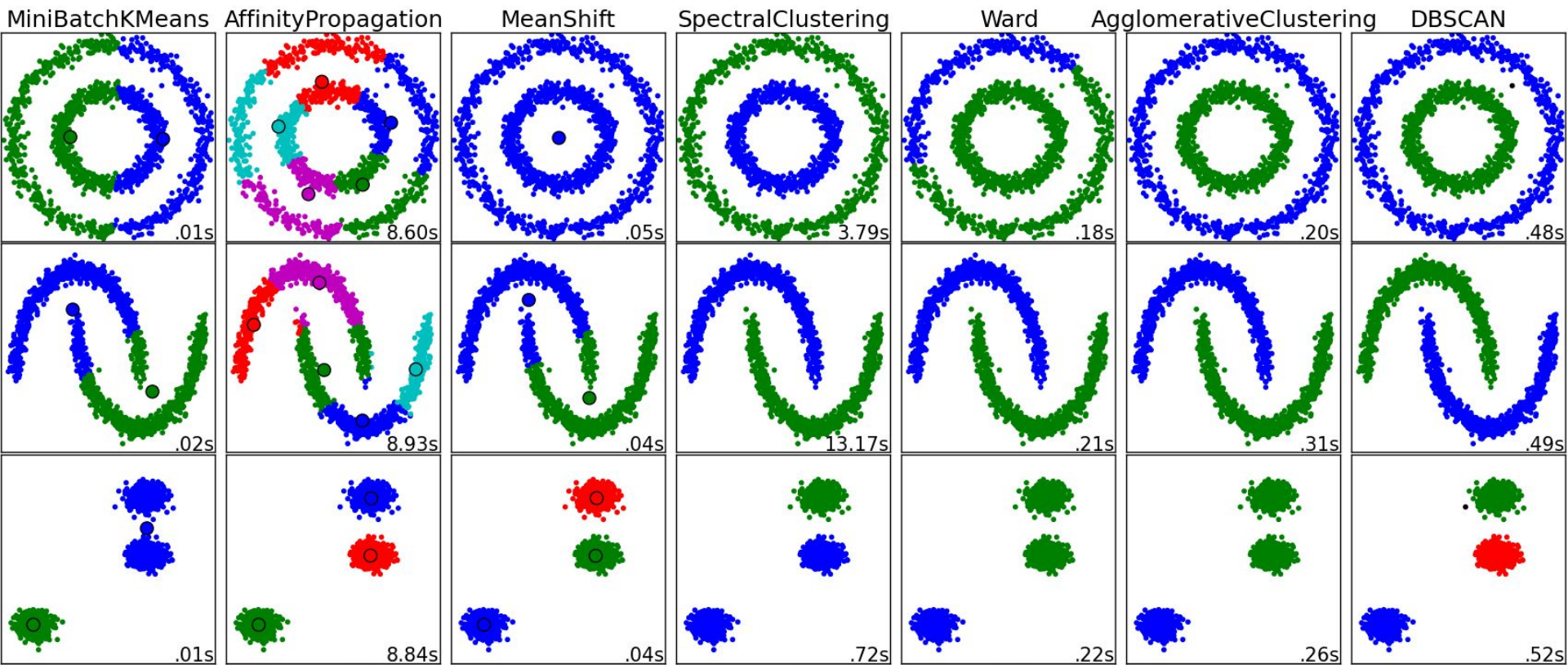


# Frontière de décision





# Identification de clusters



# Apprentissage

---

- ❑ Optimisation d'une métrique par rapport aux données disponibles
- ❑ Validation de la capacité du modèle à généraliser à de nouvelles données

# Valider la généralisation

---

Dataset complet

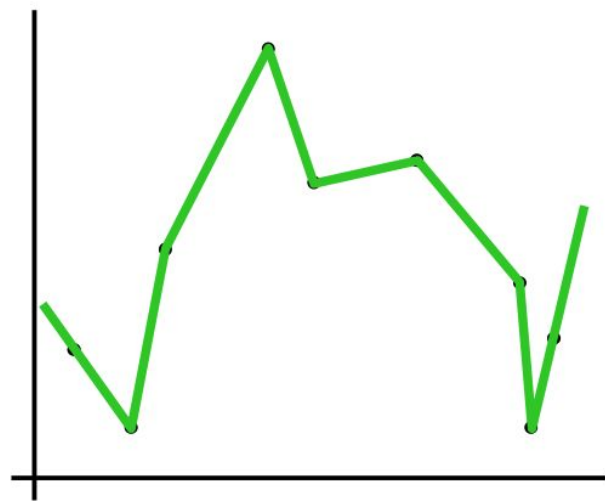
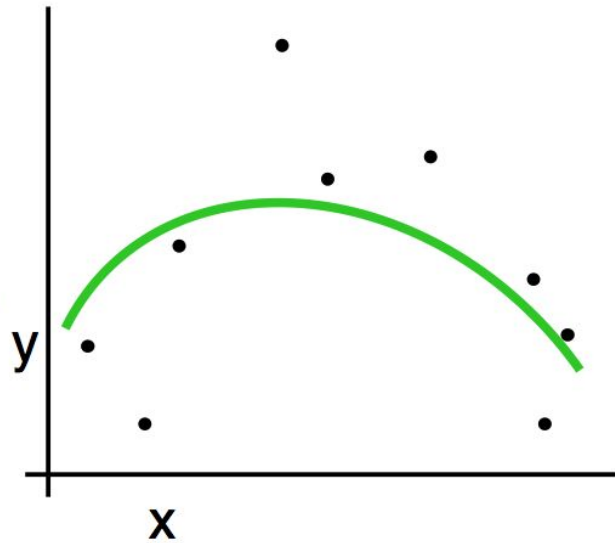
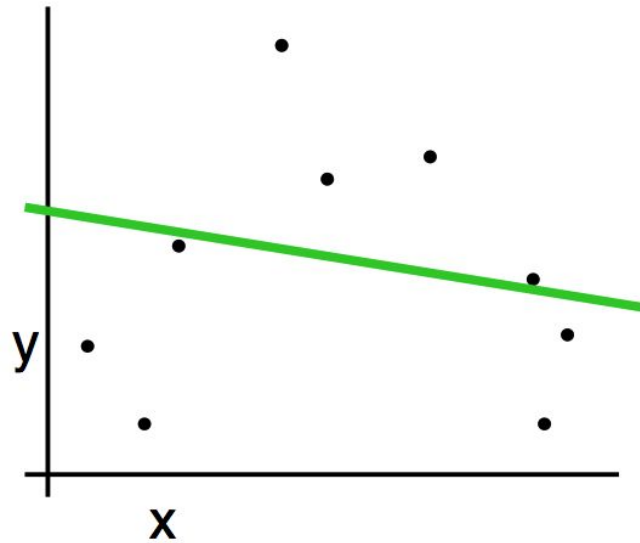
Train

Optimisation

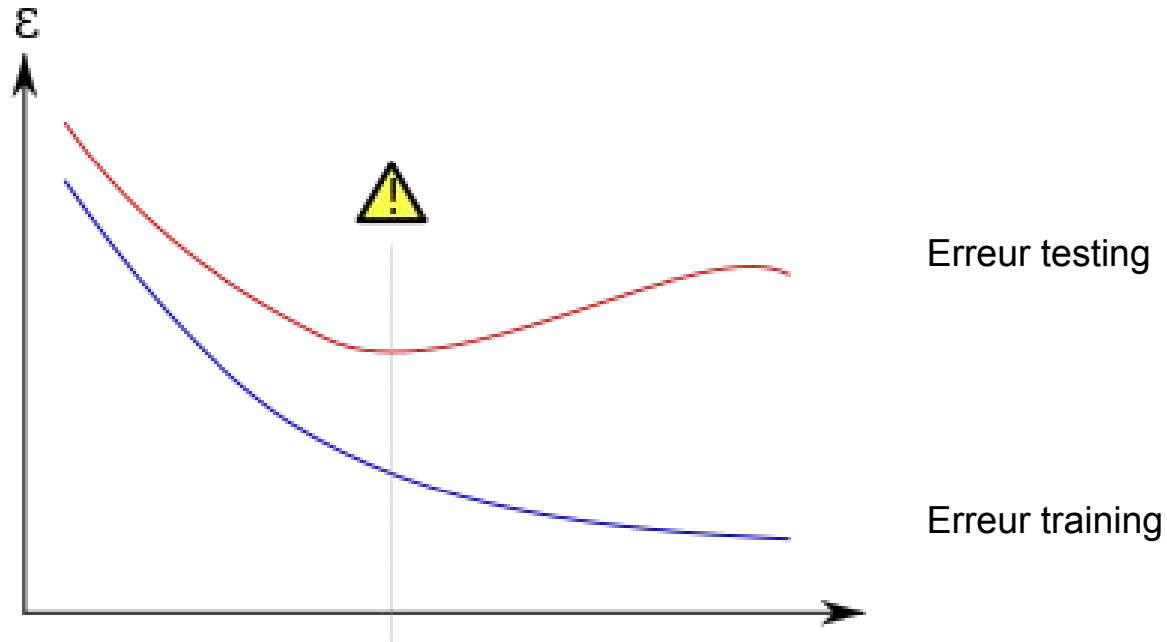
Validation

Test

# Problèmes d'overfitting



# Problèmes d'overfitting







# scikit-learn

Machine Learning in Python

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

## Classification

Identifying to which category an object belongs to.

**Applications:** Spam detection, Image recognition.

**Algorithms:** *SVM*, *nearest neighbors*, *random forest*, ...

— *Examples*

## Regression

Predicting a continuous-valued attribute associated with an object.

**Applications:** Drug response, Stock prices.

**Algorithms:** *SVR*, *ridge regression*, *Lasso*, ...

— *Examples*

## Clustering

Automatic grouping of similar objects into sets.

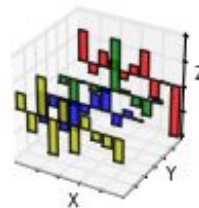
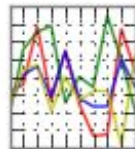
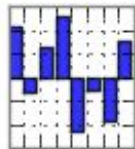
**Applications:** Customer segmentation, Grouping experiment outcomes

**Algorithms:** *k-Means*, *spectral clustering*, *mean-shift*, ...

— *Examples*

# pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



[overview](#) // [get pandas](#) // [documentation](#) // [community](#) // [talks](#)

## Python Data Analysis Library

*pandas* is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the [Python](#) programming language.

### Note

We are proud to announce that *pandas* has become a sponsored project of the ([NUMFocus organization](#)). This will help ensure the success of development of *pandas* as a world-class open-source project.

## VERSIONS

### Release

0.17.1 - November 2015

[download](#) // [docs](#) // [pdf](#)

### Development

0.18.0 - January 2015

[github](#) // [docs](#)

### Previous Releases

0.17.0 - [download](#) // [docs](#) // [pdf](#)