

Intelligence Artificielle pour l'Analyse de Données

Akka Zemmari

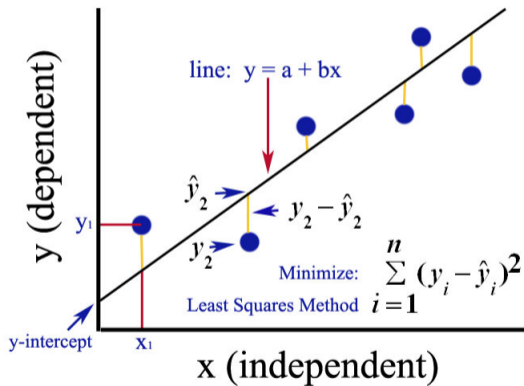
`zemmari@u-bordeaux.fr`

LaBRI, Université de Bordeaux

2019-2020

Machine Learning ?

Pour demystifier la chose :



C'est quoi le Machine Learning

C'est quoi le Machine Learning

- ▶ "The goal of machine learning is to build computer systems that can adapt and **learn** from their experience." - Tom Dietterich

C'est quoi le Machine Learning

- ▶ "The goal of machine learning is to build computer systems that can adapt and **learn** from their experience." - Tom Dietterich
- ▶ To **learn**: to acquire knowledge by study, experience or being taught.

C'est quoi le Machine Learning

- ▶ "The goal of machine learning is to build computer systems that can adapt and **learn** from their experience." - Tom Dietterich
- ▶ To **learn**: to acquire knowledge by study, experience or being taught.
- ▶ A Computer program is said to

C'est quoi le Machine Learning

- ▶ "The goal of machine learning is to build computer systems that can adapt and **learn** from their experience." - Tom Dietterich
- ▶ To **learn**: to acquire knowledge by study, experience or being taught.
- ▶ A Computer program is said to learn by **experience** E

C'est quoi le Machine Learning

- ▶ "The goal of machine learning is to build computer systems that can adapt and **learn** from their experience." - Tom Dietterich
- ▶ To **learn**: to acquire knowledge by study, experience or being taught.
- ▶ A Computer program is said to learn by **experience E** with respect to class of **tasks T**

C'est quoi le Machine Learning

- ▶ "The goal of machine learning is to build computer systems that can adapt and **learn** from their experience." - Tom Dietterich
- ▶ To **learn**: to acquire knowledge by study, experience or being taught.
- ▶ A Computer program is said to learn by **experience E** with respect to class of **tasks T** **performance measure P**,

C'est quoi le Machine Learning

- ▶ "The goal of machine learning is to build computer systems that can adapt and **learn** from their experience." - Tom Dietterich
- ▶ To **learn**: to acquire knowledge by study, experience or being taught.
- ▶ A Computer program is said to learn by **experience E** with respect to class of **tasks T** **performance measure P**, if the **performance** at the task T, measured by performance P, **improves by experience E**.

C'est quoi le Machine Learning

- ▶ **task** to be addressed by the system (e.g. recognizing handwritten characters)

C'est quoi le Machine Learning

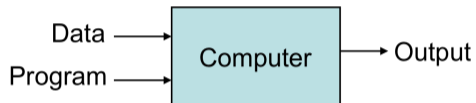
- ▶ **task** to be addressed by the system (e.g. recognizing handwritten characters)
- ▶ **performance measure** to evaluate the learned system (e.g. number of misclassified characters)

C'est quoi le Machine Learning

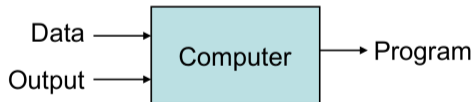
- ▶ **task** to be addressed by the system (e.g. recognizing handwritten characters)
- ▶ **performance measure** to evaluate the learned system (e.g. number of misclassified characters)
- ▶ **training experience** to train the learning system (e.g. labelled handwritten characters)

C'est quoi le Machine Learning

▶ Traditional Programming

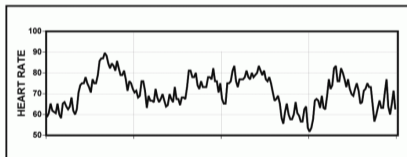


▶ Machine Learning



Quelques Exemples

- ▶ A partir des données médicales (sur 20 ans) sur des patients, peut-on dire si un patient risque d'avoir une crise cardiaque dans les 2,3 prochaines années ?



Quelques Exemples

- ▶ On dispose de données cliniques sur 20 ans. Un patient risque-t-il d'avoir une attaque cardiaque dans les 5 prochaines années ?
- ▶ Quel sera le prix de ce stock dans 6 mois ?



29.90	+12.3%	▲	593.23	120.000
15.00	+5.34%	▲	254.23	320.000
7.34	+5.97%	▲	100.00	120.000
6.89	+2.13%	▲	564.23	300.000
1.45	+6.43%	▲	765.90	600.000
0.67	-11.6%	▼	120.34	300.000
0.54	+23.1%	▲	893.23	120.000
0.39	+5.56%	▲	120.98	320.000
0.29	-3.67%	▼	432.12	750.000
0.19	+11.3%	▲	785.23	150.000
0.14	+2.54%	▲	432.24	120.000

Quelques Exemples

- ▶ On dispose de données cliniques sur 20 ans. Un patient risque-t-il d'avoir une attaque cardiaque dans les 5 prochaines années ?
- ▶ Quel sera le prix de ce stock dans 6 mois ?
- ▶ Es-ce que ce gribouillage est un 7 ?



Quelques Exemples

- ▶ On dispose de données cliniques sur 20 ans. Un patient risque-t-il d'avoir une attaque cardiaque dans les 5 prochaines années ?
- ▶ Quel sera le prix de ce stock dans 6 mois ?
- ▶ Es-ce que ce gribouillage est un 7 ?
- ▶ Est-ce que le mail que je viens juste de recevoir est un spam ?



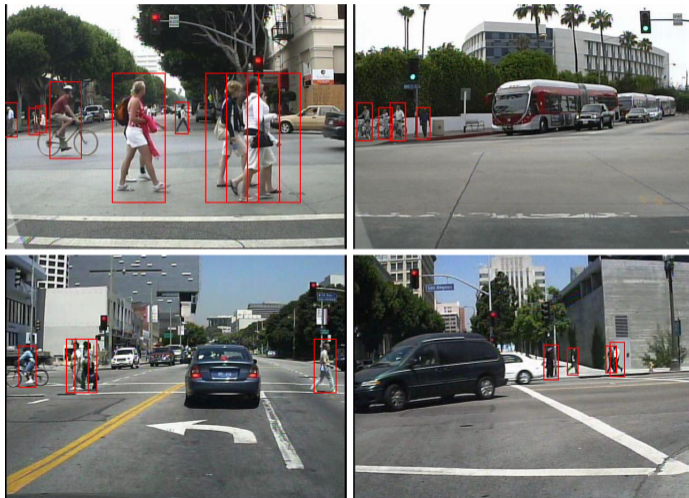
Quelques Exemples

- ▶ On dispose de données cliniques sur 20 ans. Un patient risque-t-il d'avoir une attaque cardiaque dans les 5 prochaines années ?
- ▶ Quel sera le prix de ce stock dans 6 mois ?
- ▶ Es-ce que ce gribouillage est un 7 ?
- ▶ Est-ce que le mail que je viens juste de recevoir est un spam ?
- ▶ Est-ce que je peux regrouper des consommateurs ? Des mots ? Des gènes ?



Et plus récemment ... le *Deep* ...

Reconnaissance vidéo ¹



¹<https://3c1703fe8d.site.internapcdn.net/newman/gfx/news/hires/2016/newalgorithm.jpg>

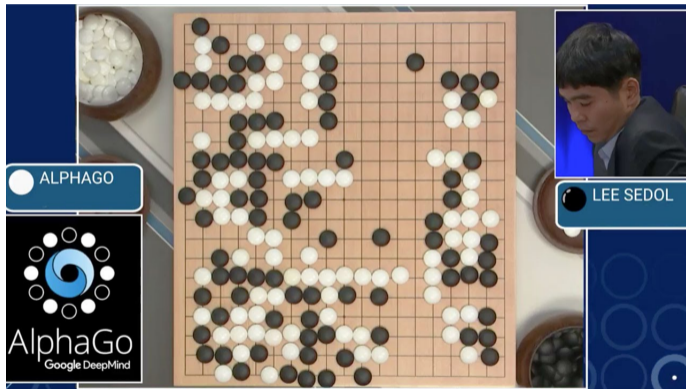
Et plus récemment ... le *Deep* ...

Reconnaissance vocale ²



²<https://technologybux.com/urdu-speech-recognition-no-longer-dream/>

Et plus récemment ... le *Deep* ...



Quelques Exemples

Et beaucoup plus, ...

<https://www.youtube.com/watch?v=wXxrmussq4E>

Compétences nécessaires

Pour faire du Machine Learning, on a besoin de :

Compétences nécessaires

Pour faire du Machine Learning, on a besoin de :

Mathématiques, Algorithmique, Programmation ...

Compétences nécessaires

Pour faire du Machine Learning, on a besoin de :

Mathématiques, Algorithmique, Programmation ...

Sinon, on risque d'utiliser le ML comme ...

Compétences nécessaires

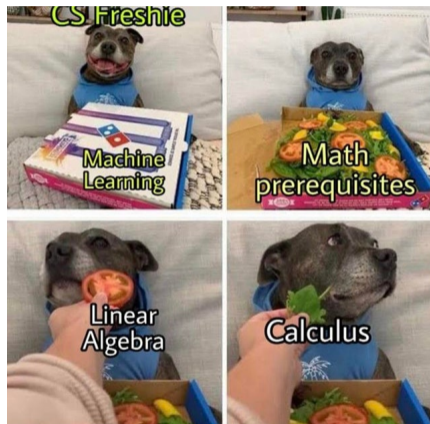
Pour faire du Machine Learning, on a besoin de :

Mathématiques, Algorithmique, Programmation ...

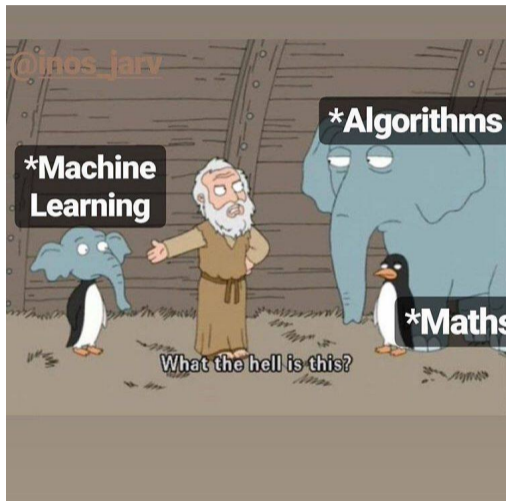
Sinon, on risque d'utiliser le ML comme ...

... une "black box" ...

Compétences nécessaires



Compétences nécessaires



Étapes à suivre pour concevoir un système ML

Étapes à suivre pour concevoir un système ML

1. Formaliser la tâche d'apprentissage.

Étapes à suivre pour concevoir un système ML

1. Formaliser la tâche d'apprentissage.
2. Collecter les données.

Étapes à suivre pour concevoir un système ML

1. Formaliser la tâche d'apprentissage.
2. Collecter les données.
3. Extraire les caractéristiques (features).

Étapes à suivre pour concevoir un système ML

1. Formaliser la tâche d'apprentissage.
2. Collecter les données.
3. Extraire les caractéristiques (features).
4. Choisir la bonne classe de modèles d'apprentissage.

Étapes à suivre pour concevoir un système ML

1. Formaliser la tâche d'apprentissage.
2. Collecter les données.
3. Extraire les caractéristiques (features).
4. Choisir la bonne classe de modèles d'apprentissage.
5. Entraîner le modèle.

Étapes à suivre pour concevoir un système ML

1. Formaliser la tâche d'apprentissage.
2. Collecter les données.
3. Extraire les caractéristiques (features).
4. Choisir la bonne classe de modèles d'apprentissage.
5. Entraîner le modèle.
6. Évaluer le modèle.

Étapes à suivre pour concevoir un système ML

Etape 1. Formaliser la tâche d'apprentissage

- ▶ Définir la tâche que l'on veut traiter avec le système d'apprentissage (exemple : reconnaître des caractères à partir d'images).

Étapes à suivre pour concevoir un système ML

Étape 1. Formaliser la tâche d'apprentissage

- ▶ Définir la tâche que l'on veut traiter avec le système d'apprentissage (exemple : reconnaître des caractères à partir d'images).
- ▶ Un problème d'apprentissage est souvent composé de plusieurs tâches :
 - ▶ Découper l'image en mots et chaque mot en caractères.
 - ▶ Identifier la langue.
 - ▶ Classer chaque caractère dans l'alphabet de la langue.

Étapes à suivre pour concevoir un système ML

Étape 1. Formaliser la tâche d'apprentissage

- ▶ Définir la tâche que l'on veut traiter avec le système d'apprentissage (exemple : reconnaître des caractères à partir d'images).
- ▶ Un problème d'apprentissage est souvent composé de plusieurs tâches :
 - ▶ Découper l'image en mots et chaque mot en caractères.
 - ▶ Identifier la langue.
 - ▶ Classifier chaque caractère dans l'alphabet de la langue.
- ▶ Choisir une mesure de performance pour évaluer le système d'apprentissage (exemple : nombre de caractères mal classés).

Étapes à suivre pour concevoir un système ML

Étape 2. Collecter les données

Étapes à suivre pour concevoir un système ML

Étape 2. Collecter les données

- ▶ Un ensemble d'entraînement doit être collecté dans un format lisible par la machine.

Étapes à suivre pour concevoir un système ML

Étape 2. Collecter les données

- ▶ Un ensemble d'entraînement doit être collecté dans un format lisible par la machine.
- ▶ La collecte de données est une tâche souvent très lourde. Elle implique une intervention manuelle, en particulier pour l'étiquetage des exemples pour l'apprentissage supervisé.

Étapes à suivre pour concevoir un système ML

Étape 3. Extraire les caractéristiques/descripteurs (features)

Voyons en quelques minutes ce qu'est une caractéristique et l'intérêt de l'extraire/la choisir ...

C'est quoi une caractéristique/descripteur ?

C'est quoi une caractéristique/descripteur ?

- ▶ Un attribut caractérisant un objet/une instance.

C'est quoi une caractéristique/descripteur ?

- ▶ Un attribut caractérisant un objet/une instance.
- ▶ Elle peut être quantifiable ou non.

C'est quoi une caractéristique/descripteur ?

- ▶ Un attribut caractérisant un objet/une instance.
- ▶ Elle peut être quantifiable ou non.
- ▶ Il s'agit d'une information qui peut potentiellement être utilisée pour la classification/prediction.

C'est quoi une caractéristique/descripteur ?

- ▶ Un attribut caractérisant un objet/une instance.
- ▶ Elle peut être quantifiable ou non.
- ▶ Il s'agit d'une information qui peut potentiellement être utilisée pour la classification/prediction.



Figure: Classification de fleurs (i.e. Rose ou Lotus).

C'est quoi une caractéristique/descripteur ?

- ▶ Un attribut caractérisant un objet/une instance.
- ▶ Elle peut être quantifiable ou non.
- ▶ Il s'agit d'une information qui peut potentiellement être utilisée pour la classification/prediction.



Figure: Classification de fleurs (i.e. Rose ou Lotus).

- ▶ **Q** : Quelle caractéristique peut-on utiliser pour la classification ?

C'est quoi une caractéristique/descripteur ?

- ▶ Un attribut caractérisant un objet/une instance.
- ▶ Elle peut être quantifiable ou non.
- ▶ Il s'agit d'une information qui peut potentiellement être utilisée pour la classification/prediction.



Figure: Classification de fleurs (i.e. Rose ou Lotus).

- ▶ **Q** : Quelle caractéristique peut-on utiliser pour la classification ?
- ▶ **Rep.** : La couleur (si rouge alors "Rose" sinon "Lotus").

C'est quoi une caractéristique ?



C'est quoi une caractéristique ?



C'est quoi une caractéristique ?



Figure: Utiliser la couleur seule ferait que le système verra une rose au lieu de la pomme et un lotus au lieu de l'ours polaire

C'est quoi une caractéristique ?



Figure: Utiliser la couleur seule ferait que le système verra une rose au lieu de la pomme et un lotus au lieu de l'ours polaire

Note

- ▶ Une seule caractéristique ne peut pas être suffisante pour la classification. Il faut utiliser un **ensemble** de caractéristiques.

C'est quoi une caractéristique ?



Figure: Utiliser la couleur seule ferait que le système verra une rose au lieu de la pomme et un lotus au lieu de l'ours polaire

Note

- ▶ Une seule caractéristique ne peut pas être suffisante pour la classification. Il faut utiliser un **ensemble** de caractéristiques.
- ▶ Un ensemble de caractéristiques forme un vecteur.

C'est quoi une caractéristique ?



Figure: Utiliser la couleur seule ferait que le système verra une rose au lieu de la pomme et un lotus au lieu de l'ours polaire

Note

- ▶ Une seule caractéristique ne peut pas être suffisante pour la classification. Il faut utiliser un **ensemble** de caractéristiques.
- ▶ Un ensemble de caractéristiques forme un vecteur.
- ▶ La dimension du vecteur et le nombre de caractéristiques.

C'est quoi une caractéristique ?



Figure: Utiliser la couleur seule ferait que le système verra une rose au lieu de la pomme et un lotus au lieu de l'ours polaire

Note

- ▶ Une seule caractéristique ne peut pas être suffisante pour la classification. Il faut utiliser un **ensemble** de caractéristiques.
- ▶ Un ensemble de caractéristiques forme un vecteur.
- ▶ La dimension du vecteur et le nombre de caractéristiques.
- ▶ **Un vecteur doit être discriminant.**

Étapes à suivre pour concevoir un système ML

Étape 3. Extraire les features

Étapes à suivre pour concevoir un système ML

Étape 3. Extraire les features

- ▶ On doit extraire un ensemble de features des données, ce seront les entrées au système d'apprentissage.

Étapes à suivre pour concevoir un système ML

Étape 3. Extraire les features

- ▶ On doit extraire un ensemble de features des données, ce seront les entrées au système d'apprentissage.
- ▶ Il est nécessaire d'avoir des **connaissances "métier"** pour savoir choisir les features appropriés à la tâche d'apprentissage considérée.

Étapes à suivre pour concevoir un système ML

Étape 3. Extraire les features

- ▶ On doit extraire un ensemble de features des données, ce seront les entrées au système d'apprentissage.
- ▶ Il est nécessaire d'avoir des **connaissances "métier"** pour savoir choisir les features appropriés à la tâche d'apprentissage considérée.
- ▶ Si on choisit trop peu de features, on peut passer à côté d'informations importantes pour la classification.

Étapes à suivre pour concevoir un système ML

Étape 3. Extraire les features

- ▶ On doit extraire un ensemble de features des données, ce seront les entrées au système d'apprentissage.
- ▶ Il est nécessaire d'avoir des **connaissances "métier"** pour savoir choisir les features appropriés à la tâche d'apprentissage considérée.
- ▶ Si on choisit trop peu de features, on peut passer à côté d'informations importantes pour la classification.
- ▶ Si on en sélectionne trop, il faut disposer de beaucoup de données pour l'entraînement.

Étapes à suivre pour concevoir un système ML

Étape 4. Choisir la bonne catégorie de modèles d'apprentissage

Étapes à suivre pour concevoir un système ML

Étape 4. Choisir la bonne catégorie de modèles d'apprentissage

- ▶ Un modèle simple comme un classifieur linéaire peut être facile à entraîner mais peut s'avérer inefficace pour des données non linéairement séparables.

Étapes à suivre pour concevoir un système ML

Étape 4. Choisir la bonne catégorie de modèles d'apprentissage

- ▶ Un modèle simple comme un classifieur linéaire peut être facile à entraîner mais peut s'avérer inefficace pour des données non linéairement séparables.
- ▶ Un modèle trop complexe peut mener à du **sur-apprentissage (overfitting)** et peut ne pas se généraliser pour de nouveaux exemples.

Étapes à suivre pour concevoir un système ML

Étape 4. Choisir la bonne catégorie de modèles d'apprentissage

- ▶ Un modèle simple comme un classifieur linéaire peut être facile à entraîner mais peut s'avérer inefficace pour des données non linéairement séparables.
- ▶ Un modèle trop complexe peut mener à du **sur-apprentissage** (**overfitting**) et peut ne pas se généraliser pour de nouveaux exemples.
- ▶ Exemple : régression **linéaire** vs régression **polynomiale** ...

Étapes à suivre pour concevoir un système ML

Étape 5. Entraîner le modèle

Étapes à suivre pour concevoir un système ML

Étape 5. Entraîner le modèle

- ▶ Chercher dans l'espace des modèles possible (en fonction de la classe du modèle recherché).

Étapes à suivre pour concevoir un système ML

Étape 5. Entraîner le modèle

- ▶ Chercher dans l'espace des modèles possible (en fonction de la classe du modèle recherché).
- ▶ → Adapter le modèle aux exemples disponibles en s'assurant que la performance du modèle est bonne.

Étapes à suivre pour concevoir un système ML

Étape 5. Entraîner le modèle

- ▶ Chercher dans l'espace des modèles possible (en fonction de la classe du modèle recherché).
- ▶ → Adapter le modèle aux exemples disponibles en s'assurant que la performance du modèle est bonne.
- ▶ La modèle entraîné doit être performant sur des données qu'il n'a jamais vues (**généralisation**). S'assurer que le modèle n'apprend pas par coeur les exemples qui lui ont été présentés (overfitting).

Étapes à suivre pour concevoir un système ML

Étape 6. Evaluer le modèle

Étapes à suivre pour concevoir un système ML

Étape 6. Evaluer le modèle

- ▶ Le modèle est évalué sur sa capacité à généraliser à des exemples qu'il n'a jamais vus.

Étapes à suivre pour concevoir un système ML

Étape 6. Evaluer le modèle

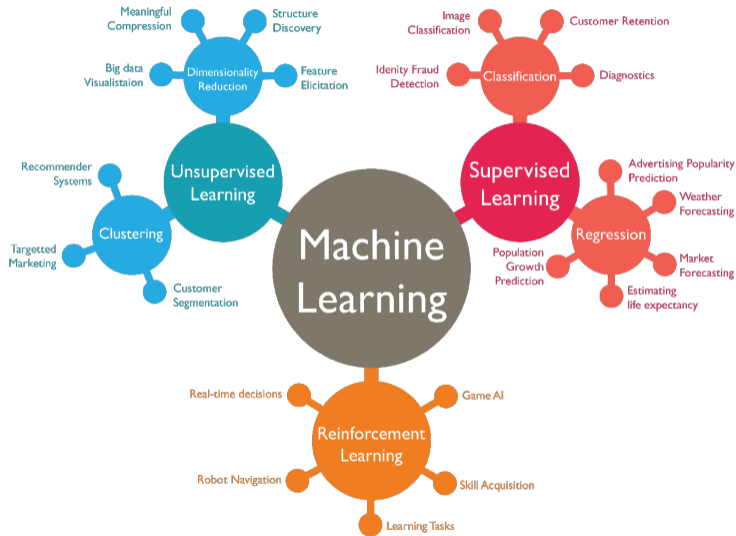
- ▶ Le modèle est évalué sur sa capacité à généraliser à des exemples qu'il n'a jamais vus.
- ▶ L'évaluation peut donner des indications sur les faiblesses du modèle et peut suggérer des modifications et des raffinements possibles ...

Étapes à suivre pour concevoir un système ML

Étape 6. Evaluer le modèle

- ▶ Le modèle est évalué sur sa capacité à généraliser à des exemples qu'il n'a jamais vus.
- ▶ L'évaluation peut donner des indications sur les faiblesses du modèle et peut suggérer des modifications et des raffinements possibles ...
- ▶ Mais attention au piège dans lequel tombent la plupart des concepteurs de solutions ML : **ne pas séparer le test de la validation** (cela mérite quelques explications ...)

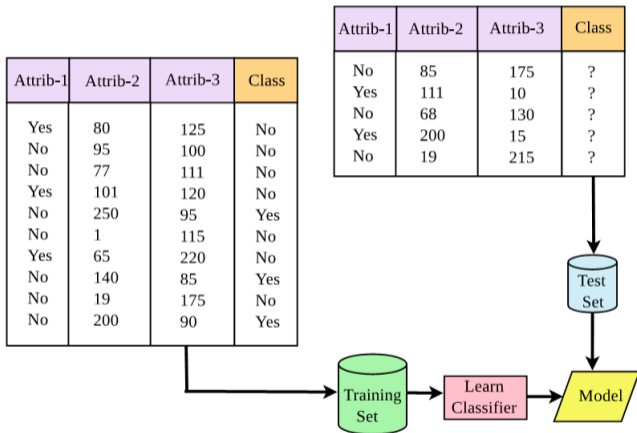
Modèles d'apprentissage



Modèles d'apprentissage

- ▶ **Apprentissage supervisé** :
 - ▶ **Supervision** : les données d'apprentissage (observations) sont accompagnés par les labels indiquant leurs classes.
 - ▶ Les nouvelles données sont classifiées en se basant sur le training set.

Apprentissage supervisé



Modèles d'apprentissage

- ▶ Apprentissage non supervisé :
 - ▶ Le label de classe des éléments observés (training set) **n'est pas connu**.
 - ▶ Le but est de déceler l'existence de classes ou groupes dans les données.

Apprentissage non supervisé

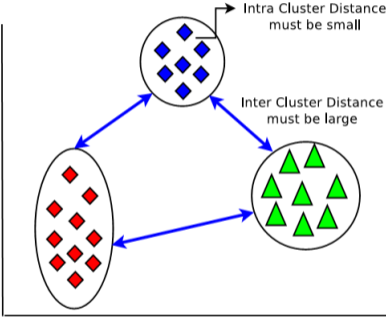


Figure: Unsupervised Learning

Apprentissage supervisé vs non supervisé

Plus formellement :

Task	Data: based on...	Target: learn...
▶ Supervized	$\mathcal{T} = \{(x_i, y_i)\}_{i=1..n}$	$f(x) = y$
▶ Unsupervised	$\mathcal{T} = \{x_i\}_{i=1..n}$	$x \in X_k$

Apprentissage supervisé

Classification.

Apprentissage supervisé

Classification.

- ▶ **Binaire** : affecter chaque exemple à une des deux classes possibles (chien ou chat, 1 ou -1, etc).

Classification.

- ▶ **Binaire** : affecter chaque exemple à une des deux classes possibles (chien ou chat, 1 ou -1, etc).
- ▶ **Multiclasse** : affecter chaque exemple à une classe parmi $n > 2$ classes possibles (affecter un gribouillage à un des chiffres de l'ensemble $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$).

Apprentissage supervisé

Classification.

- ▶ **Binaire** : affecter chaque exemple à une des deux classes possibles (chien ou chat, 1 ou -1, etc).
- ▶ **Multiclasse** : affecter chaque exemple à une classe parmi $n > 2$ classes possibles (affecter un gribouillage à un des chiffres de l'ensemble $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$).

Régression.

- ▶ Affecter une valeur **réelle** à chaque exemple (prix d'un stock, ...).

Apprentissage non supervisé

Apprentissage non supervisé

Réduction de dimension.

- ▶ Réduire la dimension des données tout en minimisant la perte d'information (exemple : Analyse en Composante Principale (ACP)).

Apprentissage non supervisé

Réduction de dimension.

- ▶ Réduire la dimension des données tout en minimisant la perte d'information (exemple : Analyse en Composante Principale (ACP)).

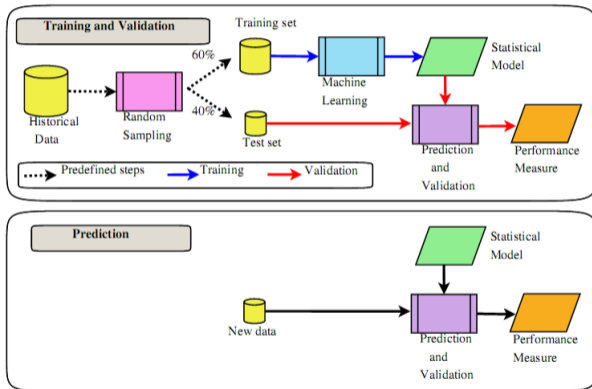
Clustering.

- ▶ Regrouper les données en groupes homogènes par rapport à une mesure de similarité.

On se limite ici (dans ce cours) à l'apprentissage supervisé.

Evaluation d'un modèle d'apprentissage (classifieur)

Méthode générale :



Evaluation d'un modèle d'apprentissage (classifieur)

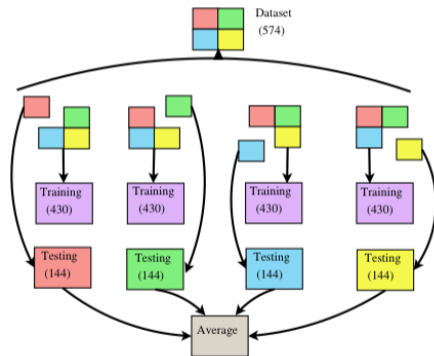
- ▶ Si l'ensemble de données est suffisamment grand :
Découper le dataset en deux sous-ensemble : **Training Set** ($\sim 80\%$) et **Test Set** ($\sim 20\%$).

Evaluation d'un modèle d'apprentissage (classifieur)

- ▶ Si l'ensemble de données est suffisamment grand :
Découper le dataset en deux sous-ensemble : **Training Set** ($\sim 80\%$) et **Test Set** ($\sim 20\%$).
- ▶ Sinon, utiliser la **cross-validation**.

Evaluation d'un modèle d'apprentissage (classifieur)

Cross-validation :



Evaluation d'un modèle d'apprentissage (classifieur)

Matrice de confusion (cas binaire) :

	Actual class (Observation)	
Predicted class (Expectation)	TP	FP
	FN	TN

Table: Paramètres d'évaluation.

- ▶ TP : True Positive
- ▶ FP : False Positive
- ▶ FN : False Negative
- ▶ TN : True Negative

Métriques d'évaluation

Classification binaire :

Accuracy

Il s'agit de la fraction des exemples bien classés par rapport à toutes les prédictions :

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

Métriques d'évaluation

Classification binaire :

Accuracy

Il s'agit de la fraction des exemples bien classés par rapport à toutes les prédictions :

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

Problème

- ▶ Si les deux classes sont fortement déséquilibrées, alors cette mesure n'apporte pas d'information pertinente ...

Métriques d'évaluation

Classification binaire :

Accuracy

Il s'agit de la fraction des exemples bien classés par rapport à toutes les prédictions :

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

Problème

- ▶ Si les deux classes sont fortement déséquilibrées, alors cette mesure n'apporte pas d'information pertinente ...
 - ▶ Les prédictions sont dominées par la classe dominante
 - ▶ Mettre tous les exemples dans la classe dominante permet d'obtenir une grande accuracy

Métriques d'évaluation

Classification binaire :

Accuracy

Il s'agit de la fraction des exemples bien classés par rapport à toutes les prédictions :

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

Problème

- ▶ Si les deux classes sont fortement déséquilibrées, alors cette mesure n'apporte pas d'information pertinente ...
 - ▶ Les prédictions sont dominées par la classe dominante
 - ▶ Mettre tous les exemples dans la classe dominante permet d'obtenir une grande accuracy
- ▶ Une solution possible : rééquilibrer les coûts (e.g. chaque élément de la classe positive compte pour N/P où $N=TN+FP$ et $P=TP+FN$)

Métriques d'évaluation

Classification binaire :

Precision

C'est la fraction des vrais positifs parmi les exemples prédits comme positifs.

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

Métriques d'évaluation

Classification binaire :

Recall ou Sensibilité

Il s'agit du rapport entre le nombre d'exemples de la classe c correctement prédits et le nombre d'exemples appartenant à la classe c .

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

Métriques d'évaluation

Classification binaire :

F-measure

$$F_{\beta} = \frac{(1 + \beta^2)(Pre * Rec)}{\beta^2 Pre + Rec}$$

- ▶ La précision et le recall sont complémentaires, augmenter la précision diminue le recall.
- ▶ *F*-measure combine les deux mesures en équilibrant les deux.
- ▶ β est un paramètre de compromis entre la précision et le recall.

Métriques d'évaluation

Classification binaire :

F_1

$$F_{\beta} = \frac{2(Pre * Rec)}{Pre + Rec}$$

- ▶ Il s'agit de la F -mesure pour $\beta = 1$.
- ▶ C'est la moyenne harmonique de la précision et du recall.

Métriques d'évaluation

Classification multiclass :

$T \setminus P$	y_1	y_2	\dots	y_k
y_1	n_{11}	n_{12}	\dots	n_{1k}
y_2	n_{21}	n_{22}	\dots	n_{2k}
\vdots	\vdots	\vdots	\dots	\vdots
y_k	n_{k1}	n_{k2}	\dots	n_{kk}

- ▶ Il s'agit d'une généralisation de la classification binaire.
- ▶ n_{ij} est le nombre d'exemples de la classe y_i prédits comme y_j .
- ▶ La diagonale contient les TP de chaque classe.
- ▶ La somme des éléments d'une colonne donne les FP pour la classe de la colonne.
- ▶ La somme des éléments d'une ligne donne les FN pour la classe de la ligne.

$$FP_i = \sum_{j \neq i} n_{ji} \quad FN_i = \sum_{j \neq i} n_{ij}$$