



INTRODUCTION AU MACHINE LEARNING

(D'APRÈS LE LIVRE « HANDS ON MACHINE LEARNING » D'OLIVIER GERON)

MERCREDI XX OCTOBRE

NICOLAS MAKAROFF

PLAN :

- I. Qu'est-ce que la machine learning ?
 - i. Machine Learning ?
 - ii. Pourquoi ?
 - iii. Conclusion
- II. Les différents systèmes de Machine Learning.
 - I. Supervisé vs. Non-Supervisé
 - II. Reinforcement Learning
- III. Les principaux challenges.



QU'EST-CE QUE LE MACHINE LEARNING ?

DATASCIENCE

NICOLAS MAKAROFF

3

MACHINE LEARNING ?

- C'est le champ d'étude qui permet aux ordinateurs d'apprendre sans être explicitement programmés. (Arthur Samuel, 1959)
- On dit qu'un programme informatique apprend à partir d'une expérience E selon une tâche T et une mesure de performance P, si la performance sur T mesurée par P s'améliore avec l'expérience E. (Tom Mitchell, 1997)

Exemple :

Le filtre anti-spam de votre boîte mail

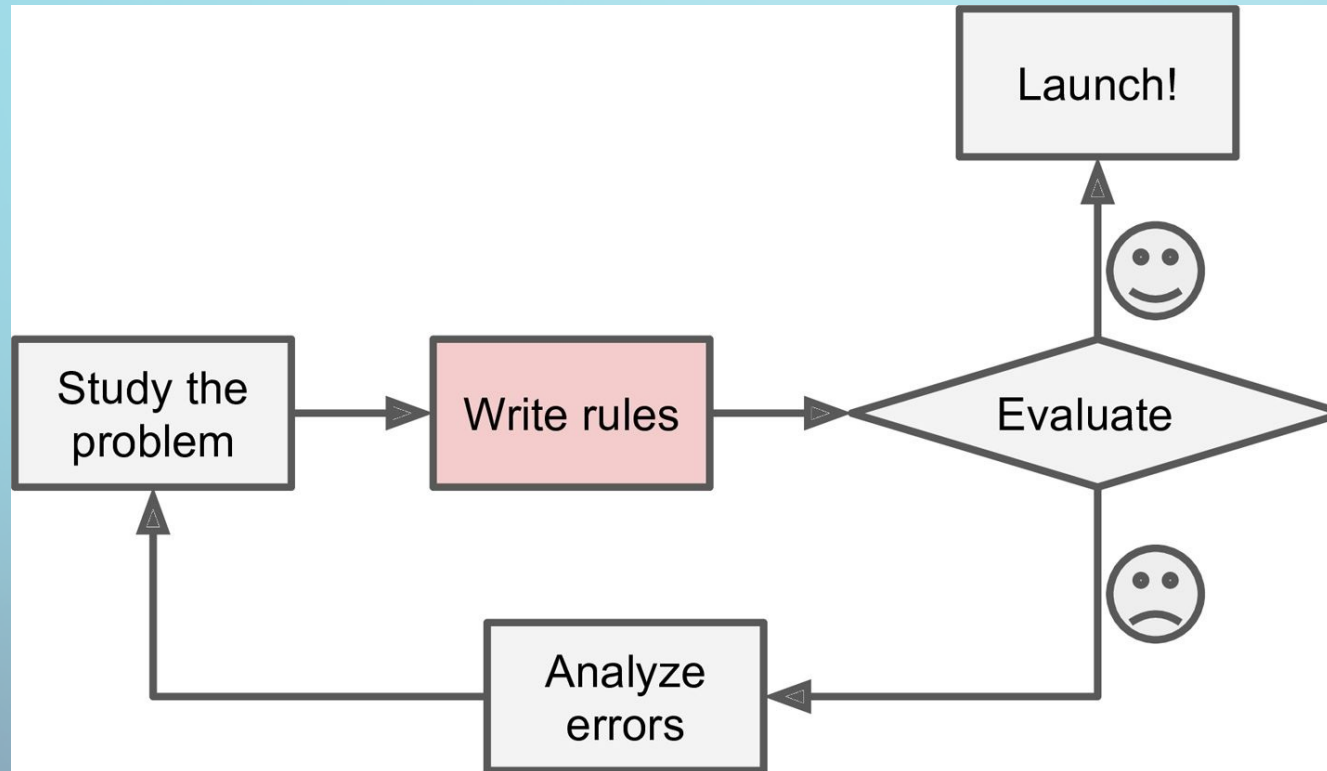
T : repérer les spams

E : training data > e-mails déjà notés comme spam

P : accuracy

POURQUOI LE MACHINE LEARNING ?

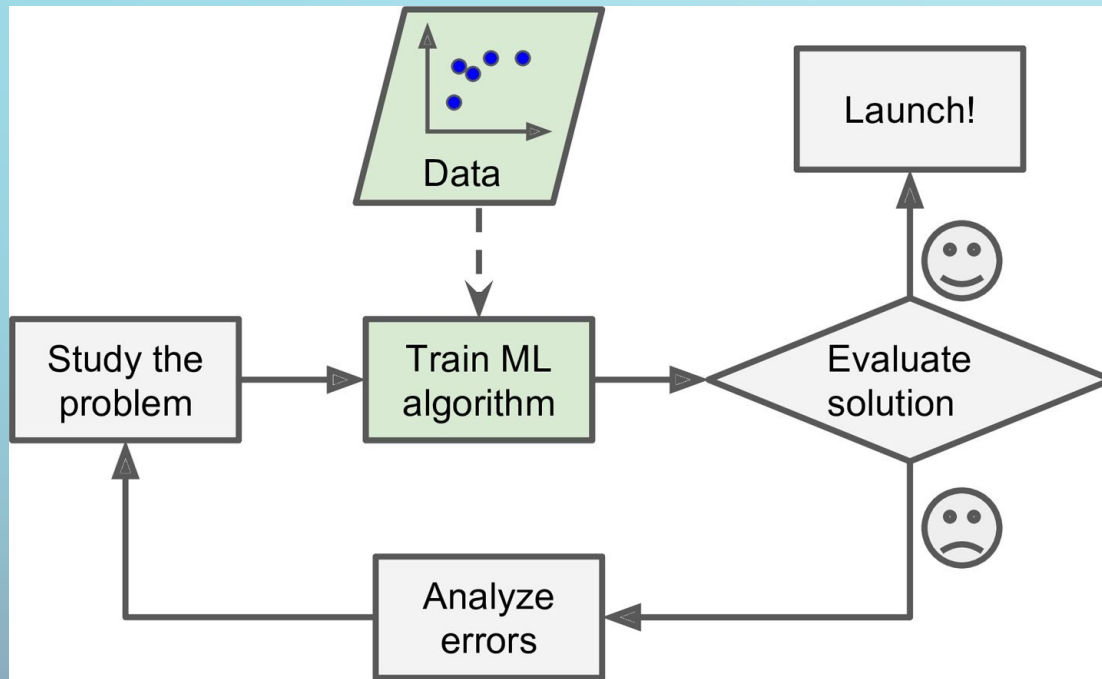
L'approche traditionnelle



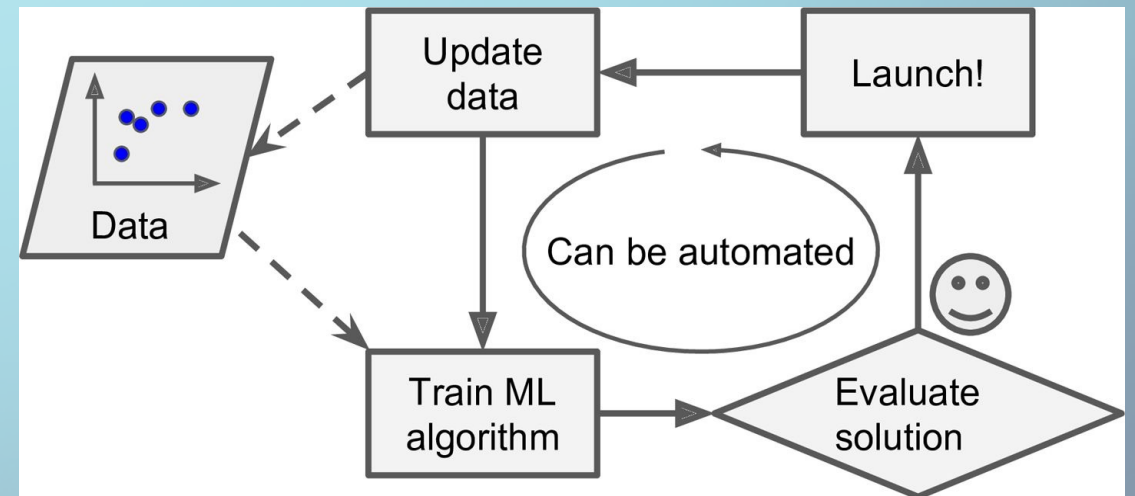
1. Non trivial
2. Longue liste de tâches complexes
3. Dure à maintenir

MACHINE LEARNING ?

Approche Machine Learning



Adaptation automatique



CONCLUSION

C'est bien pour :

- ✓ Les problèmes pour lequel la solution à besoin de faire varier énormément de paramètres ou avec une longue liste de règle.
- ✓ Les problèmes complexes pour lequel il n'existe pas de solution avec l'approche classique.
- ✓ Les environnements qui évoluent.
- ✓ Apprendre les problèmes complexes et de massives banque de données.



LES DIFFÉRENTS SYSTÈMES DE MACHINE LEARNING

DATASCIENCE

NICOLAS MAKAROFF

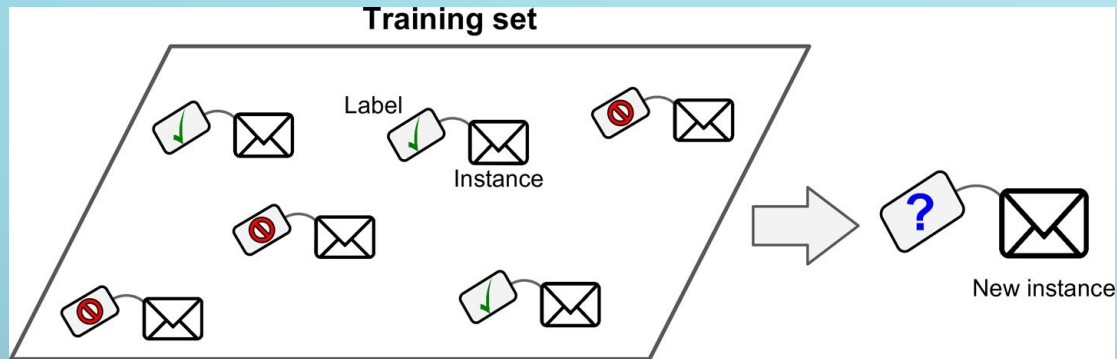
8

LES DIFFÉRENTS SYSTÈMES DE MACHINE LEARNING

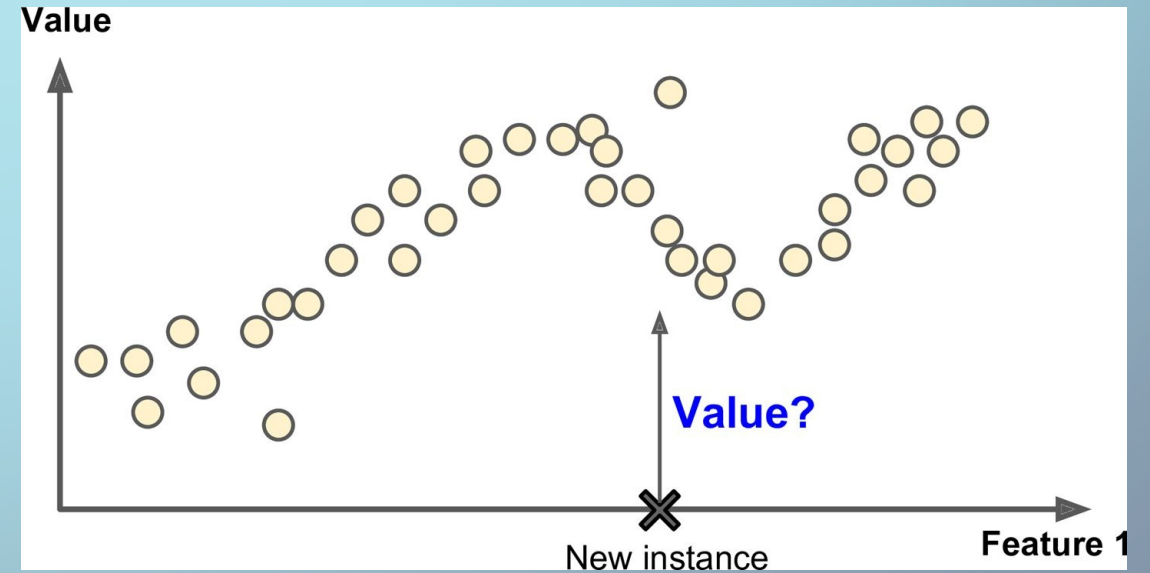
- Avec supervision humaine ou sans
- incrémental à la volée
- simple comparaison de nouvelles données avec des données déjà connues,
détecte des patterns dans les données

APPRENTISSAGE SUPERVISÉ

Classification



Régression

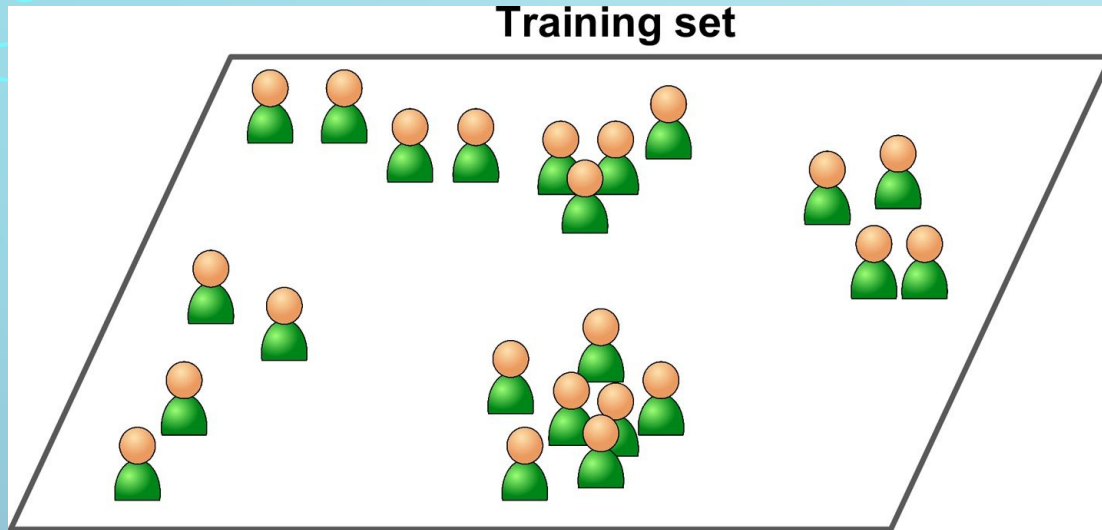


MÉTHODES D'APPRENTISSAGE SUPERVISÉ

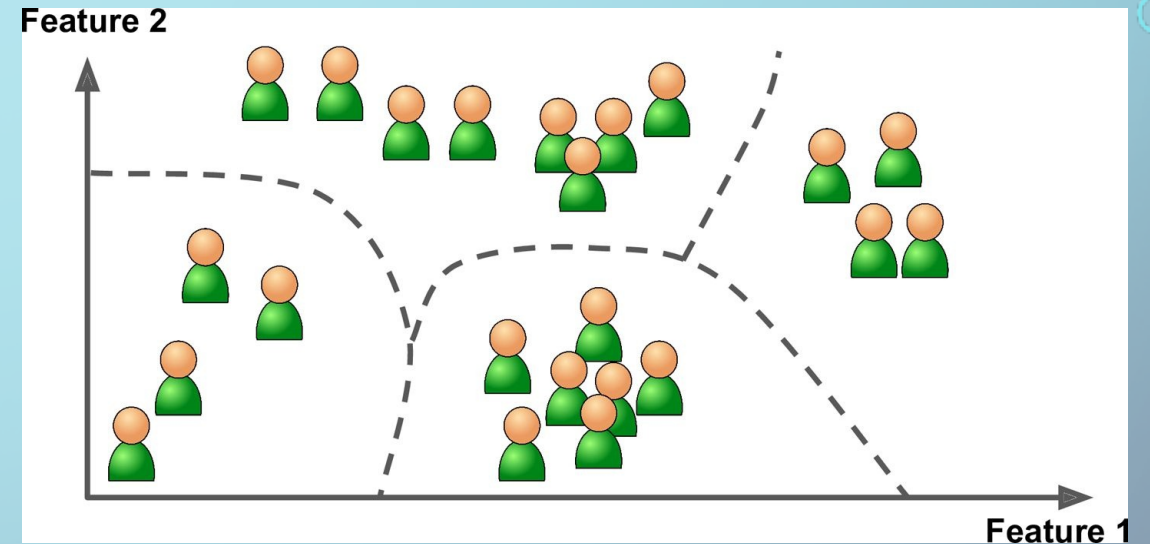
- k-Nearest Neighbors
- Régression linéaire
- Régression logistique
- Support Vector Machines (SVMs)
- Decisions Tree (Arbre de décision) and Random Forest (Forêt aléatoire)
- Neural Network (Réseau de neurones)

APPRENTISSAGE NON-SUPERVISÉ

Avant



Après



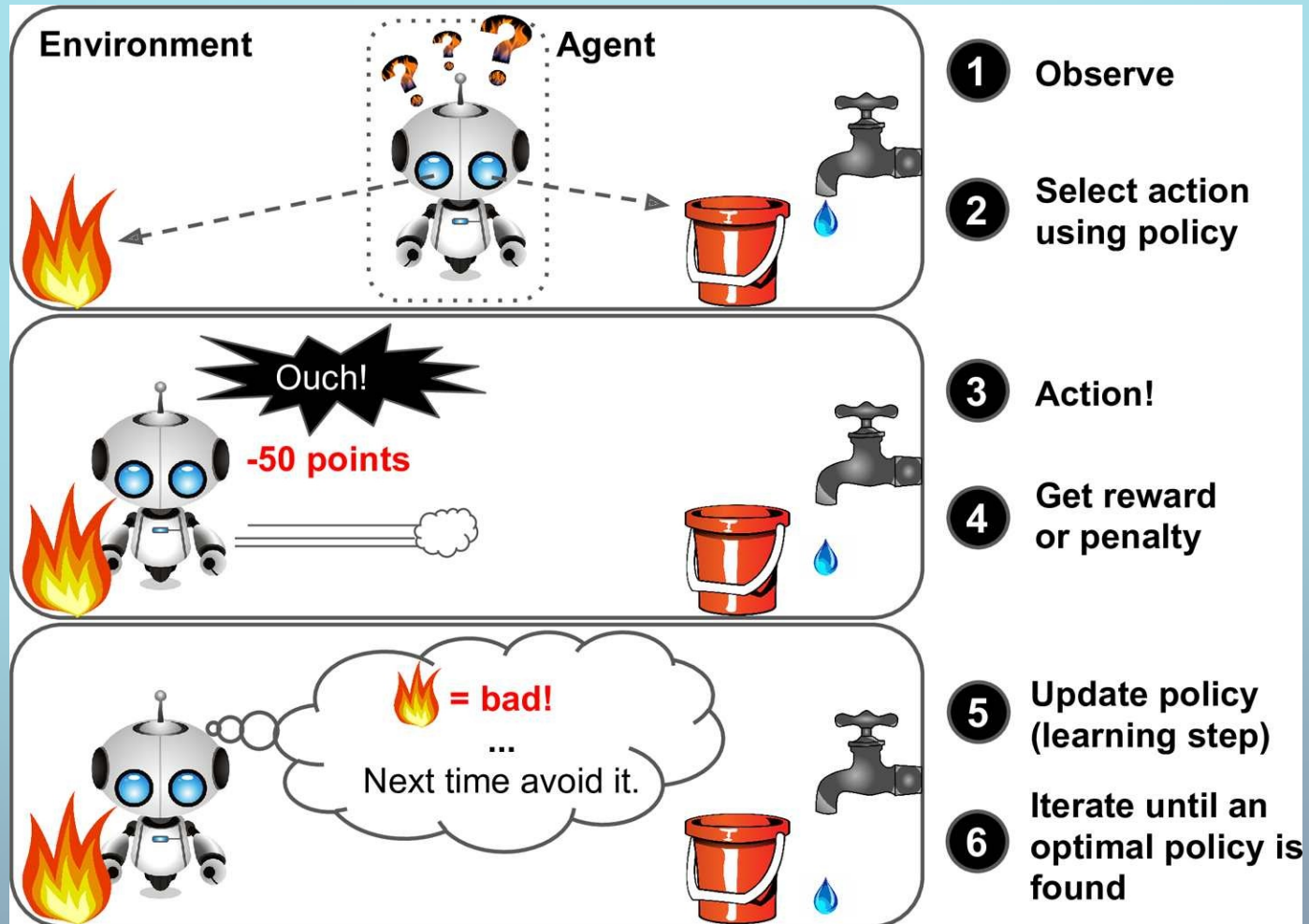
➤ Clustering

- K-Means
- Hierarchical Cluster Analysis (HCA)
- Expectation Maximization

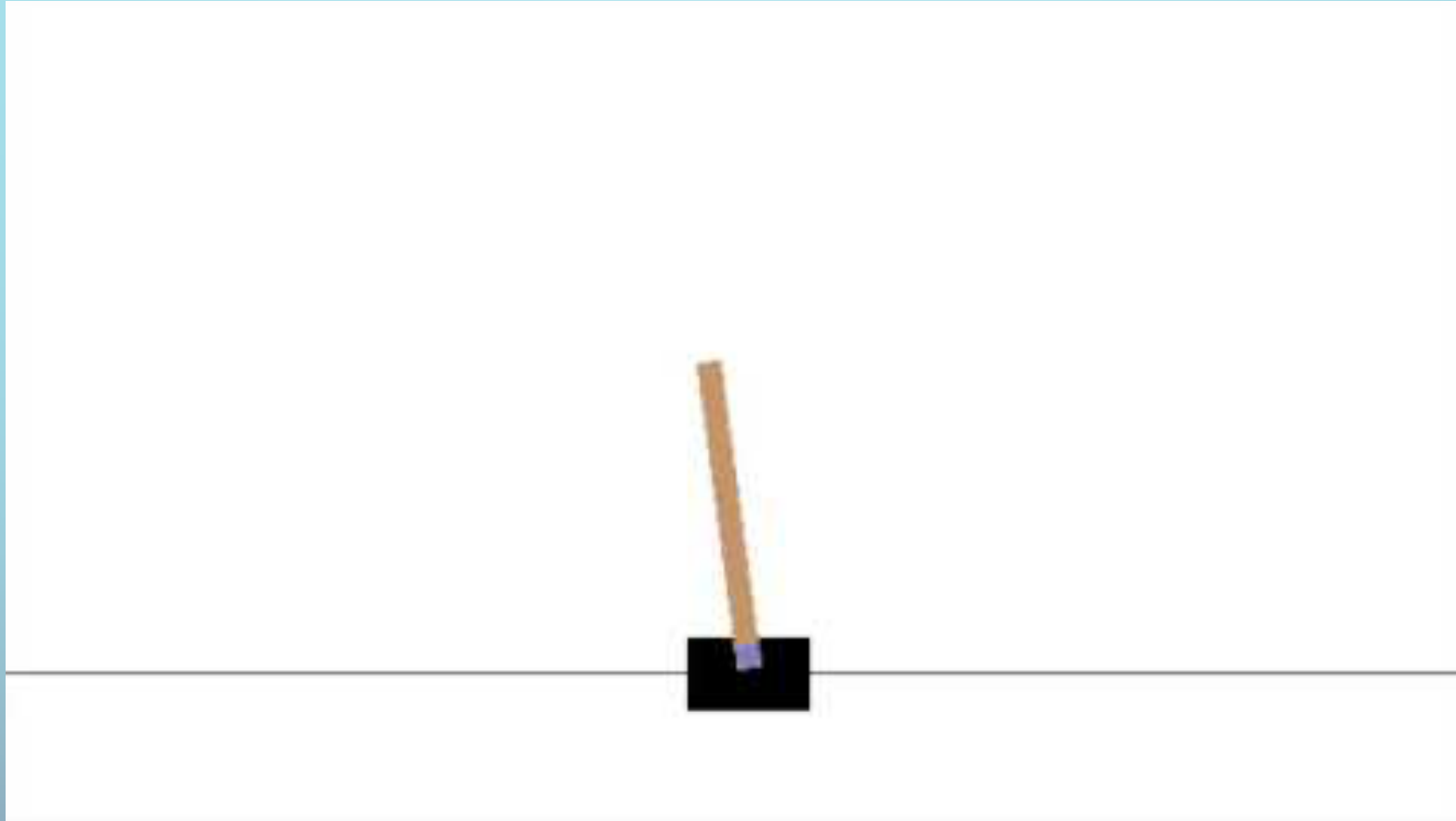
➤ Visualisation et réduction de dimension

- Principal Component Analysis (PCA)
- Kernel PCA
- Locally-Linear Embedding (LLE)
- t-distributed Stochastic Neighbor Embedding (t-SNE)

REINFORCEMENT LEARNING



EXEMPLE DU CARTPOLE

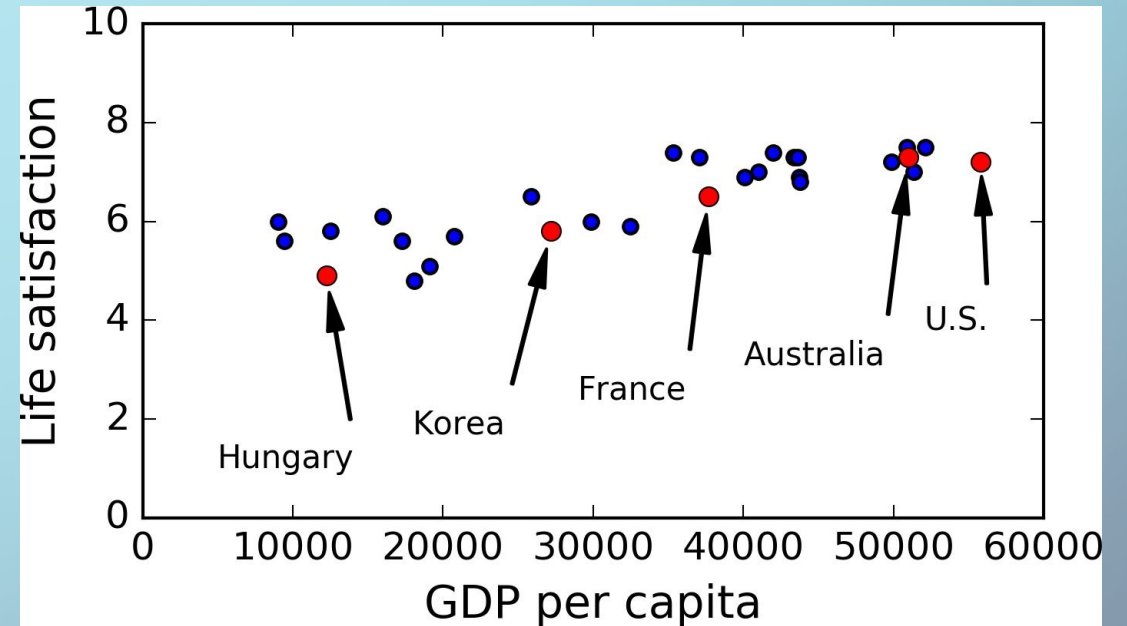


EXEMPLE DU PAC-MAN



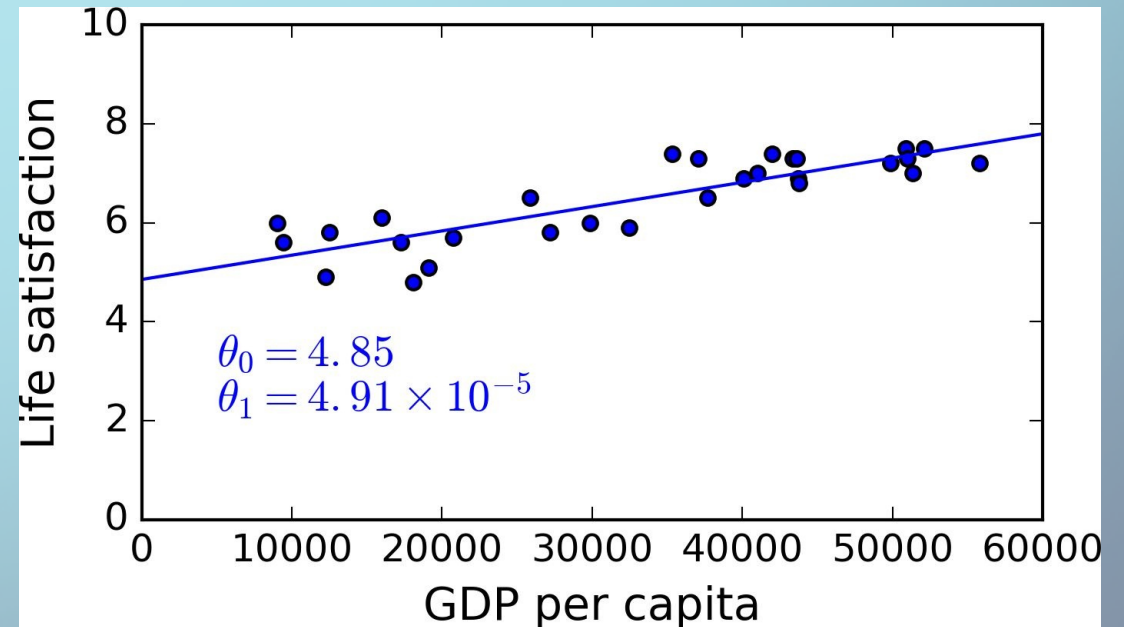
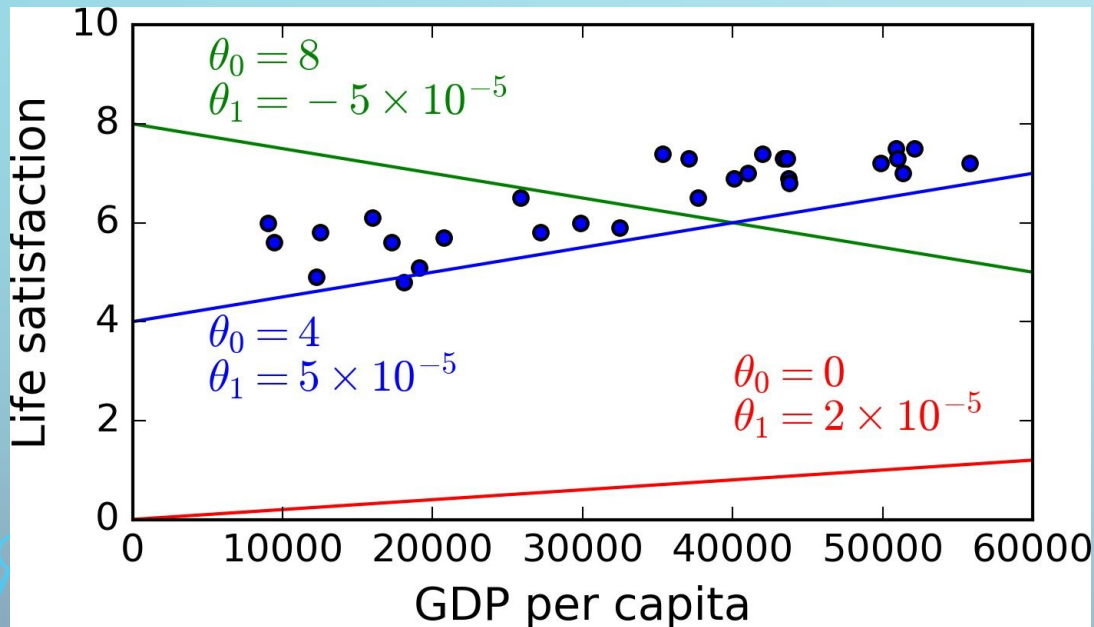
EXEMPLE CONCRET AVEC LE GDP PER CAPITA

Pays	GDP per Capita (USD)	Life satisfaction
Hongrie	12,240	4.9
Corée du Sud	27,195	5.8
France	37,675	6.5
Australie	50,962	7.3
USA	55,805	7.2



EXEMPLE CONCRET AVEC LE GDP PER CAPITA

$$life_satisfaction = \theta_0 + \theta_1 \times GDP_per_capita$$





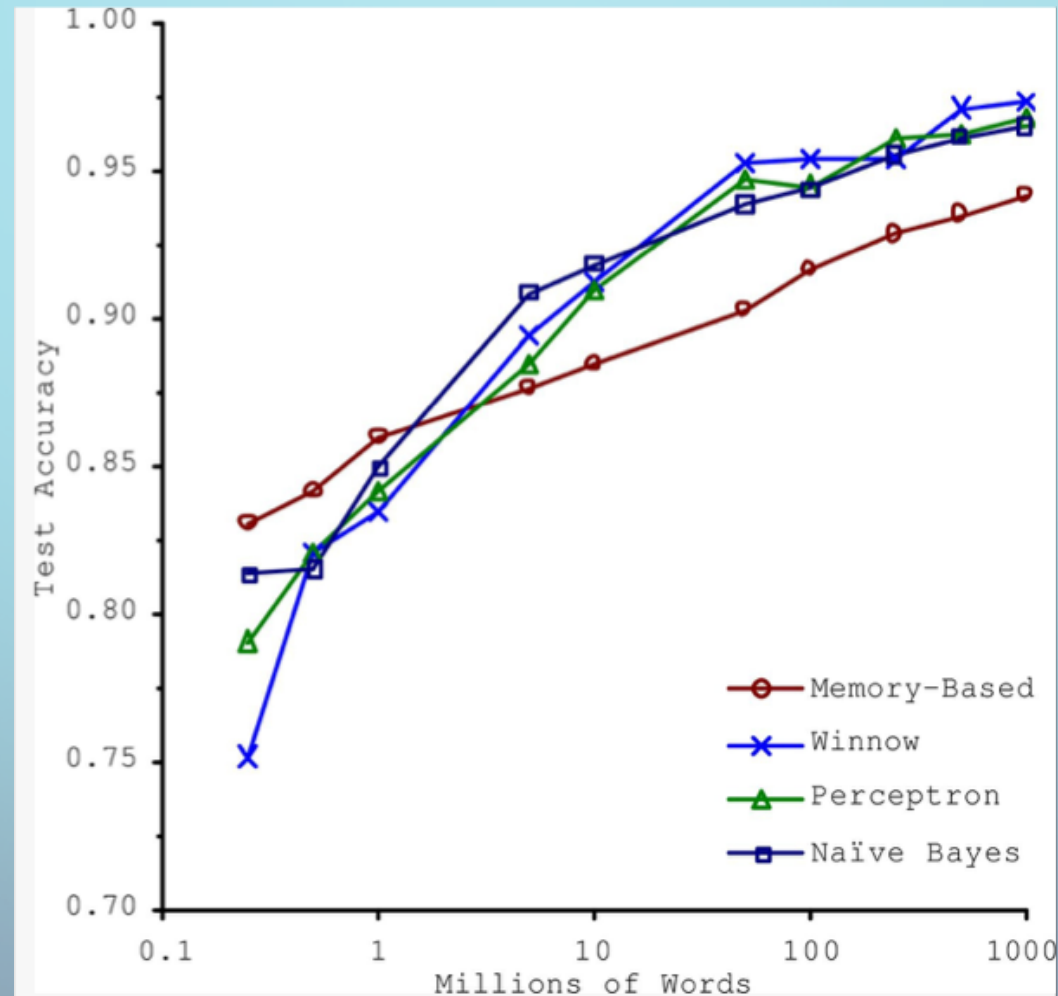
DIFFICULTÉS DU MACHINE LEARNING

DATASCIENCE

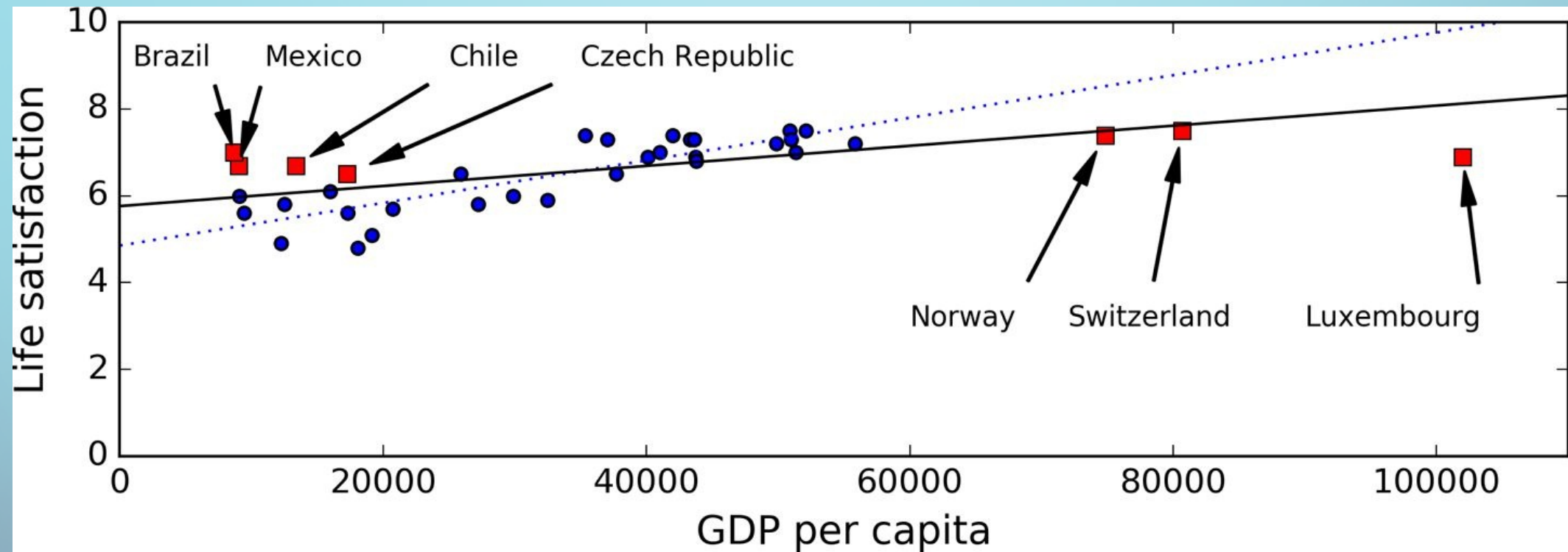
NICOLAS MAKAROFF

18

TAILLE DES BASES DE DONNÉES

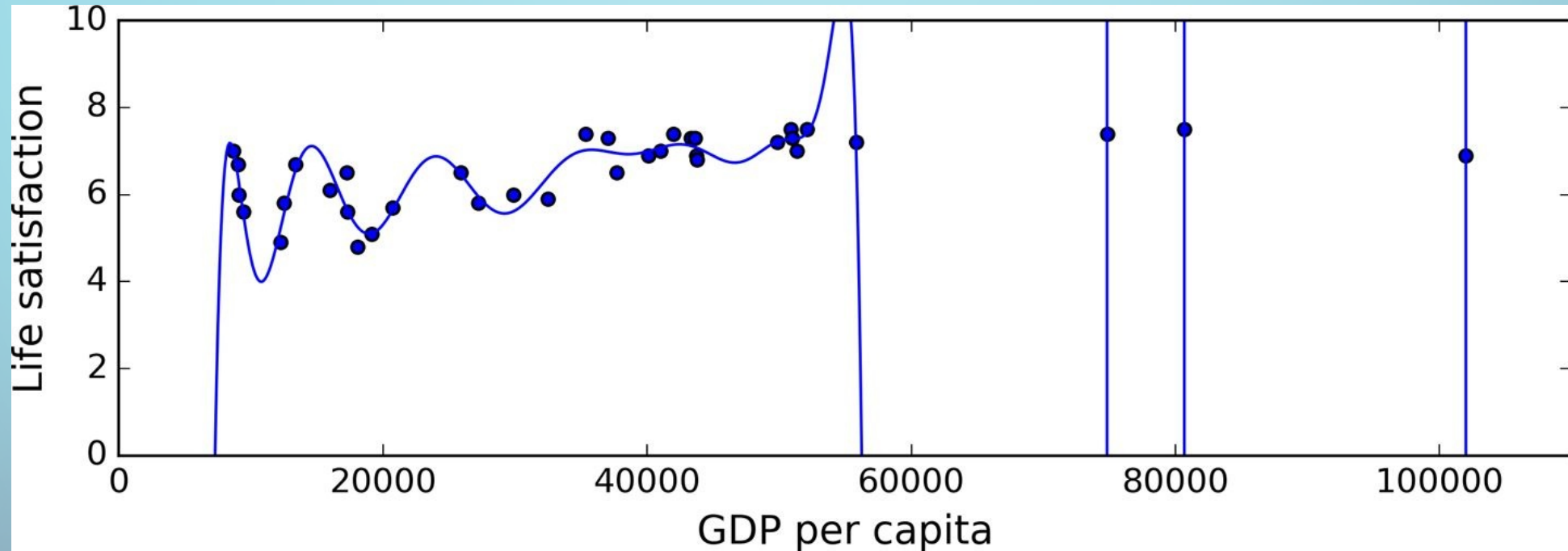


BASES DE DONNÉES NON REPRESENTATIVES



BASES DE DONNÉES NON REPRESENTATIVES

Overfitting



...Underfitting

Source :

Hands On Machine Learning with Scikit-Learn and Tensorflow, Olivier Geron

Conseil de lecture :

..le livre qui est au-dessus de préférence la 2nd édition

[pdf de la 1^{ère} édition](#)

Vidéos :

La série de 3Blue1Brown sur les réseaux de neurones : [lien vers la playlist](#)