

Introduction to Machine Learning

A. Larhlimi

Nov. 2025

Machine Learning: introduction

- Machine Learning \equiv data-based programming
 - ability of computers to learn how to perform tasks (classification, detection, translation...) without being explicitly programmed
 - study of algorithms that **improve their performance** at **some task** based on **experience**

What is Machine Learning?

Learn From Experience



Data
Learn From **Experience**



Follow Instructions



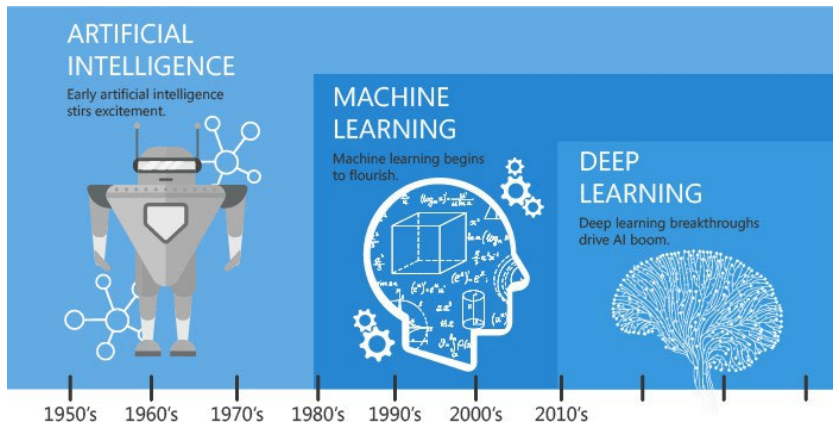
The rise

- Big Data : continuous increase in data generated
 - Twitter : 50M tweets /day (=7 terabytes)
 - Facebook : 10 terabytes /day
 - Youtube : 50h of uploaded videos /minute
 - 2.9 millions of e-mails /second

- Moore's law
- Massively distributed computing

- Interest: from product to customers.
- Data Mining \equiv discovering patterns in large data sets

Historical perspective



Today's AI is Deep Learning (a technique of Machine Learning)

<https://blog.alore.io/machine-learning-and-artificial-intelligence/>

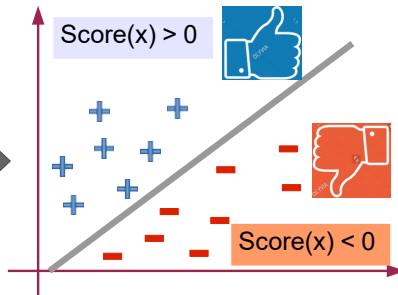
Applications: sentiment analysis

- Classify stored reviews according to users' sentiment

Here is my review about Tarantino's last movie I've watched last saturday. It was really awesome. I enjoy the characters, the script and the music. I fully recommend everyone to go to theater and enjoy the movie. |

Yeah, yeah, what to say about this movie ? It is not that bad, but it is not the top Tarantino's movies. Has some good features (music, landscapes) but also bad features

As for the previous movies of Tarantino, I was bored all along the movie. The script was quite complicated to follow up with several references to past events you should know about before. Must not see movie, pass by your way.



Product recommendation

	Feature 1	Feature 2
User 1	?	?
User 2	?	?
User 3	?	?
User 4	?	?
User 5	?	?

User
features

X

	Item 1	Item 2	Item 3	Item 4	Item 5
Feature 1	?	?	?	?	?
Feature 2	?	?	?	?	?

Product
features

Matrix factorization
Purchase history of customers

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	0 ?	3	0 ?	3	0 ?
User 2	4	0 ?	0 ?	2	0 ?
User 3	0 ?	0 ?	3	0 ?	0 ?
User 4	3	0 ?	4	0 ?	3
User 5	4	3	0 ?	4	0 ?

https://katbailey.github.io/images/matrix_factorization.png

Purchased item

Recommended products



Aigital Répéteur WiFi



AFTERSHOKZ trekz

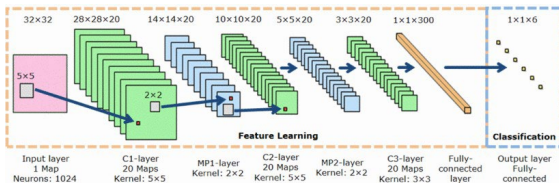


cowin E7 Casque Audio à

Image classification

Labeled training images

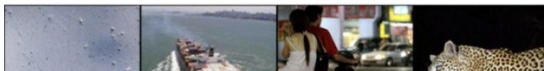
Deep classification architecture



[https://www.researchgate.net/profile/Y_Nikitin/publication/270163511/figure/download/fig5/AS:](https://www.researchgate.net/profile/Y_Nikitin/publication/270163511/figure/download/fig5/AS:295194831409153@1447391340221/MPCNN-architecture-using-alternating-convolutional-and-max-pooling-layers-13.png)

295194831409153@1447391340221/MPCNN-architecture-using-alternating-convolutional-and-max-pooling-layers-13.png

Input images and predicted category



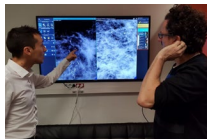
Medical diagnosis



Malignant melanoma detection

130 000 images including over 2 000 cases of cancer

Error rate 28 % (human 34 %)



Digital Mammography DREAM Challenge

640 000 mammographies (1209 participants)

false-positive rate decreased by 5 %

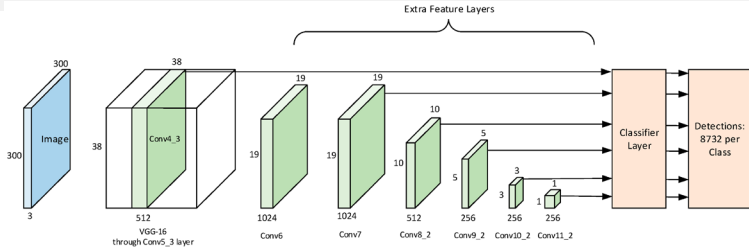


Heart rhythm analysis

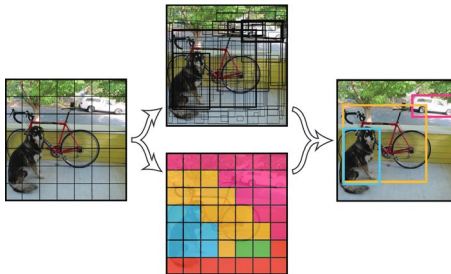
500 000 ECG

accuracy 92.6 % (human 80.0 %) sensitivity of 97 %

Object detection

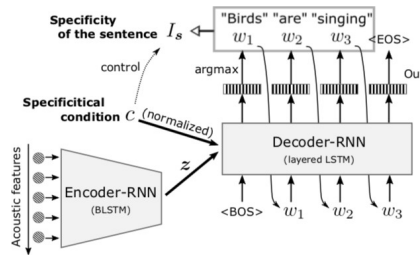
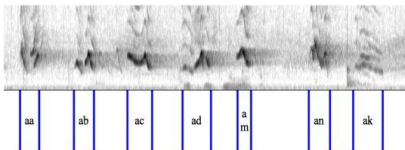


https://www.mdpi.com/applsci/applsci-09-01128/article_deploy/html/images/applsci-09-01128-g004.png



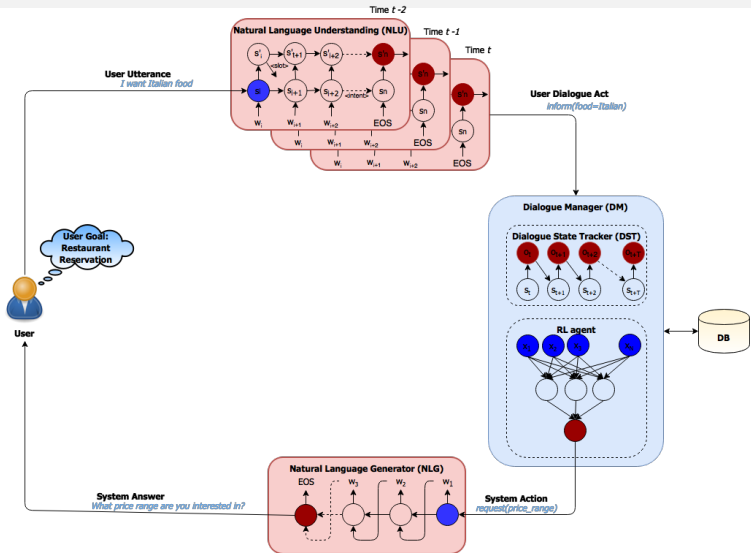
<https://arxiv.org/pdf/1506.02640.pdf>

Audio captioning



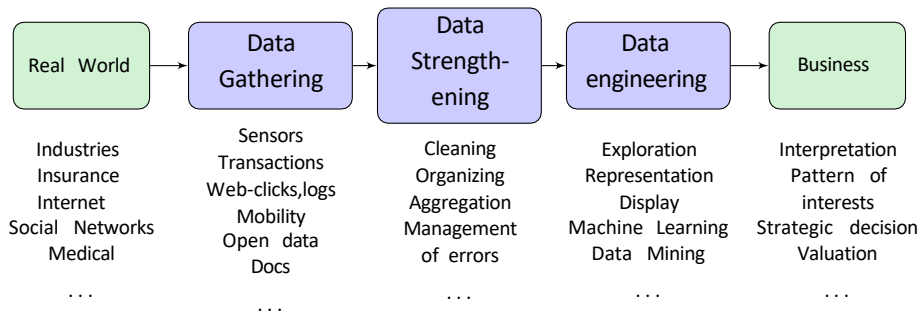
<https://ars.els-cdn.com/content/image/1-s2.0-S1574954115000151-gr1.jpg>

Chatbot

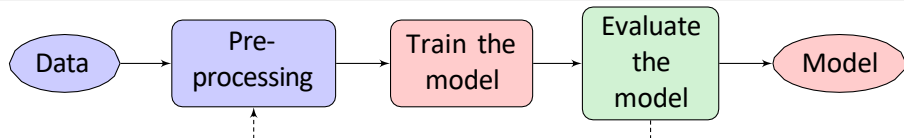


https://miro.medium.com/max/2058/1*LF5T9fsr4w2EqyFJkb_gng.png

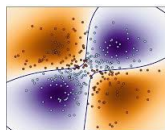
Implementing a Machine Learning project



Chain of the data engineering process



- 1 Understand and specify project goals
- 2 Pre-processing/visualize/analyze data
- 3 Which ML problem is it?
- 4 Design a solving approach
- 5 Evaluate its performance
- 6 Go to 2) if needed



Y

A matrix illustrating a recommendation system. The rows represent users (indicated by small icons on the left) and the columns represent movies (indicated by movie posters at the top). The cells contain binary values (0 or 1) representing whether a user has watched a movie. The third row from the top has a green box with a question mark in the second column, indicating a recommendation for that user.

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5	Movie 6
User 1	1	1	0	0	0	0
User 2	1	0	0	1	1	0
User 3	0	?	1	1	0	0
User 4	1	0	1	0	0	0
User 5	0	0	0	1	0	0
User 6	0	0	1	1	0	0

Course Goal: Study the steps from 2 to 5

The data

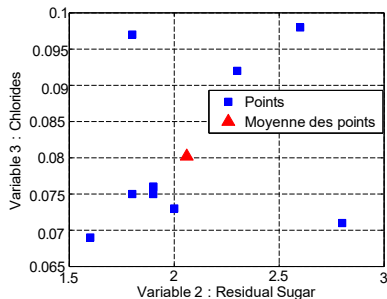
- Information (past experience) are examples with attributes
- Assume the data set consists of N samples

- An attribute is a **property** or **characteristic** of a phenomenon being observed. Also termed **feature** or **variable**

- It is an entity characterising an object; it is made up of attributes.
- Synonyms : **instance**, **point**, **vector** (usually in \mathbb{R}^d)

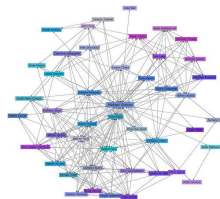
Data : visualization

Points x	citric acid	residual sugar	chlorides	sulfur dioxide
1	0	1.9	0.076	11
2	0	2.6	0.098	25
3	0.04	2.3	0.092	15
Point $x \in \mathbb{R}^4$	0.56	1.9	0.075	17
5	0	1.9	0.076	11
6	0	1.8	0.075	13
7	0.06	1.6	0.069	15
8	0.02	2	0.073	9
9	0.36	2.8	0.071	17
10	0.08	1.8	0.097	15

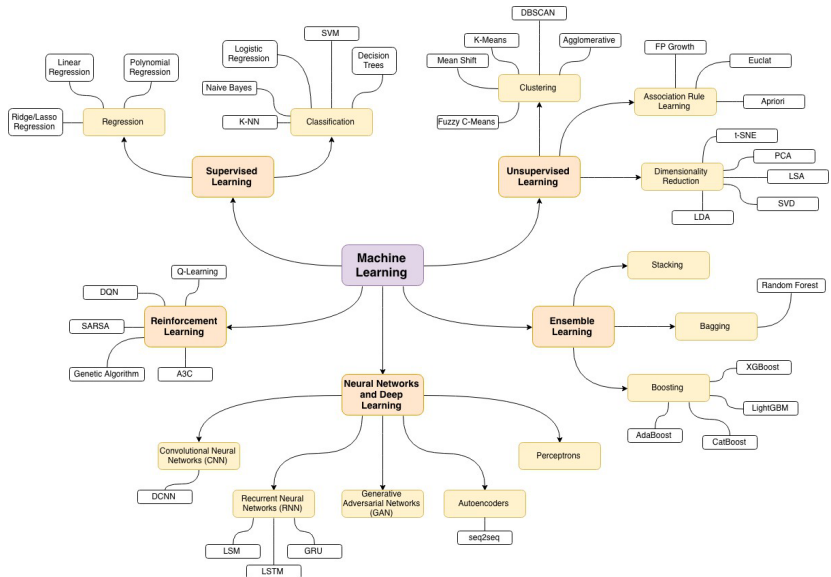


Data types

- Sensors → Quantitative and qualitative variables, ordinales, nominals
- Text → String
- Speech → Time Series
- Images → 2D Data
- Videos → 2D Data + time
- Networks → Graphs
- Stream → Logs, coupons. . .
- Labels → Expected output prediction



Approaches of Machine Learning

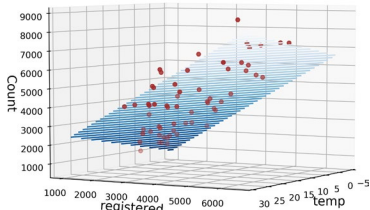


Supervised learning

- Given a set of N training examples $\{(x_i, y_i) \in X \times Y, i = \dots, N\}$, we want to estimate a prediction function $y = f(x)$.
- The *supervision* comes from the label knowledge

- Image classification, object detection, stock price prediction ...

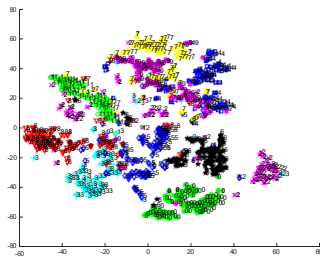
Régression linéaire avec 2 variables



Unsupervised learning

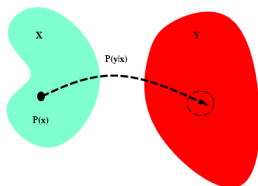
- Only the $\{x_i \in X, i = \dots, N\}$ are available. We aim to describe how data is organized and extract homogeneous subsets from it.

- Applications: Customers segmentation, image segmentation, data visualization, categorization of similar documents ...



Supervised learning : concept

- Let X and Y be two sets. Assume $p(X, Y)$ the joint probability distribution of $(X, Y) \in X \times Y$.



- Goal** : find a prediction function $f : X \rightarrow Y$ which correctly estimates the output y corresponding to x .
- f belongs to a space H called **hypothesis class**. Example of H : set of polynomial functions

Supervised learning: principle

- **Loss function** $L(Y, f(X))$

- evaluates how "close" is the prediction $f(x)$ to the true label y
- it penalizes errors: $L(y, f(x)) = \begin{cases} 0 & \text{if } y = f(x) \\ \geq 0 & \text{if } y \neq f(x) \end{cases}$

- **True risk** (expected prediction error)

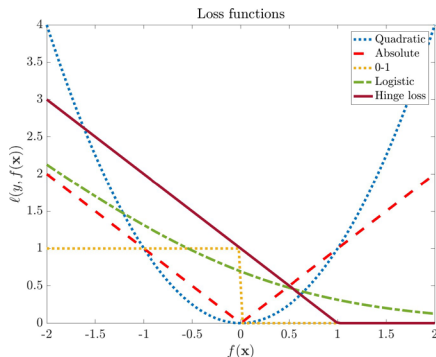
$$R(f) = E_{(X,Y)}[L(Y, f(X))] = \int_{X \times Y} L(y, f(x)) p(x, y) dx dy$$

Identify the prediction function which minimizes the true risk i.e.

$$f^* = \arg \min_{f \in H} R(f)$$

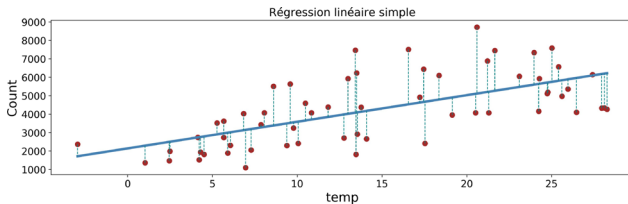
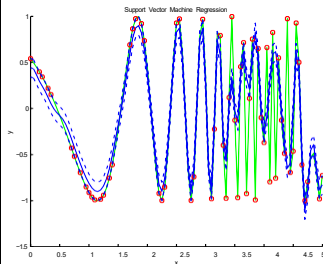
Loss functions

- Quadratic loss : $L(Y, f(X)) = (Y - f(X))^2$
- l_1 loss (absolute deviation): $L(Y, f(X)) = |Y - f(X)|$
- 0 - 1 loss: $L(y, f(x)) = \begin{cases} 1 & \text{if } y \neq f(x) \\ 0 & \text{otherwise} \end{cases}$
- Hinge loss: $L(y, f(x)) = \max(0, 1 - yf(x))$



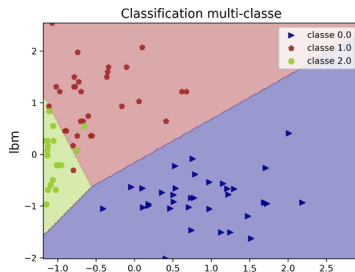
Some supervised learning problems

- We talk about **regression** when Y is a subset of \mathbb{R}^d .
- Usual related loss function: quadratic loss $(y - f(x))^2$



Some supervised learning problems

- Output space Y is an un-ordered discrete set
- Binary classification: $\text{card}(Y) = 2$
 - Example: $Y = \{-1, 1\}$
 - Loss functions: 0-1 loss, hinge loss
- Multiclass classification: $\text{card}(Y) > 2$
 - Example: $Y = \{1, 2, \dots, K\}$



From true risk to empirical risk

Minimizing the true risk is not doable (in almost all practical applications)

- The joint distribution $p(X, Y)$ is unknown!
- Only a finite **training set** $\{(x_i, y_i) \in X \times Y\}_{i=1}^N$ is available

- **Empirical risk**

$$R_{\text{emp}}(f) = \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i))$$

- Empirical risk minimization

$$\hat{f} = \arg \min_{f \in H} R_{\text{emp}}(f)$$

Overfitting

Empirical risk is not appropriate for model selection: if H is large enough, $R_{emp}(f) \rightarrow 0$ but the generalized error (true risk) is high.

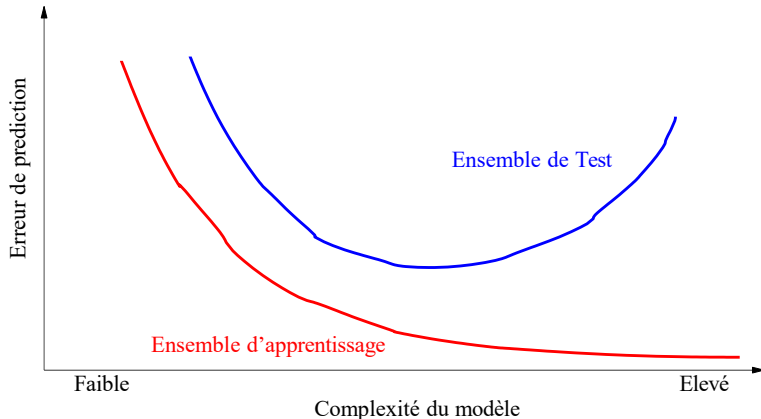
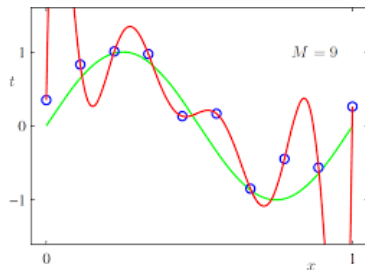
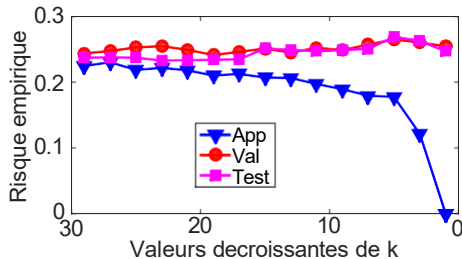


Illustration of overfitting



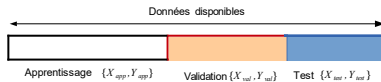
https://www.cs.princeton.edu/courses/archive/spring16/cos495/slides/ML_basics_lecture6_overfitting.pdf

Model selection

- Find in H the best function f that learned based on the training set will well generalize (low true risk)
- Example : We are looking for a polynomial function of degree α minimizing the risk : $R_{emp}(f_\alpha) = \sum_{i=1}^N (y_i - f_\alpha(x_i))^2$.
- Goal :
 - propose a model estimation method in order to choose (approximately) the best model belonging to H .
 - once the model is selected, estimate its generalization error.

Model selection : basic approach

Case 1 : N is really big (large scale D_N)



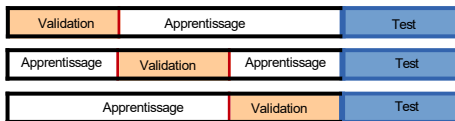
- 1 Randomly split $D_N = D_{train} \cup D_{val} \cup D_{test}$
- 2 For each α , train f_α based on D_{train}
- 3 Evaluating its performance on D_{val} $R_{val} = \frac{1}{N_{val}} \sum_{i \in D_{val}} L(y_i, f(x_i))$
- 4 Select the model with the best performance on D_{val}
- 5 Test selected model on D_{test}

• D_{test} is used once!

Model selection : Cross-validation

Case 2 : Small or medium scale D_N

- Estimate the generalization error by re-sampling.
- Principle
 - 1 Split D_N into K sets of equal size.
 - 2 For each $k = 1, \dots, K$, train a model by using the $K - 1$ remaining sets and evaluate the model on the k -th part.
 - 3 Average the K error estimates obtained to have the cross-validation error.



Conclusions

- Clearly identify and spell out the needs.
- Create or obtain data representative of the problem
- Identify the context of learning
- Analyze and reduce data size
- Choose an algorithm and/or a space of hypotheses
- Choose a model by applying the algorithm to pre-processed data
- Validate the performance of the method

Find your way...

