

# Intelligence Artificielle

Pr. Hiba Chougrad  
Année-universitaire: 2023-2024

# Plan

1. Introduction générale et Agents Intelligents
2. Logique du premier ordre
3. Machine Learning : Pré-traitement des données
4. Machine Learning : Supervised vs Unsupervised
5. Machine Learning : Construire un bon modèle
6. Machine Learning : Raisonnement probabiliste et réseaux bayésiens
7. Machine Learning: Algorithmes d'apprentissage automatique
8. Machine Learning: Apprentissage par renforcement, vision par ordinateur, NLP, Deep Learning

# Plan

1. Introduction générale et Agents Intelligents
2. Logique du premier ordre
3. Machine Learning : Pré-traitement des données
4. **Machine Learning : Supervised vs Unsupervised**
5. Machine Learning : Construire un bon modèle
6. Machine Learning : Raisonnement probabiliste et réseaux bayésiens
7. Machine Learning: Algorithmes d'apprentissage automatique
8. Machine Learning: Apprentissage par renforcement, vision par ordinateur, NLP, Deep Learning

# Machine Learning : Supervised vs Unsupervised

---

# L'exploration des données

- Les données bien propres peuvent maintenant commencer à être explorées. Cette étape vous permet de mieux comprendre les différents comportements et de bien saisir le phénomène sous-jacent.
- N'hésitez pas à afficher toutes sortes de graphiques, confronter les différentes variables les unes aux autres, tester des hypothèses de corrélation, etc.
- À la fin de l'exploration, vous devrez être en mesure de :
  - ✓ **Proposer plusieurs hypothèses sur les causes sous-jacentes à la génération de la dataset :**  
« suite à l'exploration, il y a clairement une relation entre X et Y »
  - ✓ **Proposer plusieurs pistes de modélisation statistique possible des données** qui vont permettre de résoudre la problématique de départ considérée
  - ✓ **Proposer si nécessaire de nouvelles sources de données** qui aideraient à mieux comprendre le phénomène

# L'exploration des données

- Après **le nettoyage et le prétraitement des données**, on peut enfin rentrer dans la partie la plus intéressante du métier, c'est à dire **la création du modèle statistique associé aux données** qui nous intéressent ! C'est ce qu'on appelle **Machine Learning** (ou apprentissage automatique).
- En **machine learning** et en data science plus généralement, l'objectif est de **trouver un modèle** (stochastique ou déterministe) du phénomène **à l'origine** des données. C'est à dire qu'on considère que **chaque donnée observée est l'expression d'une variable aléatoire générée par une distribution de probabilité**.

# Rappel: La notion de Modèle

- Une représentation simplifiée de la réalité

# Rappel: La notion de Modèle

- Une représentation simplifiée de la réalité
  - Une équation

$$\frac{dN(t)}{dt} = r_0 * N(t) * \left(1 - \frac{N(t)}{K}\right)$$

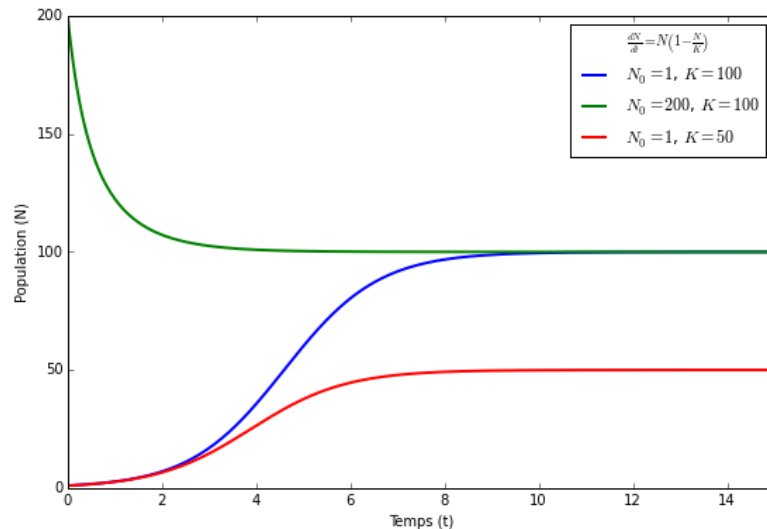
On note :

- **N(t)** le nombre d'individus dans une population donnée à un temps t.
- **r=λ-μ** en notant λ le taux de naissance et μ le taux de mortalité (c'est-à-dire le nombre de naissance -respectivement de morts- par individus lors de Δt
- **K** la capacité biotique (la taille maximale de la population qu'un milieu donné peut supporter)



# Rappel: La notion de Modèle

- Une représentation simplifiée de la réalité
  - Une équation

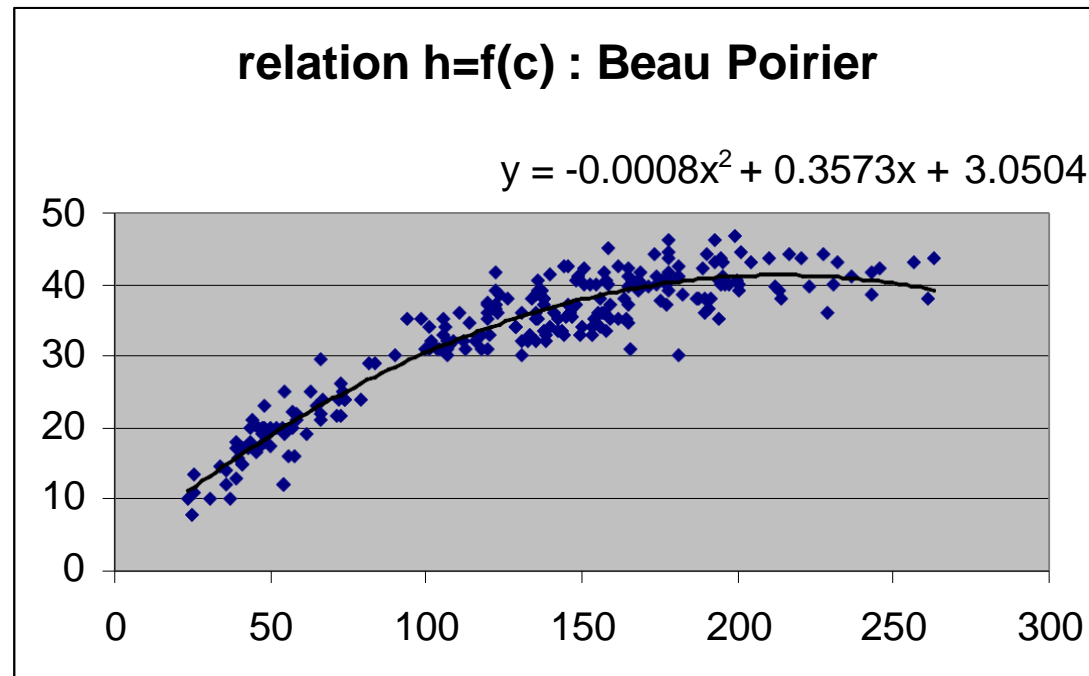


$$\frac{dN(t)}{dt} = r_0 * N(t) * \left(1 - \frac{N(t)}{K}\right)$$

Le taux de croissance de la population **sature** avec le nombre d'individu, ainsi la population ne peut pas atteindre une taille infinie: elle est limitée par une taille maximale appelée **capacité biotique**. La valeur de la capacité biotique dépend de l'environnement dans lequel évolue la population, elle est généralement liée à un épuisement des ressources.

# Rappel: La notion de Modèle

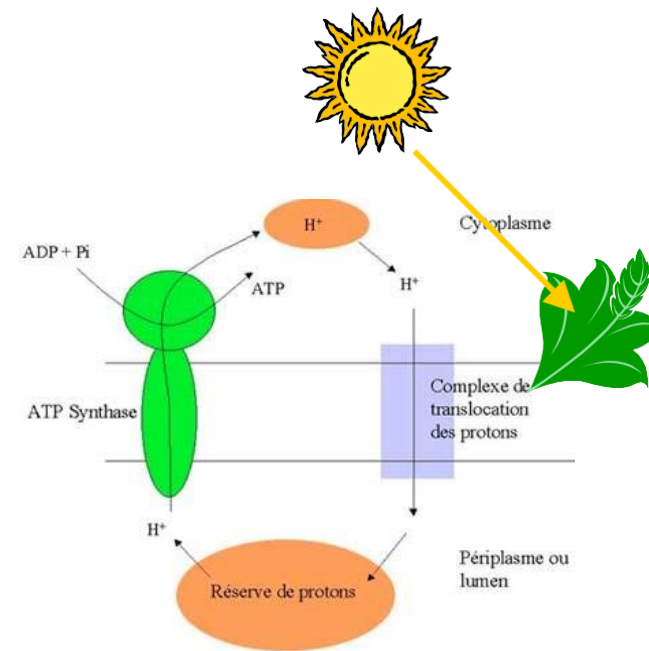
- Une représentation simplifiée de la réalité
  - Une équation
  - Une courbe ajustée aux données



# Rappel: La notion de Modèle

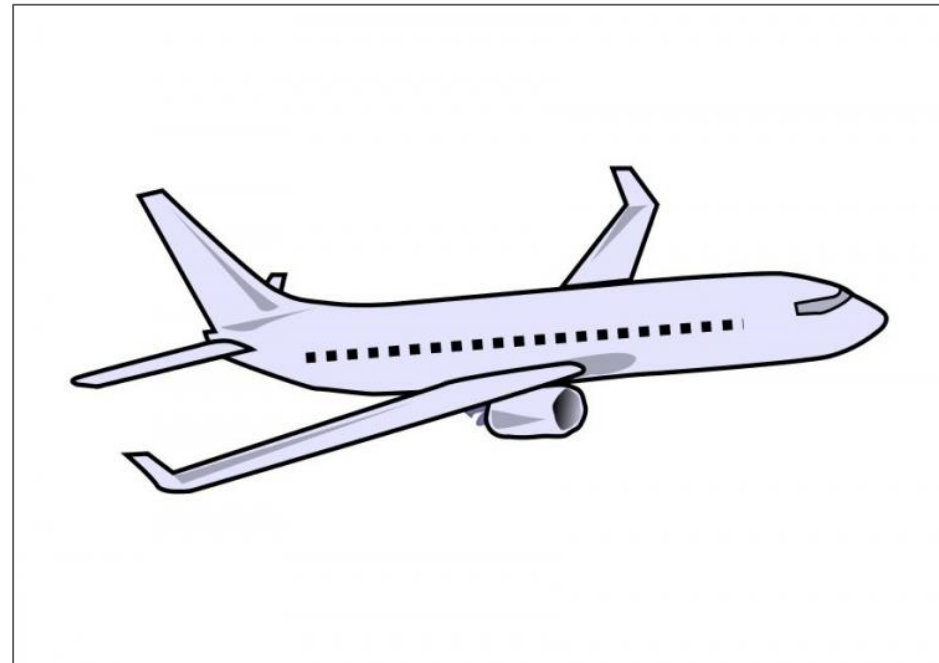
- Une représentation simplifiée de la réalité
  - Une équation
  - Une courbe ajustée aux données
  - Un schéma

**La photosynthèse qui permet à des organismes synthétiser de la matière organique en utilisant l'énergie lumineuse.**



# Rappel: La notion de Modèle

- Une représentation simplifiée de la réalité
  - Une équation
  - Une courbe ajustée aux données
  - Un schéma
  - Un dessin

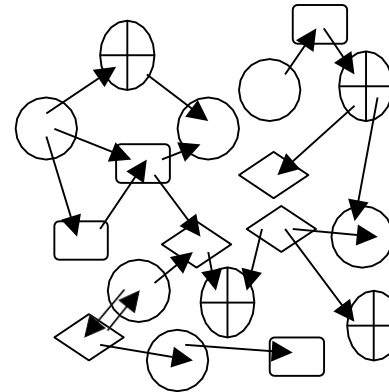


# Rappel: La notion de Modèle

- Un modèle répond à un objectif donné

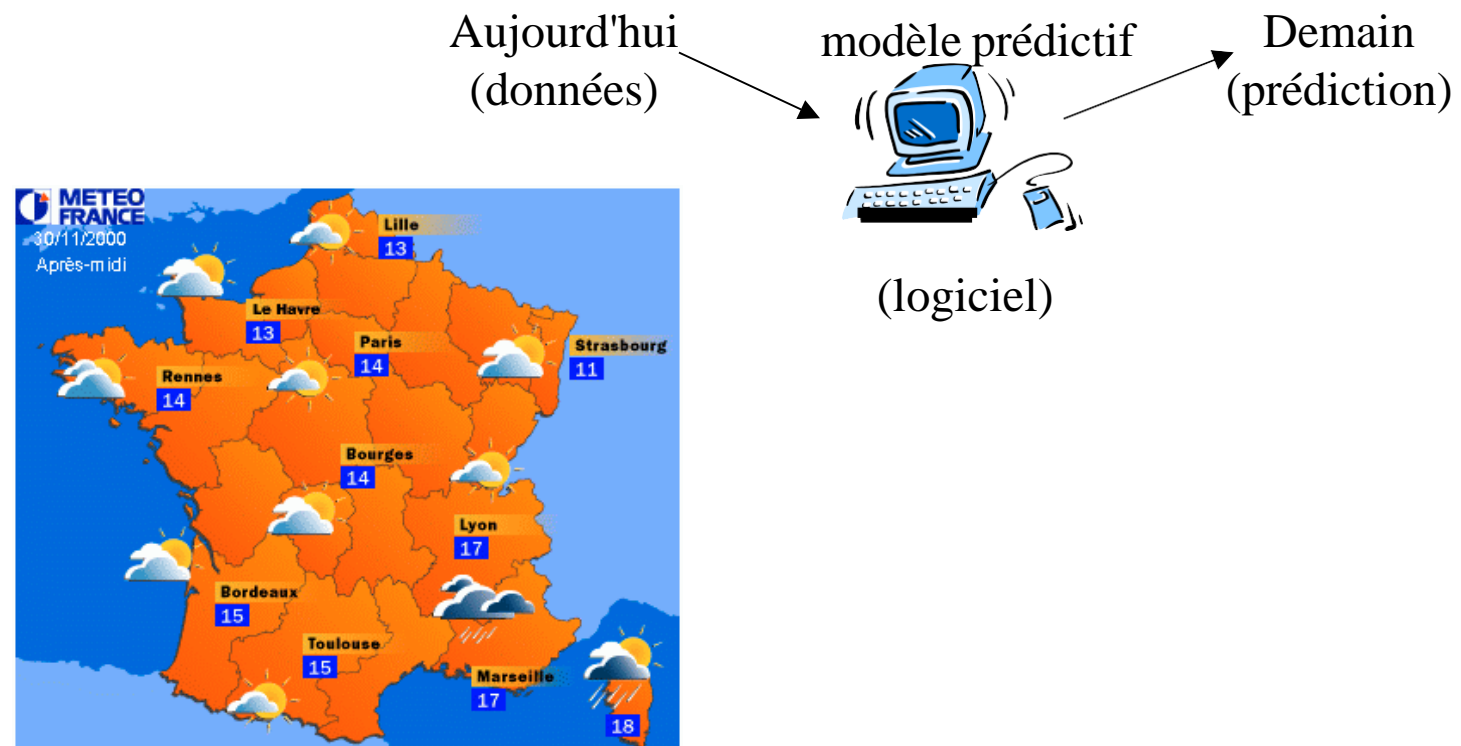
# Rappel: La notion de Modèle

- Un modèle répond à un objectif donné
  - Comprendre
    - Faire une synthèse des connaissances
    - Tester des hypothèses



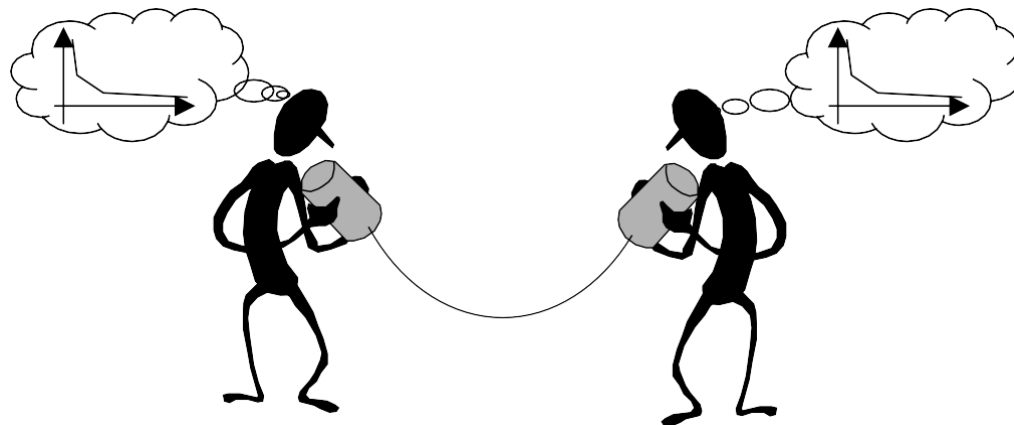
# Rappel: La notion de Modèle

- Un modèle répond à un objectif donné
  - Comprendre
  - **Prédire**



# Rappel: La notion de Modèle

- Un modèle répond à un objectif donné
  - Comprendre
  - Prédire
  - Communiquer
    - Entre scientifiques
    - Lors de négociations





# Rappel: Déterministe vs Stochastique

- **Processus Déterministe:**

L'état présent du système est entièrement déterminé par l'état précédent. Une même cause produit le même effet.

- **Processus stochastique:**

L'état présent du système est aléatoire et ne dépend pas de l'état précédent. Si on note  $\mathbf{x}(t)$  l'état du système au temps  $t$ . Le processus est dit stochastique, lorsque d'une manière consécutive  $(t_1, t_2, t_3, \dots, t_n)$  produit des états  $\mathbf{x}(t_i)$  aléatoires.  $\mathbf{x}(t_i)$  est une réalisation d'une variable aléatoire  $\mathbf{X}(t)$ .

# Rappel: Déterministe vs Stochastique

- $Y$  étant la variable d'état (Effectif de la population correspondant à la variable modélisée)
- $t$  est le temps
- $a$  est un paramètre

- **Processus déterministe**

Le paramètre  $a$  est constant et correspond au taux d'accroissement relatif de la population.

$$Y(t) = Y_0 e^{at}$$

- **Processus stochastique**

Le paramètre  $a$  n'est pas constant mais prend des valeurs  $a_i(t_i)$  selon une loi uniforme dans un ensemble de valeurs  $\{a_i\}$ .

$$Y(t) = Y_0 e^{(a \in \{a_i\})t}$$

# Rappel: Déterministe vs Stochastique

- **Un modèle déterministe**
  - partant d'un état initial, un seul état final
  - parfaitement connu
  - aucun phénomène aléatoire

# Rappel: Déterministe vs Stochastique

- Un modèle déterministe
  - partant d'un état initial, un seul état final
  - parfaitement connu
  - aucun phénomène aléatoire
- **Un modèle stochastique**
  - intègre une part d'aléatoire

=> Pourquoi ? Comment ?

# Exemple

- Dans la réalité, il y a de l'aléatoire !
  - une panne de réveil
  - un retard de train
  - être écrasé par un camion
  - gagner au loto !

Heureusement sinon la vie serait triste !

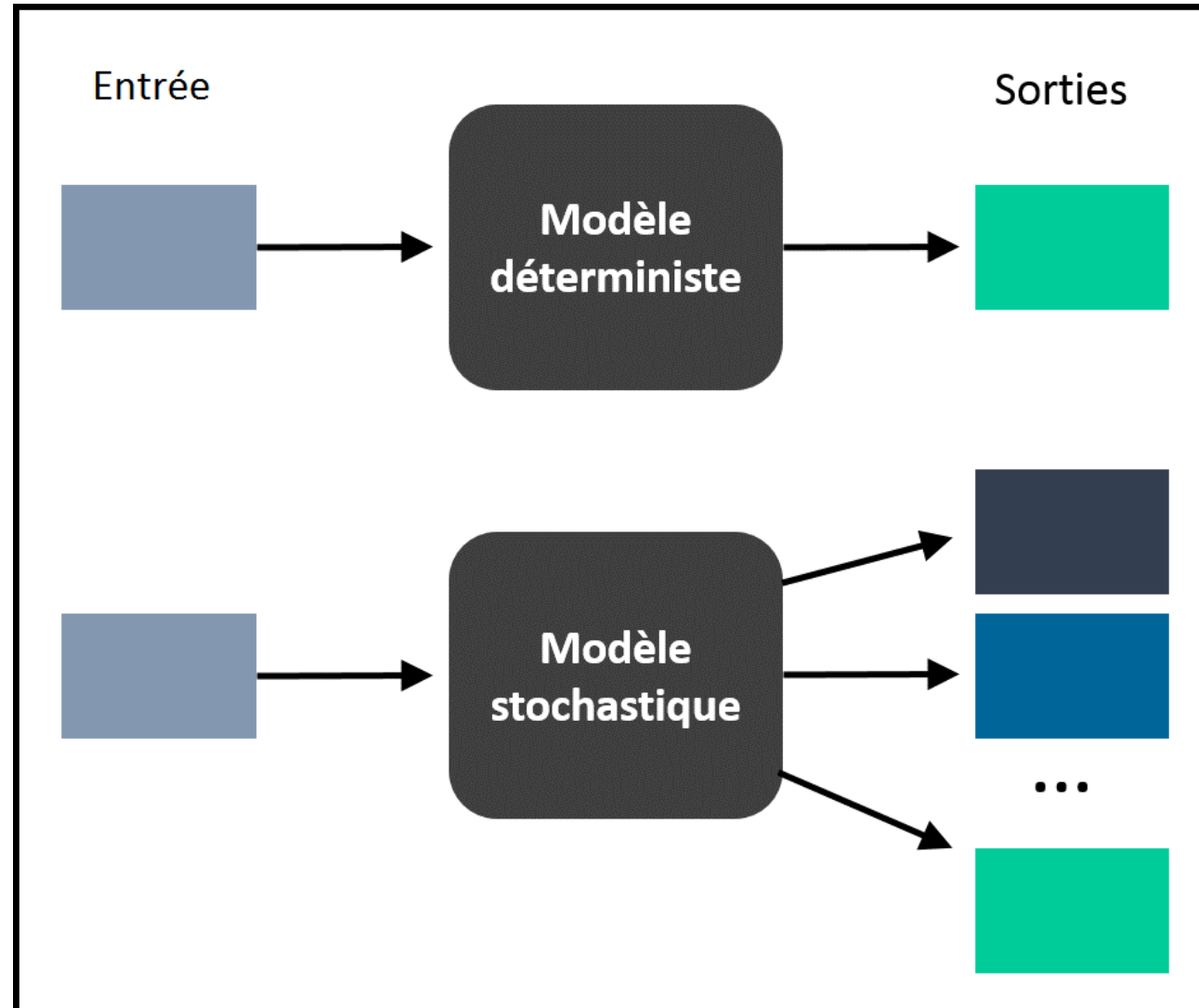


**La roulette  
(jeu d'hasard)**

# Rappel: La notion de stochasticité

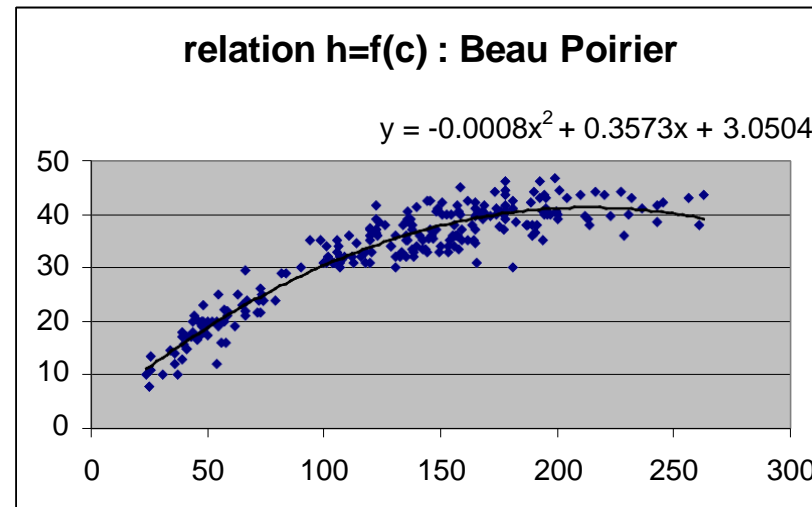
- Dans la réalité, il y a de l'aléatoire !
  - partant d'un état initial, **plusieurs états finaux**
  - **connaissance imparfaite**
  - il y a des **phénomènes aléatoire**
- **On parle de stochasticité**
  - pour tout ce qui échappe au déterminisme

# Rappel: Déterministe vs Stochastique



# Rappel: La notion de stochasticité

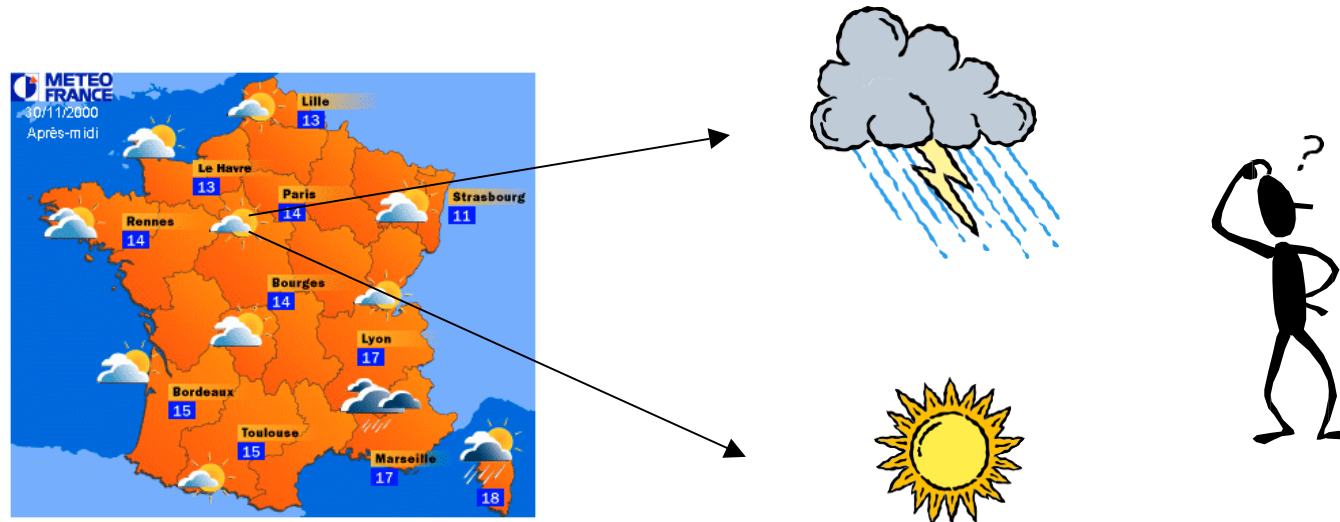
- L'aléatoire peut jouer un rôle important !
  - une variabilité dans les mesures





# Rappel: La notion de stochasticité

- L'aléatoire peut jouer un rôle important !
  - Une variabilité dans les mesures
  - Plusieurs avenir possibles : s'y préparer !
    - Comment calculer un risque ?



# Rappel: La notion de stochasticité

- L'aléatoire peut jouer un rôle important !
  - Une variabilité dans les mesures
  - Plusieurs avenir possibles : s'y préparer !
  - Des effets catastrophiques
    - Effet papillon
    - Extinction / survie d'espèces



# Rappel: La notion de stochasticité

- L'aléatoire peut jouer un rôle important !
  - Une variabilité dans les mesures
  - Plusieurs avenir possibles : s'y préparer !
  - Des effets catastrophiques
- Donc on veut pouvoir le reproduire / estimer
  - Des modèles stochastiques

# Rappel: Le modèle stochastique

- Un modèle stochastique
  - Prend en compte les phénomènes aléatoires
  - Fait intervenir le hasard (tirage aléatoire)
  - Partant d'un état initial, plusieurs états finaux



# Rappel: Le modèle stochastique

- Différentes sources de stochastité
  - Erreurs de mesures
  - Phénomènes inconnus, non modélisés
    - Interactions au sein d'un système complexe
    - Etat initial inexact
    - Fluctuation environnement
    - Variations individuelles (génétique, ...)

# Machine Learning: Objectif

- En Machine Learning l'objectif est de **trouver un modèle** (stochastique ou déterministe) du phénomène à l'origine des données.
- C'est à dire qu'on considère que **chaque donnée observée est l'expression d'une variable aléatoire générée par une distribution de probabilité.**

# Exemple

- Partons d'un exemple:

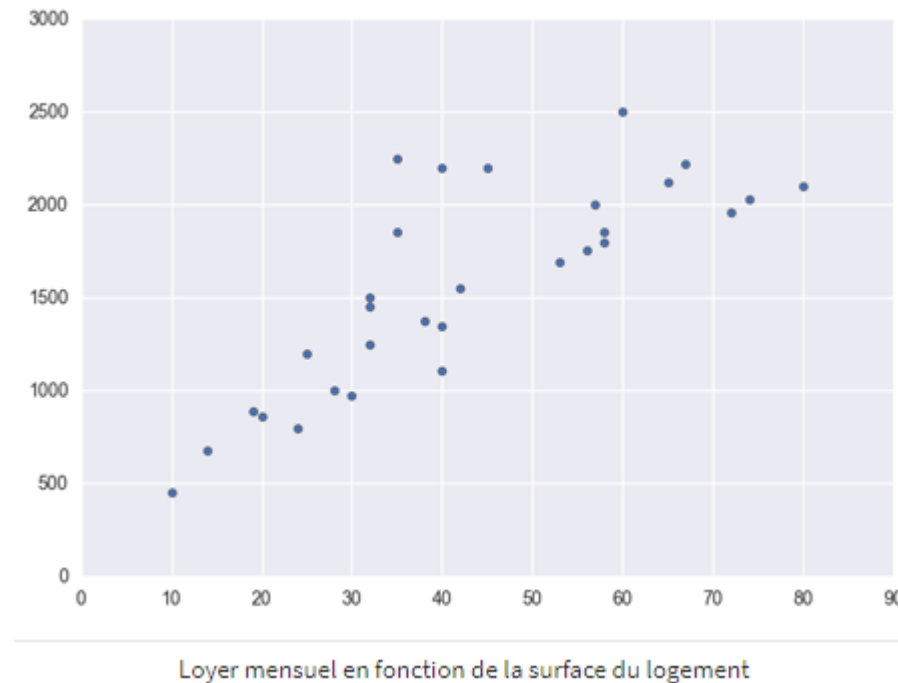
Imaginez que vous voulez savoir si vous payez trop cher votre loyer. Vous avez récupéré sur un site de location une trentaine de prix des locations disponibles, ainsi que la surface associée :

loyer mensuel (en €)	surface (en $m^2$ )
1500	32
2120	65
2500	60
...	...

# Exemple

- Partons d'un exemple:

Si on affiche maintenant ces différents points sur un graphe qui représente le montant du loyer en fonction de la surface, on obtient la figure suivante :

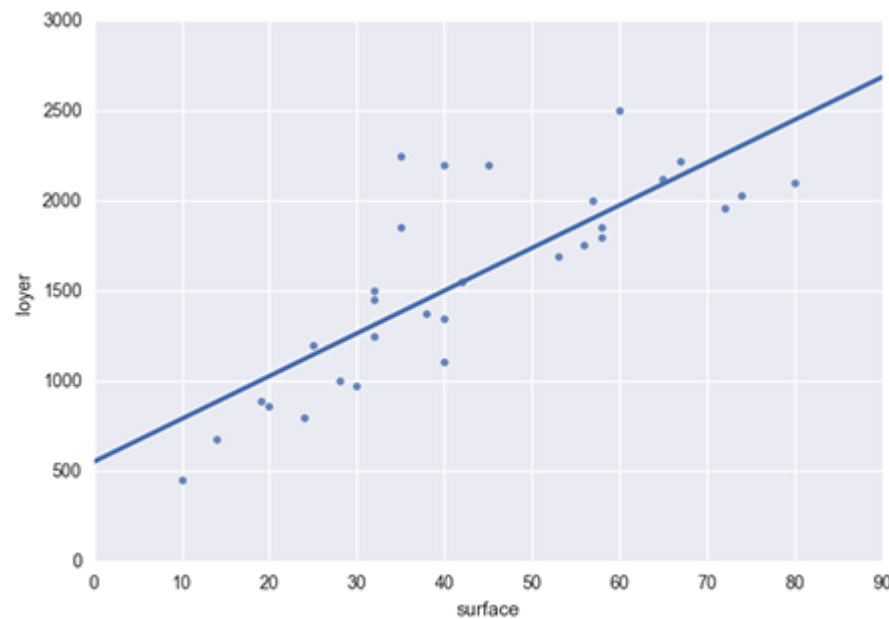




# Exemple

- Partons d'un exemple:

Comme on pouvait s'y attendre, on remarque une augmentation relativement **linéaire** du loyer par rapport à la surface de l'appartement. Une première **modélisation** simple du phénomène (le prix du loyer) serait donc simplement de considérer la droite la plus "proche" de l'ensemble des points.

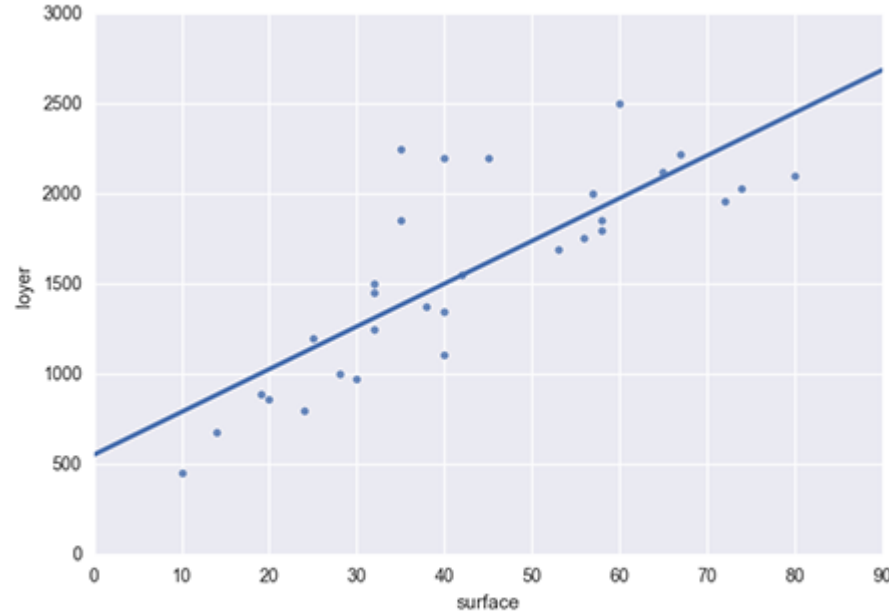


La droite de régression correspondant à la modélisation du nuage de points

# Exemple

- Partons d'un exemple:

Comme on pouvait s'y attendre, on remarque une augmentation relativement **linéaire** du loyer par rapport à la surface de l'appartement. Une première **modélisation** simple du phénomène (le prix du loyer) serait donc simplement de considérer la droite la plus "proche" de l'ensemble des points.

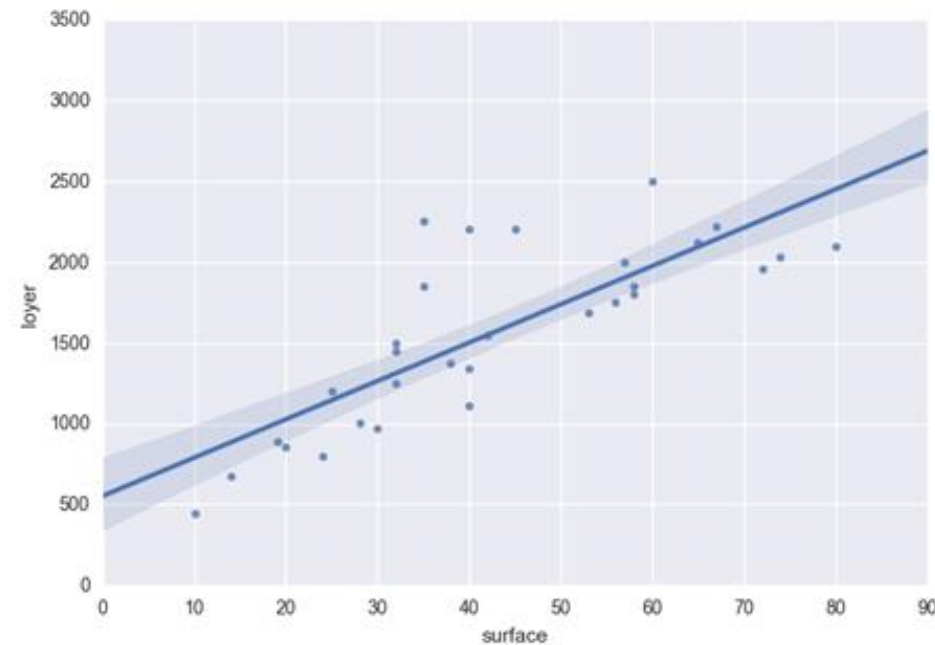


La droite de régression correspondant à la modélisation du nuage de points

# Exemple

- Partons d'un exemple:

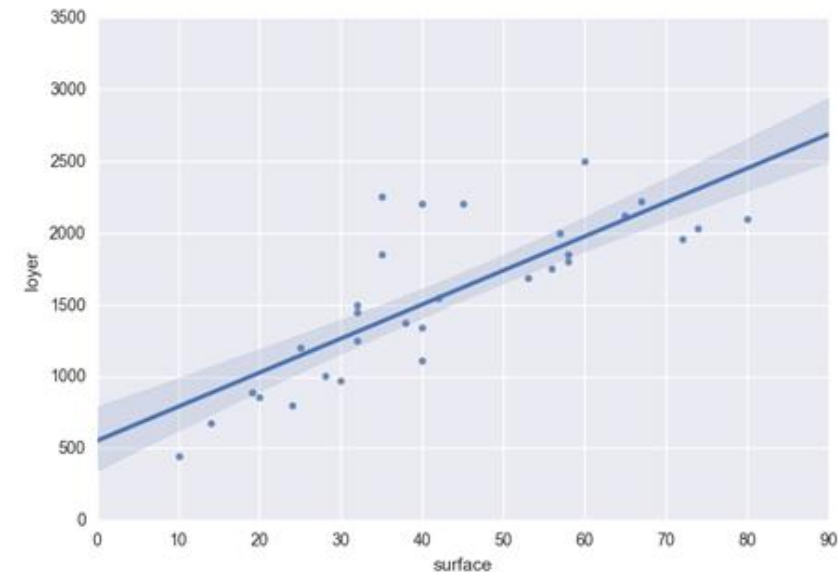
La droite représente donc notre **modèle** du phénomène, auquel nous pouvons ajouter l'intervalle de confiance dans lequel on pense que se trouve la droite.



L'intervalle de confiance (à 90%)

# Exemple

- Pour résumer:
- La **modélisation** consiste à trouver le bon modèle statistique (ici la droite et son intervalle de confiance) qui *colle le mieux aux données d'exemple*. Le **Machine Learning** en particulier intervient pour trouver ce modèle de manière **automatisée**.



L'intervalle de confiance (à 90%)

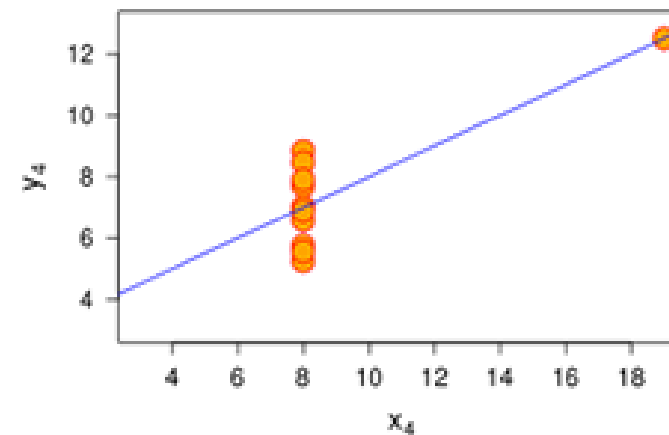
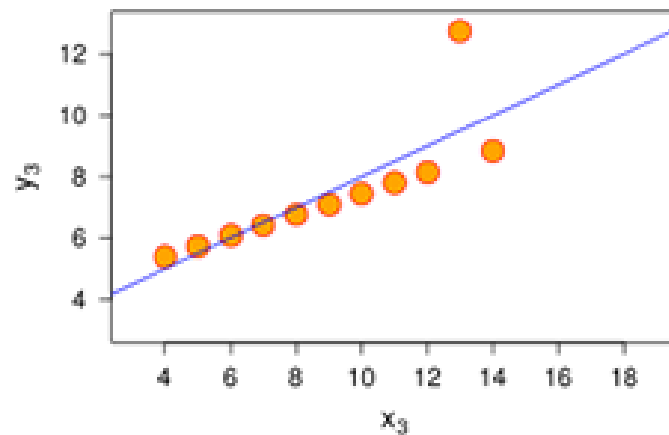
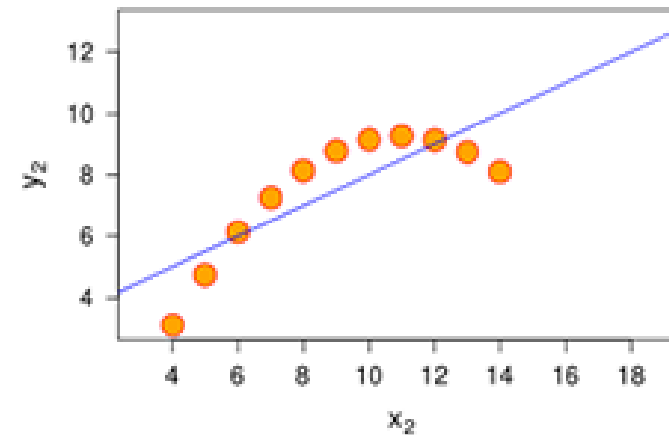
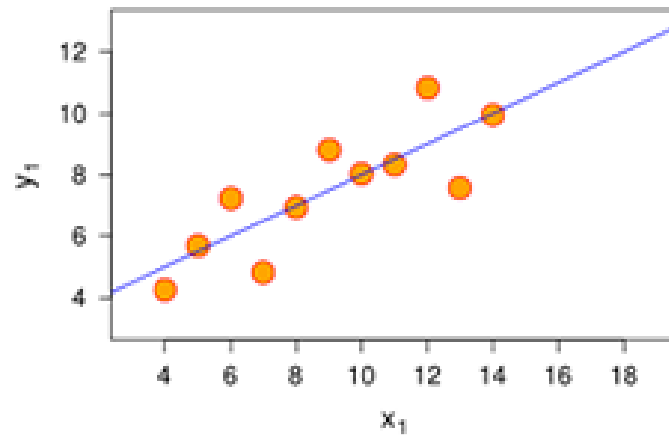
# Exemple

- Evaluation du modèle:

Une fois un premier travail de modélisation effectué, la suite de l'étude s'effectue par **l'évaluation de la qualité de notre modèle**, c'est à dire sa capacité à représenter avec exactitude notre phénomène.

# Exemple

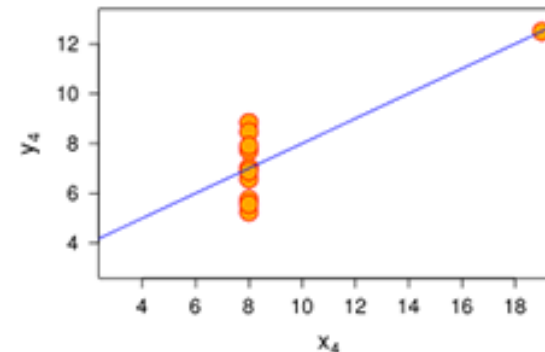
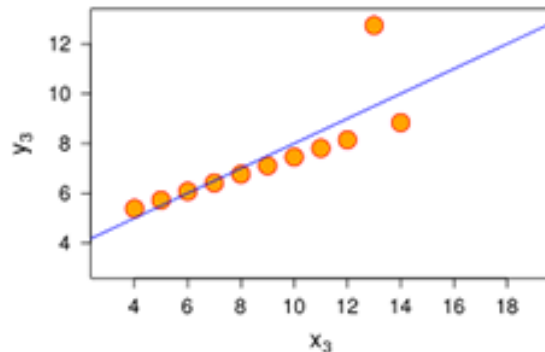
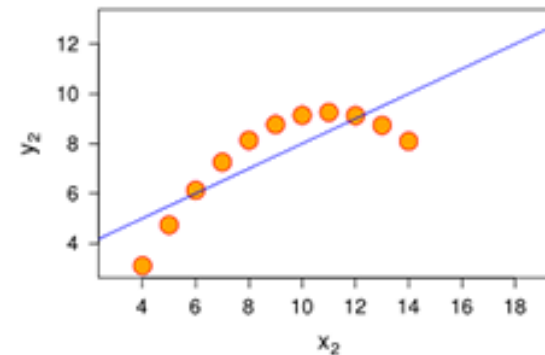
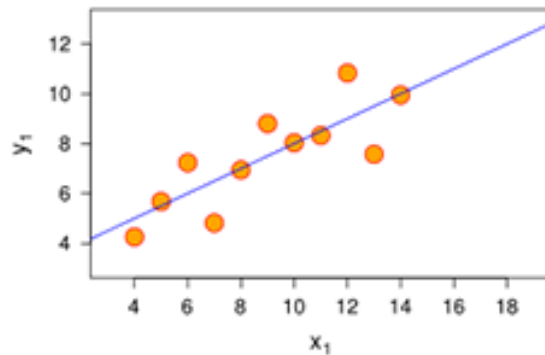
- Evaluation du modèle:



# Exemple

- Evaluation du modèle:

Dans les cas ci-dessus, il faudrait clairement changer le modèle d'une droite que nous avons décidé au départ !



# Exemple

- C'est donc un jeu d'allers-retours entre **modélisation** et **évaluation** qui s'effectue pour obtenir les performances les plus satisfaisantes possibles.
- Il est même possible dans certains cas de remettre en question certaines hypothèses de départ et de repartir dans une phase d'exploration pour mieux comprendre les données.



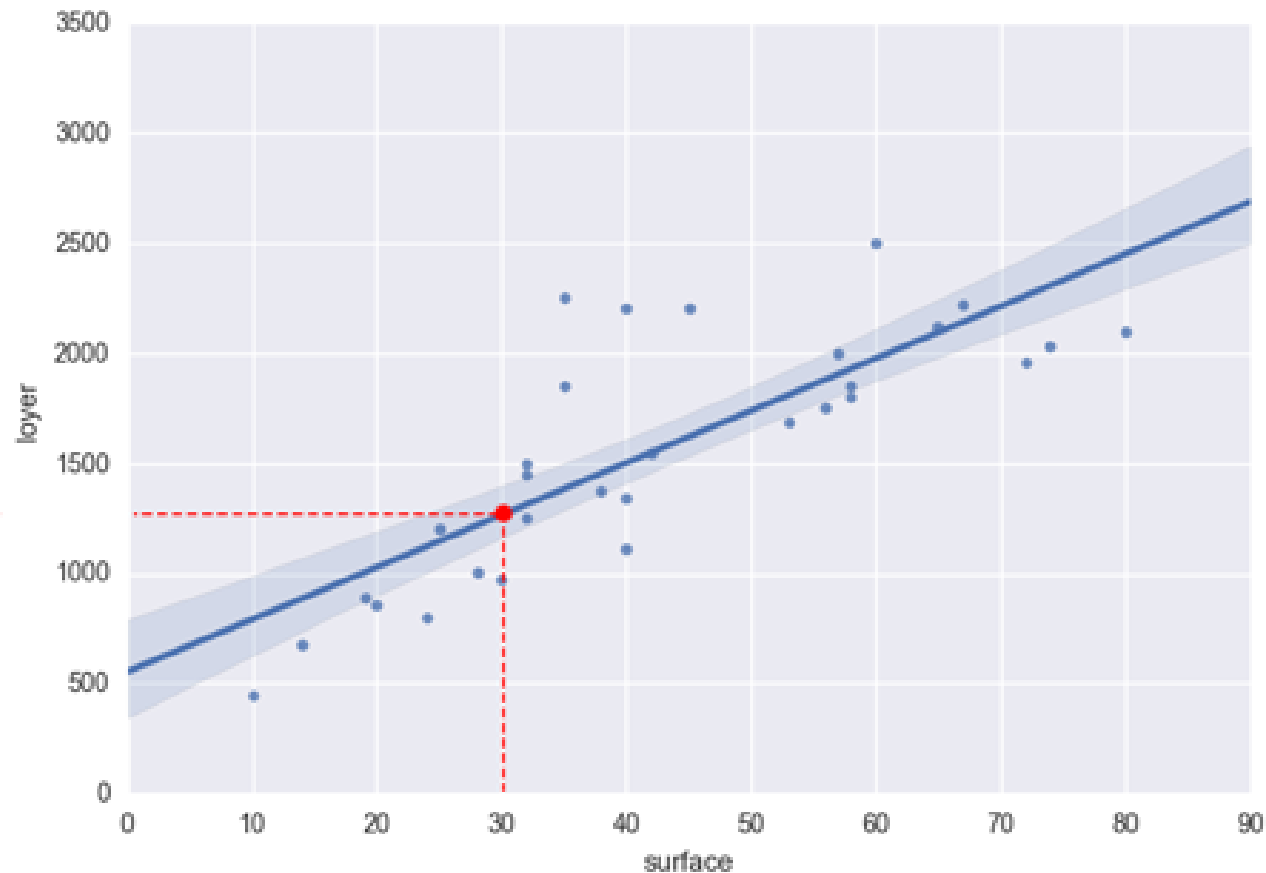
# Exemple

- Déployez le modèle en production:

**Une fois qu'on est satisfait de la qualité des performances de notre modèle, on va pouvoir passer à l'étape suivante, qui est le rendu de nos résultats et le potentiel déploiement du modèle en production.**

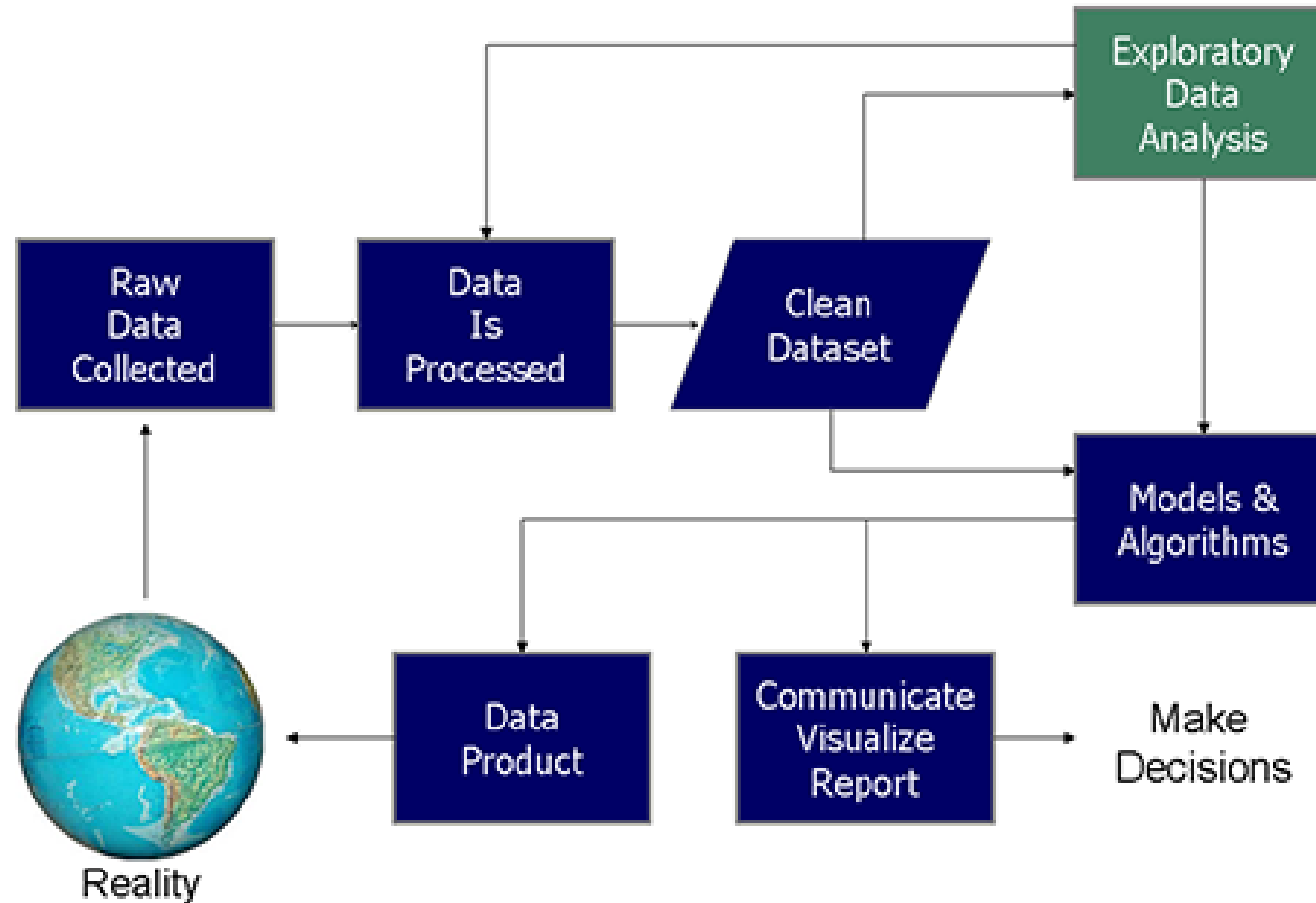
- Imaginez que vous trouvez que votre modèle d'évaluation des loyers est très performant, et mériterait d'être partagé.
- Vous décidez donc de le déployer sur un serveur ou tout le monde pourra obtenir une estimation de son loyer selon votre modèle, et ainsi déterminer s'il paie plus ou moins que les prix du marché !
- Cela l'aidera sûrement dans sa décision de déménager.

# Exemple



Imaginez qu'un appartement a une surface de 30 mètres carrés (point en rouge),  
une estimation légitime du loyer se situerait aux alentours de 1300 euros selon  
notre modèle

# Le cycle de travail



# Le cycle de travail

- Le cycle final de travail comprend notamment:
  - la récupération des données utiles à l'étude
  - le nettoyage des données pour les rendre exploitables
  - une longue phase d'exploration des données afin de comprendre en profondeur l'articulation des données
  - la modélisation des données
  - l'évaluation et interprétation des résultats
  - la conclusion de l'étude : prise de décision ou déploiement en production du modèle

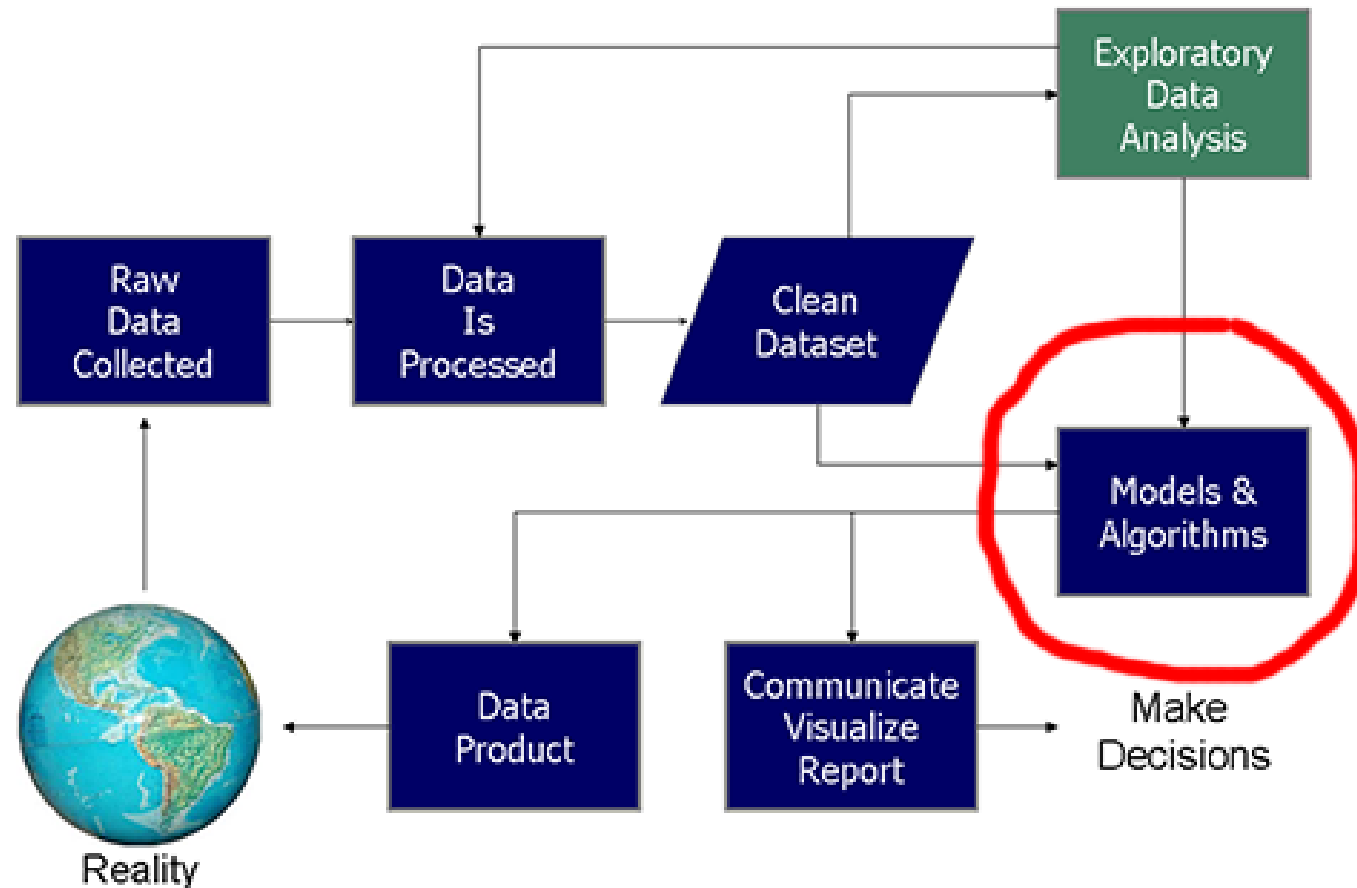
Au sein de ce cycle, le **machine learning** désigne l'ensemble des méthodes de **modélisation statistique à partir des données**, et se situe bien au coeur du travail du data scientist.

# La modélisation

---

# La modélisation

- Le **machine learning** permet de **modéliser des phénomènes, dans le but de prendre des décisions stratégiques.**
- **Machine Learning:** Méthode permettant de **construire un modèle de la réalité à partir de données**, soit en améliorant un modèle partiel, soit en le créant complètement.



# La modélisation

- **"Modéliser"** veut dire **représenter le comportement d'un phénomène** afin de pouvoir directement **aider à la résolution d'un problème concret de l'entreprise.**
- Ce problème concret peut par exemple être :
  - Identifier des transactions frauduleuses,
  - Recommander un article personnalisé à un utilisateur,
  - Prédire le prix optimal d'un produit pour maximiser le nombre de ventes,
  - etc.

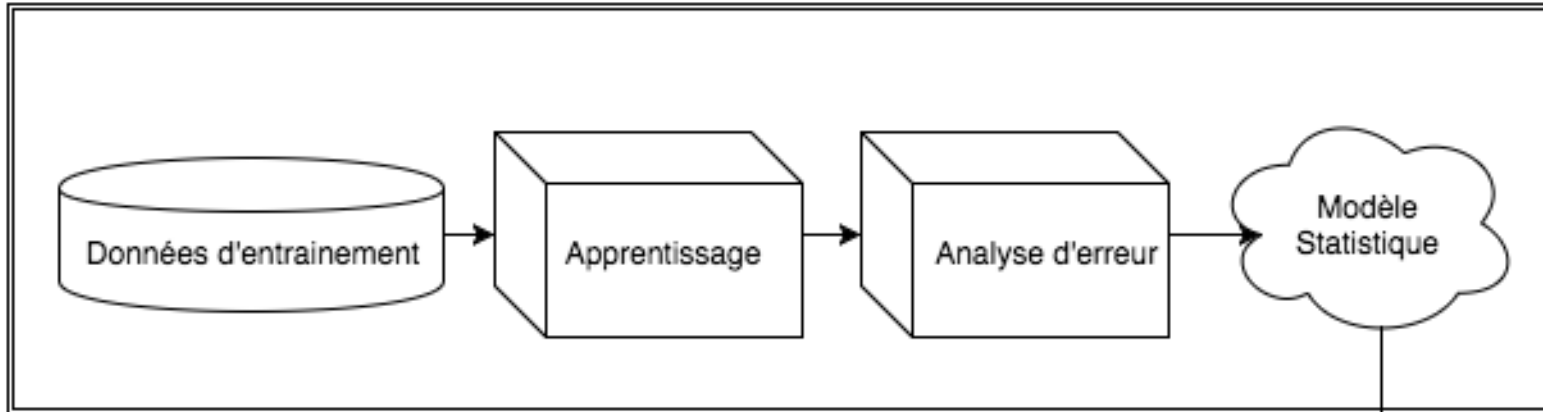
# L'apprentissage

- En **machine learning**, l'idée est que **l'algorithme construit une "représentation interne" tout seul** afin de pouvoir effectuer la tâche qui lui est demandée (prédiction, identification, etc).
- Pour cela, il va d'abord falloir lui entrer une dataset d'exemples afin qu'il puisse **s'entraîner** et s'améliorer, d'où le mot **apprentissage**. Cette dataset va contenir un **training set** (données d'entraînement).
- Et on appelle une entrée/exemple dans cette dataset : une **instance** ou une **observation**.

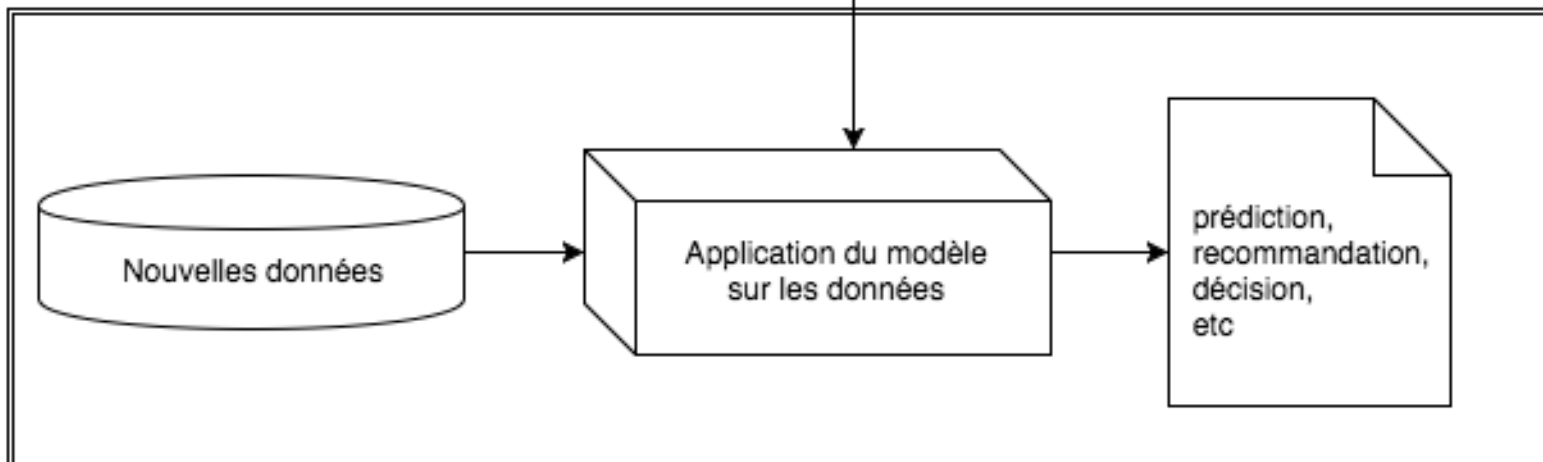


# Processus du Machine Learning en 2 phases

## Phase 1: Apprentissage

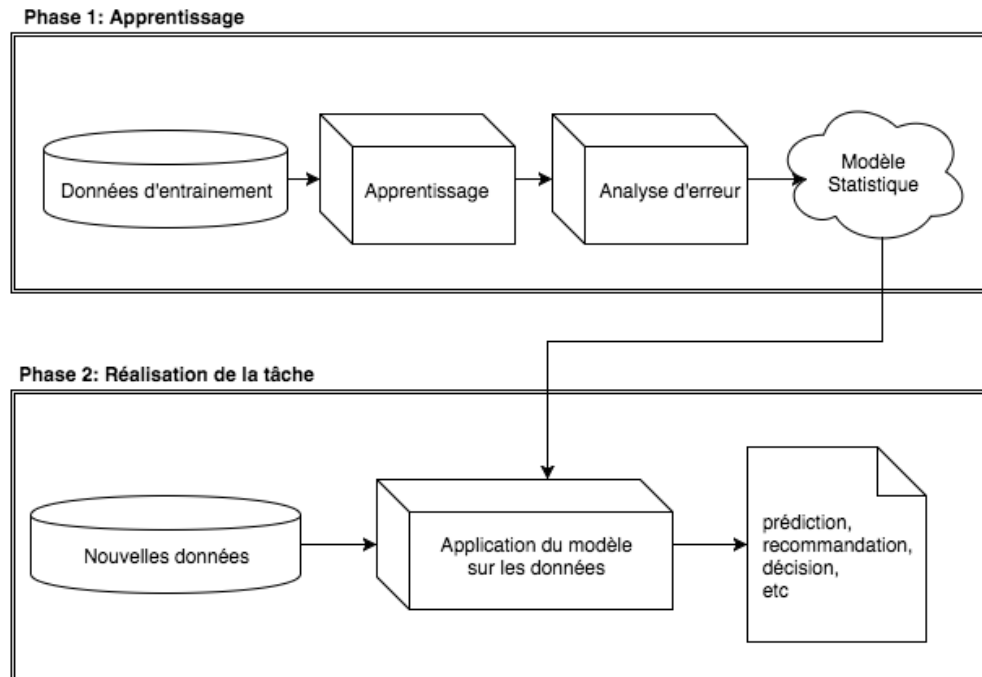


## Phase 2: Réalisation de la tâche



# Processus du Machine Learning en 2 phases

- Nous allons donc être responsable de la première phase: choix du training set et de l'entraînement de l'algorithme d'apprentissage.
- Alors que le traitement de la tâche spécifique sera appris à partir du training set et ensuite effectué par l'algorithme lui même dans la deuxième phase.



# Machine Learning

- En résumé, en **Machine Learning** le travail consiste à **sélectionner les bonnes données** test, **choisir et entraîner le bon algorithme** en vérifiant grâce à l'**analyse d'erreurs** que le modèle devient de plus en plus performant et robuste.

Si les performances s'améliorent lorsqu'on lui fourni les données d'entraînement, on dit alors que la machine "**apprend**".

- Une fois le modèle correctement paramétré sur les données d'entraînement, le modèle peut ensuite être déployer pour traiter de nouvelles données, pour accomplir **la tâche spécifique** (prédiction, recommandation, décision...).

# Machine Learning

- Le problème d'apprentissage machine comporte ainsi 4 éléments spécifiques :
  - **Les données** (les données d'entraînement mais aussi les nouvelles données)
  - **La tâche spécifique** à accomplir (prédire, recommander, décider quelque chose, etc.)
  - **L'algorithme d'apprentissage** en lui-même
  - **L'analyse d'erreur** (ou **mesure des performances** du modèle)

# Rappel: Les données

- Les données et leurs prétraitement est l'une des plus importantes étapes. C'est le cœur de la data science.
- L'idée est que **plus vous aurez une bonne compréhension de vos données, plus vous serez en mesure de bien les utiliser lors de la phase d'entraînement de votre modèle d'apprentissage.**



# La tâche à accomplir

- La **tâche spécifique** à accomplir **correspond au problème qu'on cherche à résoudre grâce à la modélisation du phénomène.**
- On peut distinguer un certain nombre de cas qui reviennent souvent dans un environnement business, tels que:
  - les **recommandations** de produits.
  - l'**identification** de transactions frauduleuses,
  - la **prédiction** du prix optimal d'un produit pour maximiser le nombre de ventes.
- Chaque tâche se traduira différemment et nécessitera bien sûr le choix d'algorithmes différents.
- Nous examinerons par la suite quelques problématiques courantes et comment elles se traduisent en problèmes de Machine Learning.

# L'algorithme d'apprentissage

- **L'algorithme d'apprentissage** constitue **la méthode avec laquelle le modèle statistique va se paramétrer à partir des données d'exemple.**
- Il existe de nombreux algorithmes différents ! On choisira un type d'algorithme particulier en fonction du type de tâche que l'on souhaite accomplir et du type de données dont on dispose. Plus précisément, selon *l'entrée* et *la sortie* de l'algorithme.
- Algorithmes de bases.
  - Regression linéaire/multiple/polynomiale
  - Naïve Bayes
  - Decision Trees
  - K-Nearest Neighbors
  - K-Means
  - Neural Networks (Réseaux de Neurones)

# La mesure de performance

- **Mesurer les performances** fait partie intégrante du travail de modélisation. Il faut en général déterminer une mesure principale, souvent spécifique à la tâche à accomplir.
- Le choix de cette métrique est très important! **On veut utiliser une métrique pertinente pour être sûr de mesurer correctement la qualité de l'algorithme.**



# Recapitulatif

- Le **Machine Learning** est l'apprentissage d'un modèle statistique par la machine, grâce à des données d'entraînement.
- Un problème de Machine Learning comporte plusieurs éléments spécifiques :
  - **Des données**
  - **Une tâche à accomplir**
  - **Un algorithme d'apprentissage**
  - **Une mesure des performances**

un ordinateur **apprend** à partir de **données** pour résoudre une **tâche** en faisant attention à **mesurer les performances**. S'il **améliore** les performances sur cette tâche, lorsqu'on lui fourni les données d'entraînement, on dit alors qu'il **apprend**.

# Machine Learning: les types de problème

- Une première grande distinction à faire en **machine learning** est la différence entre apprentissage **supervisé** et **non-supervisé**.
- En anglais ces deux notions se nomment respectivement ***supervised learning*** et ***unsupervised learning***.

# Apprentissage supervisé vs non-supervisé

- **L'apprentissage supervisé** est une technique d'apprentissage automatique où l'on cherche à produire automatiquement des règles à partir d'une base de données d'apprentissage contenant des «exemples». (**Classification, Régression**)
- **L'apprentissage non-supervisé** est une méthode d'apprentissage automatique. Un algorithme va diviser un groupe hétérogène de données, en sous-groupes de manière que les données considérées comme les plus similaires soient associées au sein d'un groupe homogène et qu'au contraire les données considérées comme différentes se retrouvent dans d'autres groupes distincts. (**Clustering**)

# Apprentissage supervisé vs non-supervisé

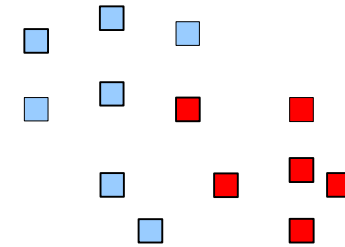
- **Apprentissage supervisé**

- On dispose d'**éléments déjà classés**
- Exemple : articles en rubrique soit politique ou sport
- On veut **classer un nouvel élément**
- Exemple : lui attribuer une étiquette parmi économie, politique, sport, culture...

- **Apprentissage non supervisé**

- On dispose d'**éléments non classés**
- Exemple : mots d'un texte
- On veut les **regrouper en classes**
- Exemple : si deux mots ont la même écriture, ils sont en rapport avec une même thématique...

# Apprentissage supervisé vs non-supervisé



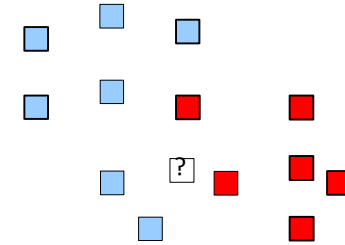
- **Apprentissage supervisé**

- On dispose d'**éléments déjà classés**
- Exemple : articles en rubrique soit politique ou sport
- On veut **classer un nouvel élément**
- Exemple : lui attribuer une étiquette parmi les deux: politique ou sport

- **Apprentissage non supervisé**

- On dispose d'**éléments non classés**
- Exemple : mots d'un texte
- On veut les **regrouper en classes**
- Exemple : si deux mots ont la même écriture, ils sont en rapport avec une même thématique...

# Apprentissage supervisé vs non-supervisé



- **Apprentissage supervisé**

- On dispose d'**éléments déjà classés**
- Exemple : articles en rubrique soit politique ou sport
- On veut **classer un nouvel élément**
- Exemple : lui attribuer une étiquette parmi les deux: politique ou sport

- **Apprentissage non supervisé**

- On dispose d'**éléments non classés**
- Exemple : mots d'un texte
- On veut les **regrouper en classes**
- Exemple : si deux mots ont la même écriture, ils sont en rapport avec une même thématique...

# Apprentissage supervisé vs non-supervisé

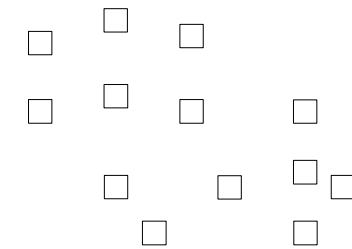


- **Apprentissage supervisé**

- On dispose d'**éléments déjà classés**
- Exemple : articles en rubrique soit politique ou sport
- On veut **classer un nouvel élément**
- Exemple : lui attribuer une étiquette parmi les deux: politique ou sport

- **Apprentissage non supervisé**

- On dispose d'**éléments non classés**
- Exemple : mots d'un texte
- On veut les **regrouper en classes**
- Exemple : si deux mots ont la même écriture, ils sont en rapport avec une même thématique...



# Apprentissage supervisé vs non-supervisé

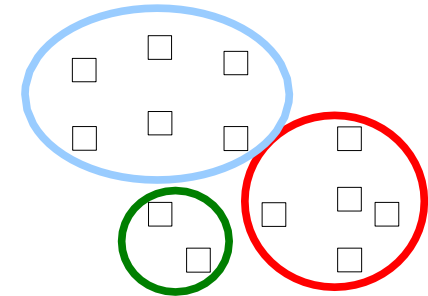


- **Apprentissage supervisé**

- On dispose d'**éléments déjà classés**
- Exemple : articles en rubrique soit politique ou sport
- On veut **classer un nouvel élément**
- Exemple : lui attribuer une étiquette parmi les deux: politique ou sport

- **Apprentissage non supervisé**

- On dispose d'**éléments non classés**
- Exemple : mots d'un texte
- On veut les **regrouper en classes**
- Exemple : si deux mots ont la même écriture, ils sont en rapport avec une même thématique...





# Apprentissage supervisé vs non-supervisé

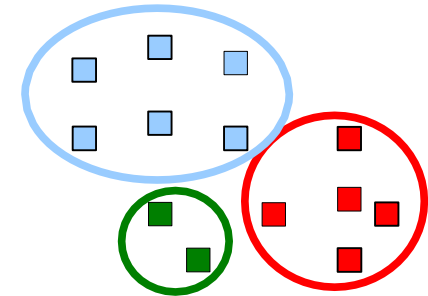


- **Apprentissage supervisé**

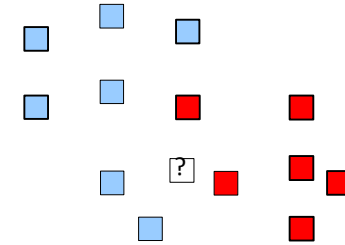
- On dispose d'**éléments déjà classés**
- Exemple : articles en rubrique soit politique ou sport
- On veut **classer un nouvel élément**
- Exemple : lui attribuer une étiquette parmi les deux: politique ou sport

- **Apprentissage non supervisé**

- On dispose d'**éléments non classés**
- Exemple : mots d'un texte
- On veut les **regrouper en classes**
- Exemple : si deux mots ont la même écriture, ils sont en rapport avec une même thématique...



# Apprentissage supervisé vs non-supervisé

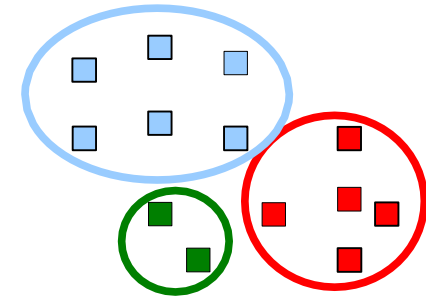


- **Apprentissage supervisé**

- On dispose d'**éléments déjà classés**
- Exemple : articles en rubrique soit politique ou sport
- On veut **classer un nouvel élément**
- Exemple : lui attribuer une étiquette parmi les deux: politique ou sport

- **Apprentissage non supervisé**

- On dispose d'**éléments non classés**
- Exemple : mots d'un texte
- On veut les **regrouper en classes**
- Exemple : si deux mots ont la même écriture, ils sont en rapport avec une même thématique...



# Apprentissage supervisé vs non-supervisé

- Prenons un exemple:

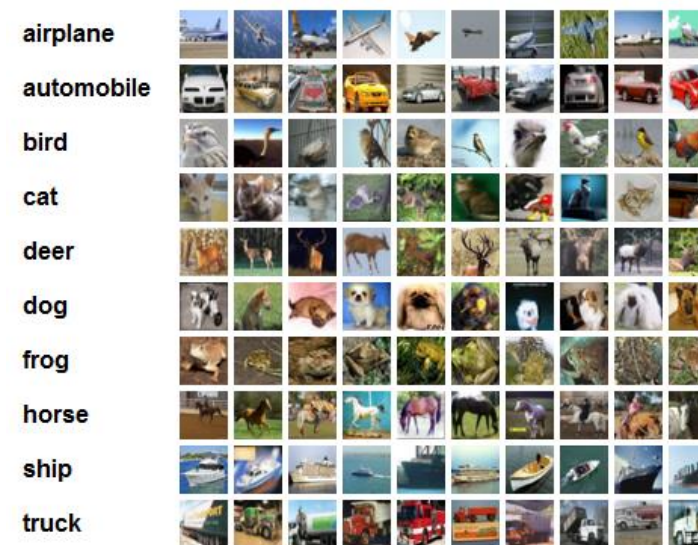
supposez que vous avez une nouvelle base de photos à catégoriser...

# Apprentissage supervisé vs non-supervisé

- Prenons un exemple:

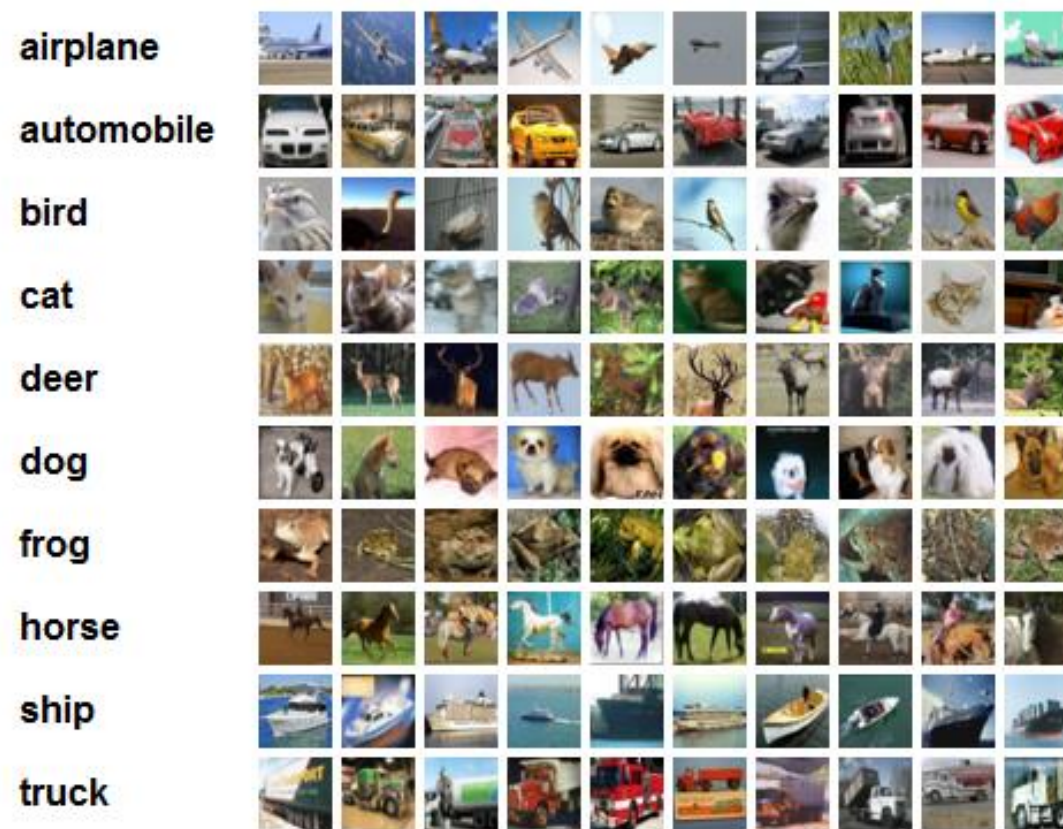
En **apprentissage supervisé**, vous allez récupérer des **données dites annotées** pour entraîner le modèle, c'est à dire que les données utilisées ont un **label associé** et appartiennent à une **classe cible** et ***on veut que l'algorithme devienne capable de la prédire sur de nouvelles données non annotées une fois entraîné.***

Dans ce cas, les données d'entrée sont des images, et la cible (ou *target* en anglais) est la catégorie de photos.



Chaque image utilisée pour entraîner l'algorithme est étiquetée avec sa catégorie

# Apprentissage supervisé vs non-supervisé



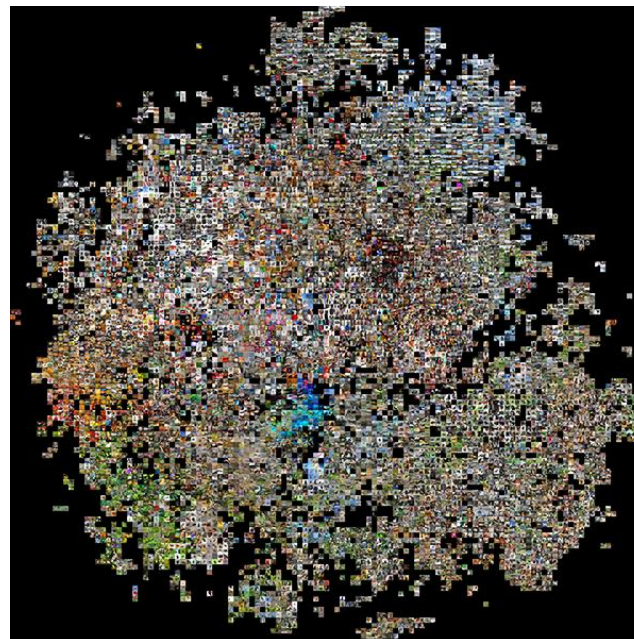
Chaque image utilisée pour entraîner l'algorithme est étiquetée avec sa catégorie

# Apprentissage supervisé vs **non-supervisé**

- Prenons un exemple:

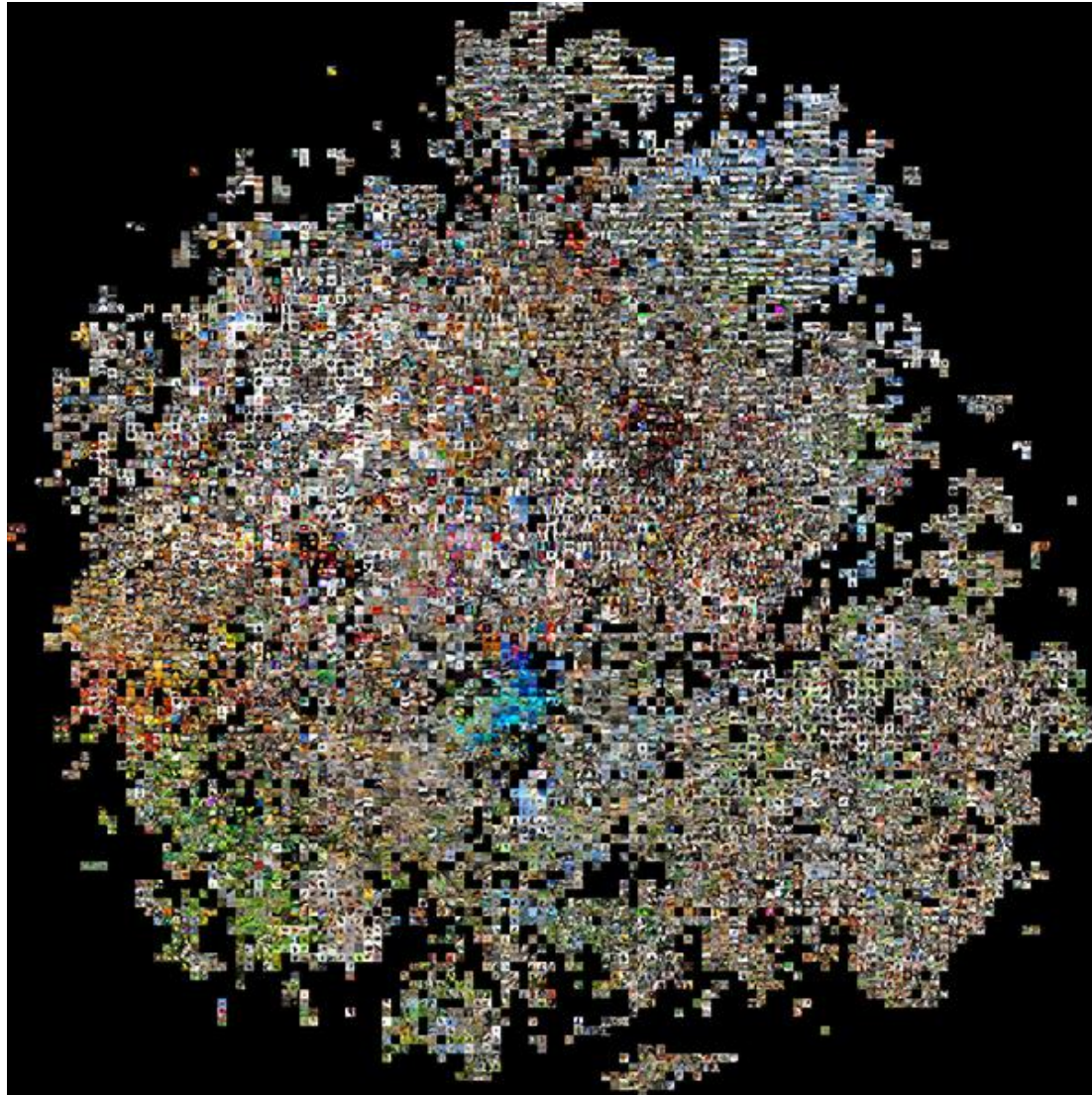
En **apprentissage non-supervisé**, les données d'entrées ne sont pas annotées. Comment cela peut fonctionner ? Et bien, ***l'algorithme d'entraînement s'applique dans ce cas à trouver seul les similarités et distinctions au sein de ces données***, et à regrouper ensemble celles qui partagent des caractéristiques communes.

Dans ce cas, les photos similaires seraient ainsi regroupées automatiquement au sein d'une même catégorie.





# Apprentissage supervisé vs **non-supervisé**



# Apprentissage supervisé vs non-supervisé

- En **Apprentissage supervisé**, on reçoit des données d'exemple annotées:  $(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots$  et on espère **prédire** la sortie sur de nouvelles observations :  $x^* \rightarrow y^*$
- En **Apprentissage non supervisé**, on reçoit uniquement des observations brutes de variables aléatoires:  $x_1, x_2, x_3, x_4, \dots$  et on espère découvrir la relation avec des variables latentes structurelle :  $x_i \rightarrow y_i$



## Apprentissage supervisé vs non-supervisé: Cas d'utilisation

- Dans le cas où vous avez un problème ou vous pouvez **annoter précisément** pour chaque observation la cible que vous voulez en sortie, vous pouvez utiliser ***l'apprentissage supervisé.***
- Dans le cas où vous essayez de **mieux comprendre votre dataset où d'identifier des comportements intéressants**, vous pouvez utiliser ***l'apprentissage non-supervisé.***

# Apprentissage supervisé vs non-supervisé: Types de problèmes

## **Reconnaître le problème:**

- Regrouper automatiquement mes photos personnels en catégories
- Apprendre à aider au diagnostic d'une maladie à partir d'images médicales
- Etant donné le profil d'un client, est ce qu'en tant que banque je peux lui octroyer un prêt?
- Etant donné une bande musicale, séparer la voix du chanteur du reste de la musique

# Apprentissage supervisé vs non-supervisé: Types de problèmes

## Reconnaître le problème:

- Regrouper automatiquement mes photos personnels en catégories
  - **non-supervisé**
- Apprendre à aider au diagnostic d'une maladie à partir d'images médicales
- Etant donné le profil d'un client, est ce qu'en tant que banque je peux lui octroyer un prêt?
- Etant donné une bande musicale, séparer la voix du chanteur du reste de la musique

# Apprentissage supervisé vs non-supervisé: Types de problèmes

## Reconnaître le problème:

- Regrouper automatiquement mes photos personnels en catégories
  - non-supervisé
- Apprendre à aider au diagnostic d'une maladie à partir d'images médicales
  - **supervisé**
- Etant donné le profil d'un client, est ce qu'en tant que banque je peux lui octroyer un prêt?
- Etant donné une bande musicale, séparer la voix du chanteur du reste de la musique

# Apprentissage supervisé vs non-supervisé: Types de problèmes

## Reconnaître le problème:

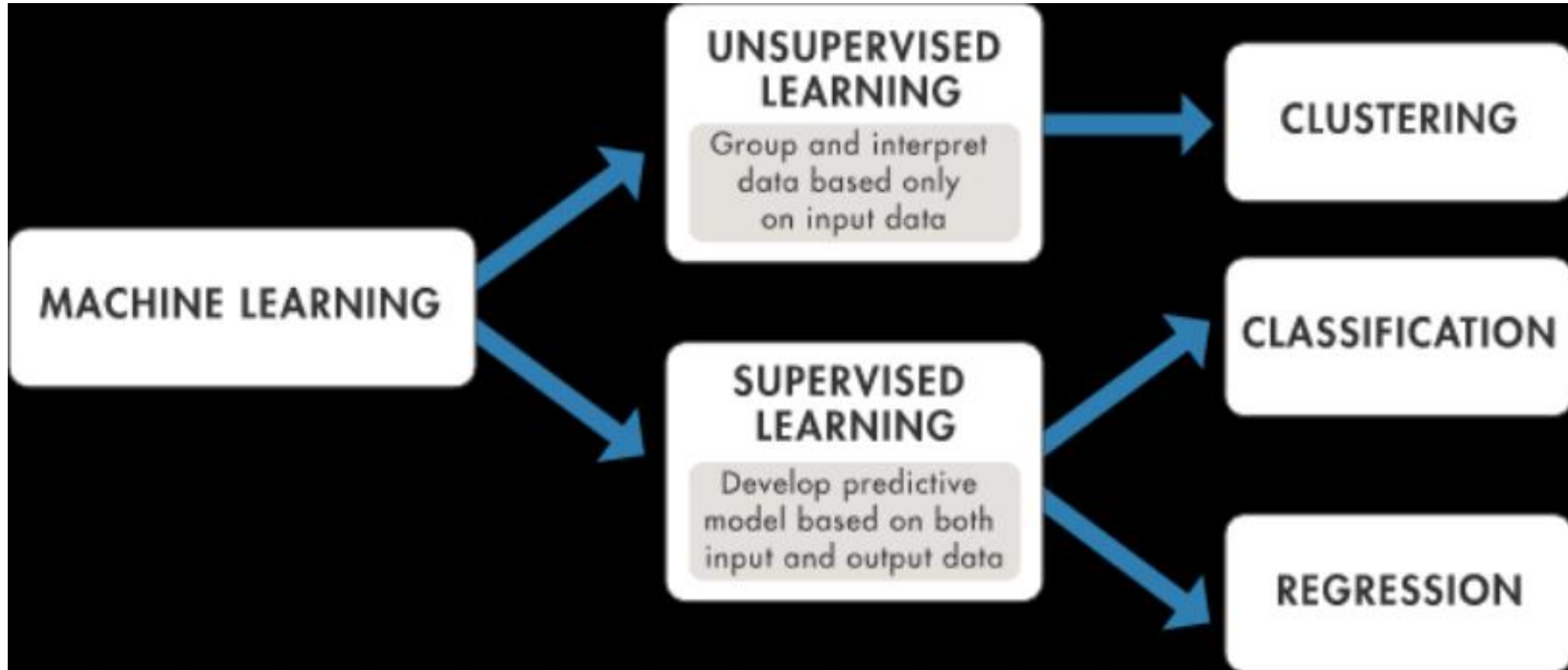
- Regrouper automatiquement mes photos personnels en catégories
  - non-supervisé
- Apprendre à aider au diagnostic d'une maladie à partir d'images médicales
  - supervisé
- Etant donné le profil d'un client, est ce qu'en tant que banque je peux lui octroyer un prêt?
  - **supervisé**
- Etant donné une bande musicale, séparer la voix du chanteur du reste de la musique

# Apprentissage supervisé vs non-supervisé: Types de problèmes

## Reconnaître le problème:

- Regrouper automatiquement mes photos personnels en catégories
  - non-supervisé
- Apprendre à aider au diagnostic d'une maladie à partir d'images médicales
  - supervisé
- Etant donné le profil d'un client, est ce qu'en tant que banque je peux lui octroyer un prêt?
  - supervisé
- Etant donné une bande musicale, séparer la voix du chanteur du reste de la musique
  - **non-supervisé**

# Formes d'apprentissage



# Apprentissage Supervisé

---



# Régression vs Classification

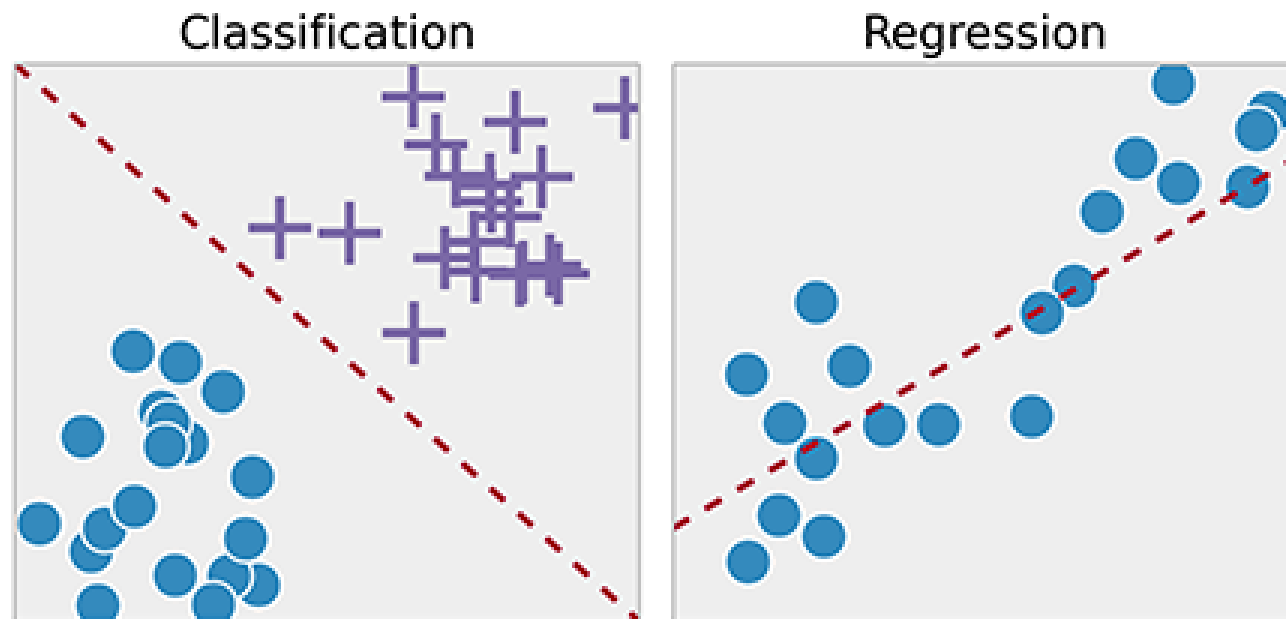
Une autre distinction qui vous aidera dans le choix d'un algorithme de machine learning est **le type de sortie** que l'on attend de notre programme :

- Est-ce une valeur **continue** (un nombre) ou bien une **catégorie** ?
- Le premier cas est appelé une **régression**, le second une **classification**.

# Régression vs Classification

Par exemple:

- Si je veux déterminer le coût par clic d'une publicité web, j'effectue **une régression**.
- Si je veux déterminer si une photo est un chat ou un chimpanzé, j'effectue **une classification**.



# Régression vs Classification

- **Régression** : lorsque le résultat supervisé que l'on cherche à estimer est une valeur dans un ensemble **continu** de réels.
- **Classification** : lorsque l'ensemble des valeurs de sortie est **discret**. Ceci revient à attribuer une **classe** (aussi appelée *étiquette* ou *label*) pour chaque vecteur d'entrée.

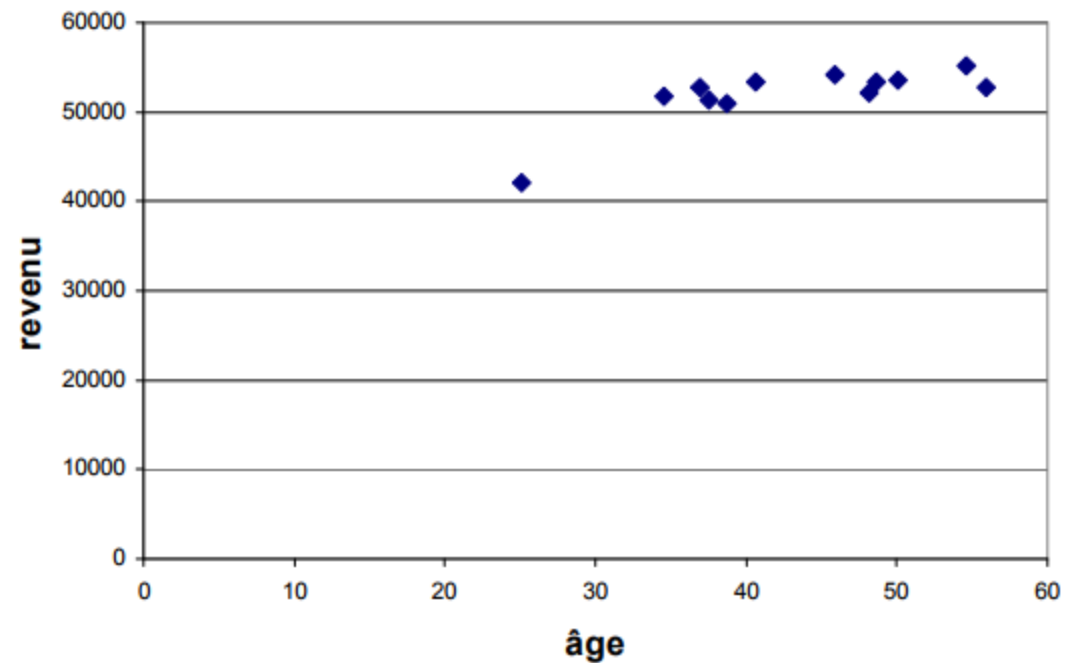
# Régression (Exemple)

- Le syndic s'intéresse au rapport entre l'âge et le revenu des résidents de Statville. Il sélectionne un échantillon aléatoire simple de taille  $n = 12$ .

*données de l'échantillon :*

ind.	revenu	âge
1	52125.0	48.1
2	50955.9	38.7
3	53382.9	48.6
4	51286.9	37.5
5	55243.6	54.7
6	53384.7	40.7
7	53488.2	50.1
8	54134.1	45.9
9	52706.4	55.9
10	42144.3	25.1
11	52665.2	36.9
12	51656.7	34.5
<b>Moyenne</b>	<b>51931.2</b>	<b>43.1</b>
<b>Ecart type</b>	<b>3314.9</b>	<b>9.1</b>

*nuage de points :*



# Classification (Exemple)

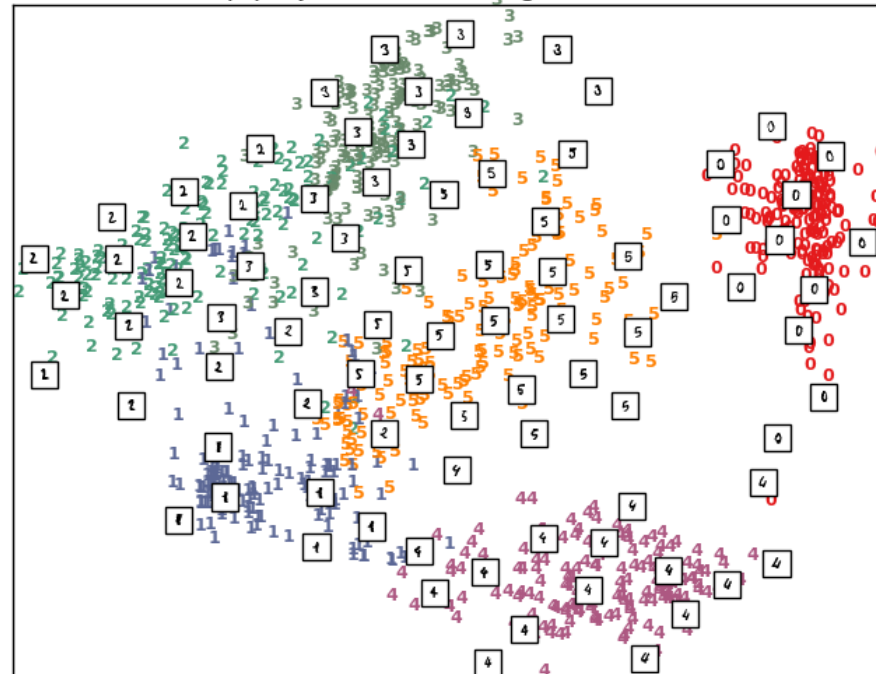
## Classification de chiffres écrits :

On veut faire la Reconnaissance des codes postaux pour le tri automatique du courrier. On veut prédire le chiffre correct la prochaine fois qu'on voit une image.

A selection from the 64-dimensional digits dataset



Isomap projection of the digits (time 1.10s)



# Apprentissage non-supervisé

---

# Clustering

- Le **Clustering** permet de **partitionner** les données en **sous-groupes**, ou *clusters*, de manière non supervisée. Intuitivement, ces sous-groupes regroupent entre elles des observations **similaires**.
- Les algorithmes de clustering dépendent donc fortement de la façon dont on définit cette notion de **similarité**, qui est souvent spécifique au domaine d'application.

# Clustering

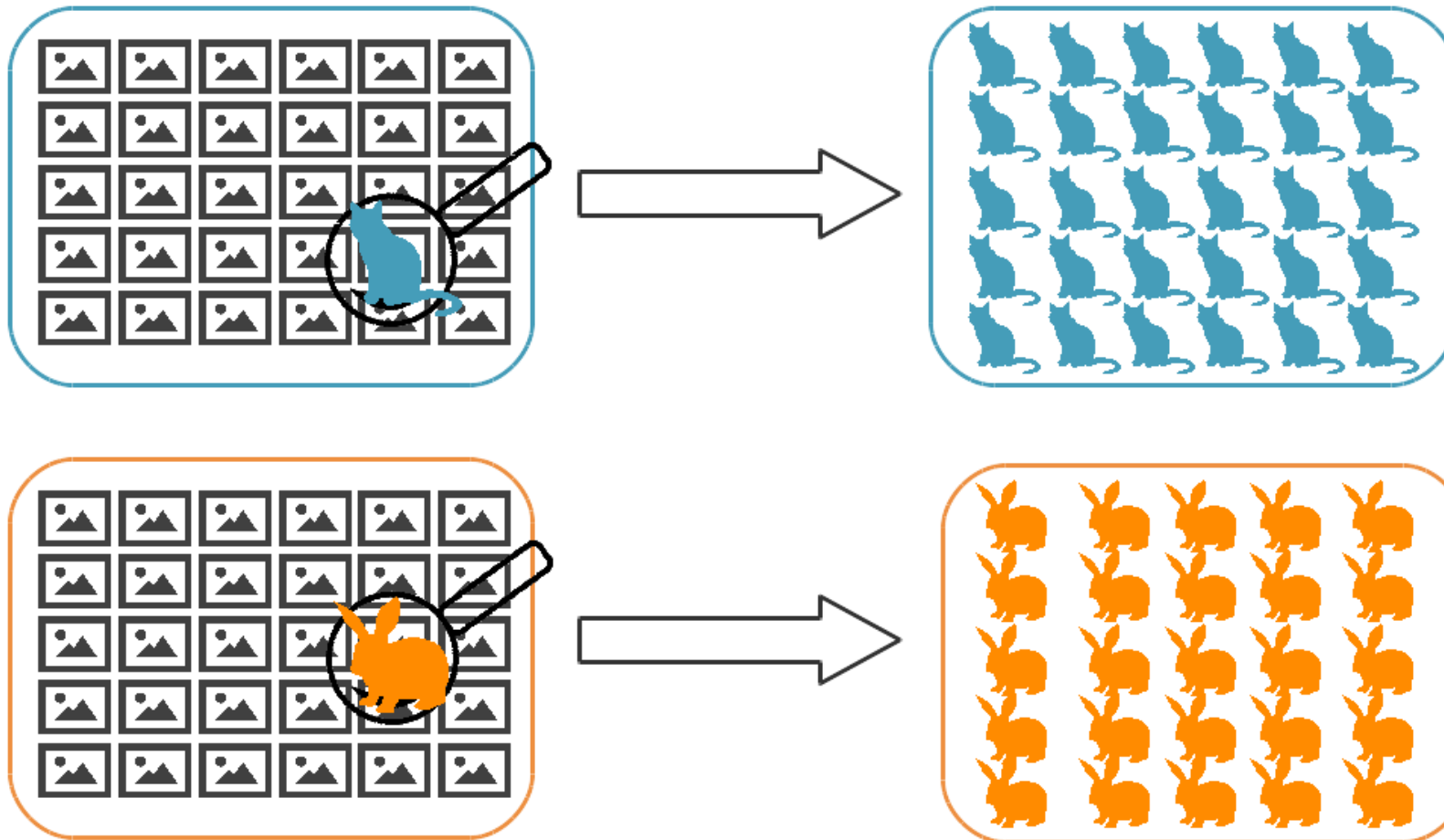
## Cas d'utilisation des algorithmes de Clustering

- **Comprendre les données:** Pour explorer les données, par exemple on veut identifier :
  - des clients qui ont des comportements similaires (**segmentation de marché**);
  - des communautés dans des réseaux sociaux ;
  - des motifs récurrents dans des transactions financières.
- **Déduire des propriétés depuis les données:** dans le cas où il est coûteux d'étiqueter les données, par exemple on veut identifier:
  - des images similaires, susceptibles de représenter le même objet, le même animal ou la même personne ;
  - des textes similaires, susceptibles de parler du même sujet ;
  - dans une image, les points qui appartiennent au même objet (on parle alors plus spécifiquement de **segmentation**).



# Clustering

- Une fois les images regroupées dans des clusters, il suffit d'identifier que l'image du haut représente un chat pour inférer (dédire) que toutes les images du cluster du haut représentent vraisemblablement des chats.



Traduire  
des **Problématiques business**  
en  
des **Problèmes Machine Learning**

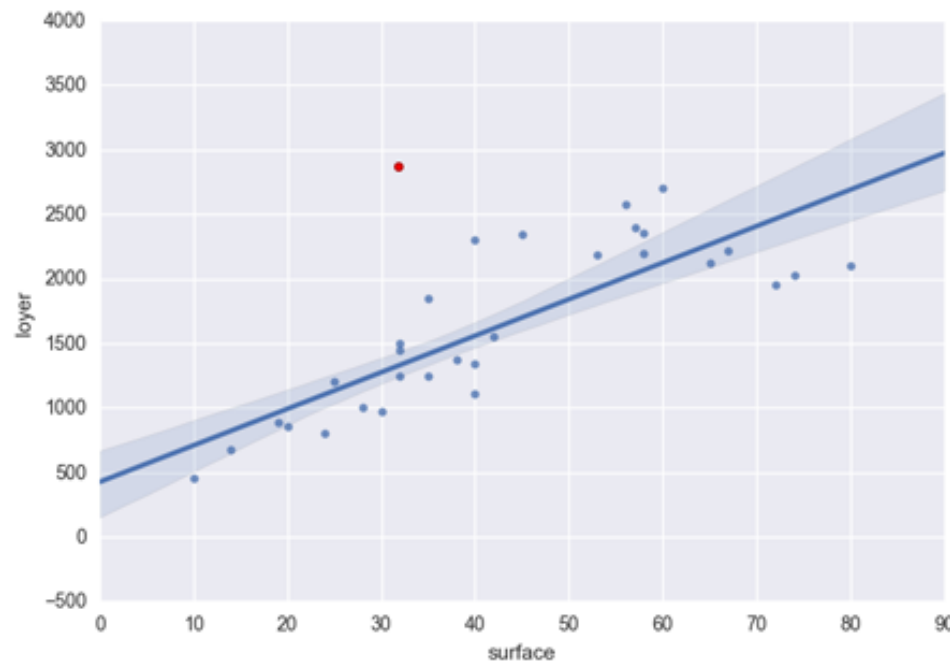
---

# Exemple1: Identification des événements rares

- L'identification d'événements appelés événements rares : les spams d'email, les transactions frauduleuses, les textes injurieux...
  1. On peut traiter ce problème comme un problème de **classification supervisé** en annotant les données d'entraînement "événement rare" et "événement normal"
  2. On effectue une première modélisation (supervisée ou non-supervisée) qui va permettre de modéliser le comportement *habituel*. Ensuite, on fixe simplement un critère de **distance** qui permet de déterminer si une entrée est trop éloignée de la modélisation. Et ainsi on identifie les événements rares comme étant des **outliers**.

# Exemple1: Identification des événements rares

2. On effectue une première modélisation (supervisée ou non-supervisée) qui va permettre de modéliser le comportement *habituel*. Ensuite, on fixe simplement un critère de **distance** qui permet de déterminer si une entrée est trop éloignée de la modélisation. Et ainsi on identifie les événements rares comme étant des **outliers**.



Détecter des événements aberrants (**outliers**) comme le point rouge

## Exemple2: Affecter une catégorie catalogue à un produit

- Lorsque le catalogue de produits d'un distributeur grand, la problématique de catalogage se fait ressentir, car elle demande un travail minutieux qui prendra beaucoup de temps.
- Comment peut traduire ce problème en Machine Learning?

## Exemple2: Affecter une catégorie catalogue à un produit

- Lorsque le catalogue de produits d'un distributeur grandi, la problématique de catalogage se fait ressentir, car elle demande un travail minutieux qui prendra beaucoup de temps.
- Comment peut traduire ce problème en Machine Learning?
  - Avec les bonnes entrées d'entraînement, l'automatisation de cette tâche peut se faire avec des algorithmes de **classification**.
  - En effet, nous sommes typiquement ici dans une problématique de **classification supervisée** : les produits déjà classés peuvent faire office de données d'entraînement et les caractéristiques de produits, voire même directement les photos, peuvent servir de variable d'entrée pour notre algorithme de classification.

## Exemple3: Recommander un produit à un client

- La recommandation est une problématique qui revient très souvent pour les data scientists :
  - Suggérer d'autres produits à acheter sur Amazon,
  - Suggérer des vidéos à regarder sur Youtube,
  - Suggérer des musiques à écouter sur Spotify,
  - ... etc.
- C'est vraiment utile, à la fois pour l'utilisateur final et pour l'entreprise qui peut ainsi proposer le contenu le plus pertinent.

## Exemple3: Recommander un produit à un client

- Une recommandation, c'est une proposition de contenu **similaire** aux produits qu'à déjà aimé l'utilisateur. C'est la notion de **similarité** qui est donc à traduire.
- Alors, c'est de la classification ? de la régression ? supervisé ? non-supervisé ?




























## Exemple3: Recommander un produit à un client

- Une recommandation, c'est une proposition de contenu **similaire** aux produits qu'à déjà aimé l'utilisateur. C'est la notion de **similarité** qui est donc à traduire.
- Alors, c'est de la classification ? de la régression ? supervisé ? non-supervisé ?
  - Le problème peut être formulé d'énormément de manières différentes. Une technique largement répandue est le "**collaborative filtering**", qui se base sur :
    - des similarités entre utilisateurs, ou bien
    - des similarités entre produits.
  - Dans ces deux cas, c'est un problème **non-supervisé**: on procure toutes nos données à l'algorithme et on le laisse essayer de déterminer les relations entre les différentes entités.

## Exemple3: Recommander un produit à un client


























- L'ensemble des visiteurs ont voté en faveur ou en défaveur de produits sur le site. À nous maintenant de prédire à quel point l'utilisateur de la dernière ligne va apprécier le produit qu'il n'a pas encore noté.

## Exemple3: Recommander un produit à un client

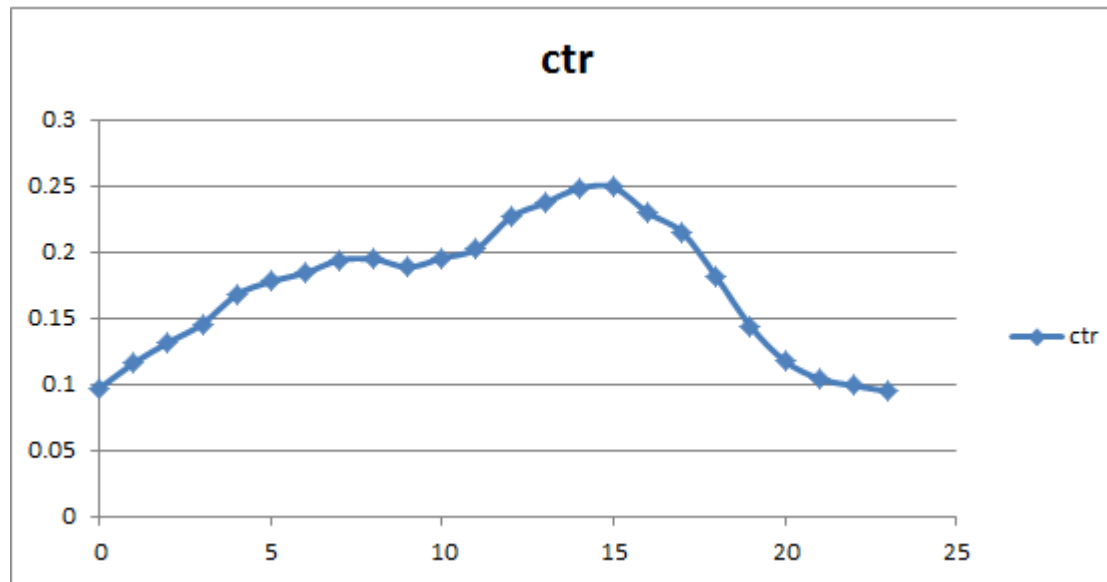
- on regarde par exemple ce qu'ont voté les utilisateurs similaires, c'est à dire ceux qui ont déjà voté la même chose sur d'autres produits (surlignés en vert). On peut alors prédire ce qu'aurait voté notre utilisateur sur le produit cherché, et ne proposer que les produits sur lesquels il aurait mis un pouce vert.

- Les utilisateurs similaires (en vert) n'ont pas aimé le produit que notre utilisateur n'a pas encore noté.
- L'algorithme aura donc tendance à prédire une mauvaise note (**Pouce rouge**) et à ne pas recommander le produit ici

## Exemple4: Prédire la rentabilité d'une campagne marketing

- Les entreprises ont souvent besoin d'évaluer le ROI (retour sur investissement) d'une campagne avant de la lancer, afin d'estimer si ça vaut la peine ou pas.



**Click-Through Rate:**  
Prédiction du CTR (# de visiteurs qui cliquent un lien) selon l'heure de la journée.

- Ce type de prédiction fait appel à des méthodes de **régression**, puisqu'on essaie de prédire une valeur numérique.

# Un 1<sup>er</sup> algorithme d'apprentissage

---

# Linear Regression

---

# La régression linéaire

- L'algorithme de **régression linéaire** est un algorithme d'apprentissage supervisé c'est-à-dire qu'à partir de la variable cible ou de la variable à expliquer ( $Y$ ), le modèle a pour but de faire **une prédiction** grâce à des variables dites explicatives ( $X$ ) ou prédictives.
- Un modèle de régression linéaire est un modèle de Machine Learning dont la variable cible ( $Y$ ) est **quantitative** tandis que la variable  $X$  peut être **quantitative** ou **qualitative**.
- L'objectif est de trouver une fonction dite de prédiction ou une fonction coût qui décrit la relation entre  $X$  et  $Y$  c'est-à-dire qu'à partir de valeurs connues de  $X$ , on arrive à donner une prédiction des valeurs de  $Y$ , tel que :  $Y=f(X)$
- L'ensemble des couples observés  $E = \{(x_1, y_1), \dots, (x_n, y_n)\}$  est appelé **ensemble d'entraînement** ou d'apprentissage (ou ensemble des données annotées ou étiquetées).

# Formalisation du problème

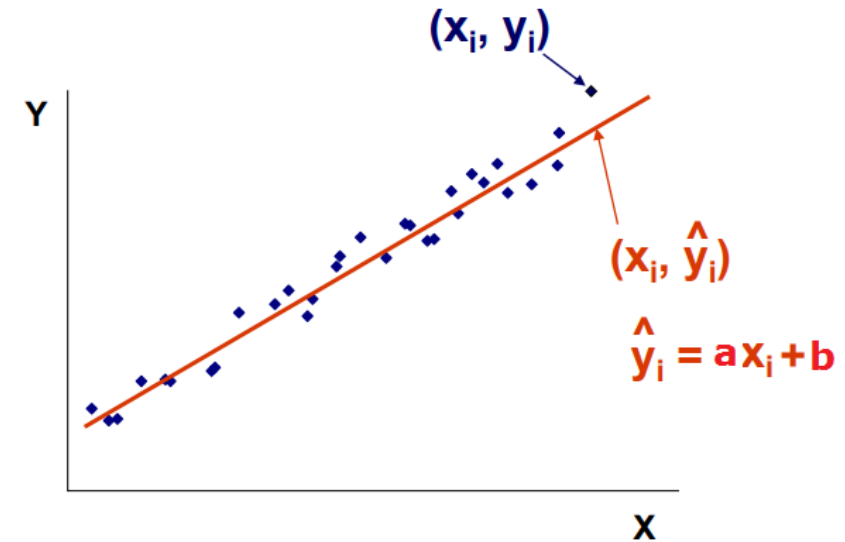
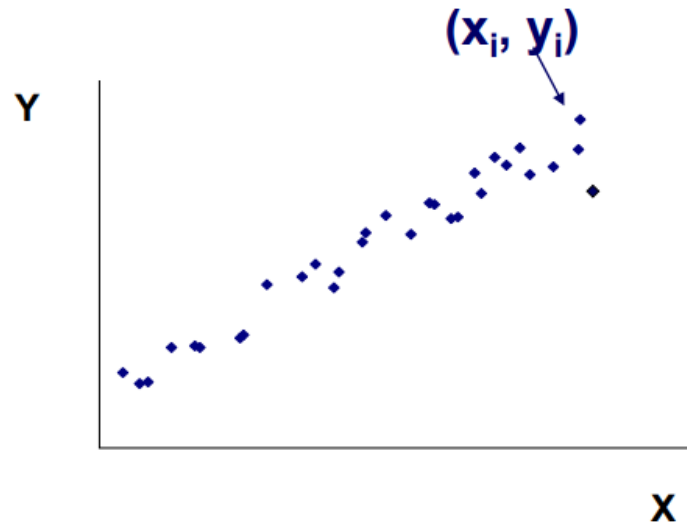
- Etant donné un ensemble d'entraînement  $E$ , on cherche à déterminer  $f: \mathbf{X} \rightarrow \mathbf{Y}$  une fonction **modélisant** la relation entre les  $\mathbf{X}$  et la variable cible  $\mathbf{Y}$
- En revanche, ne connaissant pas la vraie nature de la relation entre  $\mathbf{X}$  et  $\mathbf{Y}$  et les données observées étant soit bruitées, soit incomplètes ; il n'est pas raisonnable de supposer une relation déterministe. Alors, il est davantage raisonnable de poser le problème en les termes suivants :  $f(\mathbf{X}) = \mathbf{Y} + \varepsilon$  où  $\varepsilon$  est l'erreur ou le résidu.
- Autrement dit, il s'agit d'**approximer**  $f$  en commettant le **moins d'erreurs possibles** sur  $E$  tout en faisant de **bonnes prédictions** pour des valeurs de  $\mathbf{X}$  non encore observées.



# La régression linéaire

- Un modèle de régression linéaire simple est de la forme :

$$Y = aX + b + \varepsilon \quad , \text{ où } f(X) = aX + b \quad \text{avec :}$$



$Y$ , la variable cible, aléatoire dépendante

$a$  (**Slope** ou pente) et  $b$  (**Intercept** ou l'ordonnée à l'origine) les coefficients à estimer

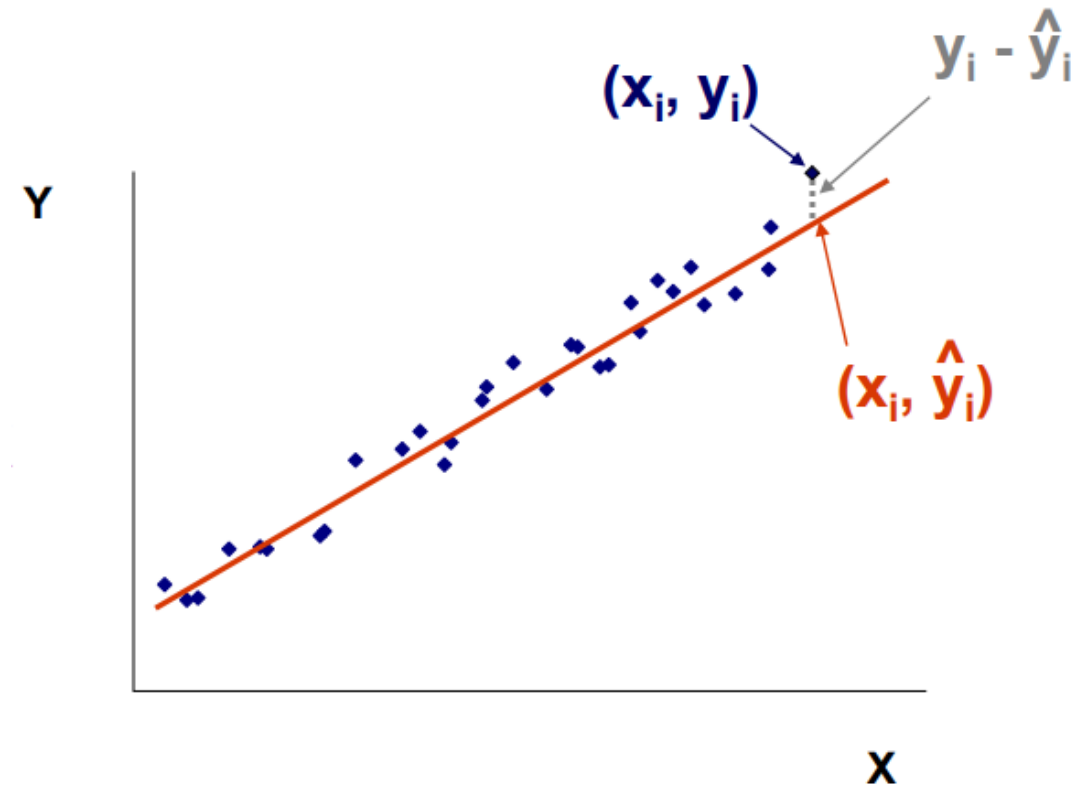
$X$ , la variable explicative

$\varepsilon$ , une variable aléatoire qui représente l'erreur.

# La régression linéaire

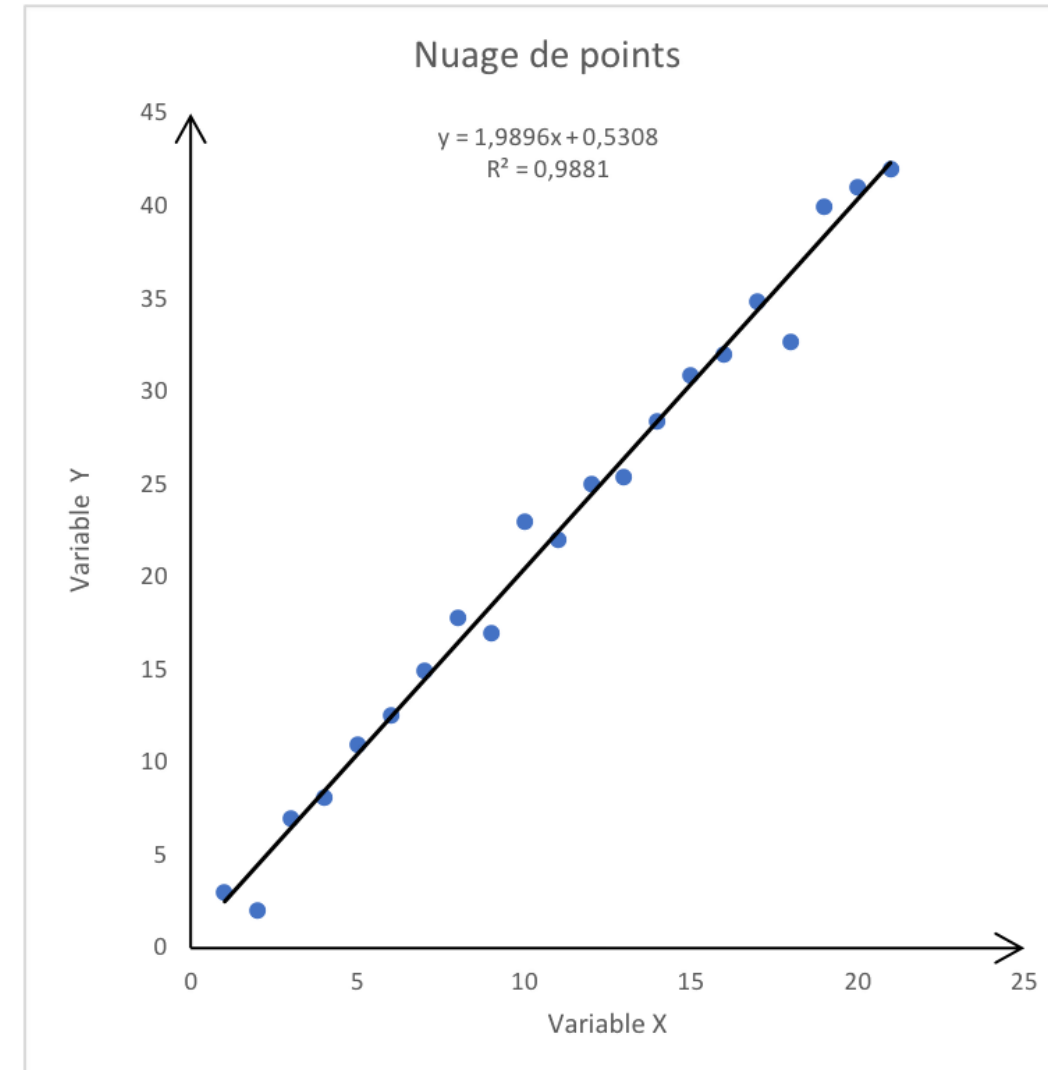
- L'objectif est d'estimer une fonction  $Y = f(X) + \varepsilon$  qui représente la relation entre  $Y$  et  $X$  afin de prédire la valeur  $\hat{y}_i$  pour une valeur de  $x_i$  quelconque.
- $\varepsilon$  est appelé résidus c'est l'erreur commise, c'est-à-dire l'écart entre la valeur  $y_i$  observée et la valeur  $a x_i + b$  donnée par la relation linéaire.
- En effet, même si une relation linéaire est effectivement présente, les données mesurées ne vérifient pas en général cette relation exactement. En statistique une méthode de mesure est les **Moindres Carrés Ordinaires** que l'on note RSS :

$$\text{RSS}(f) = \sum_{i=0}^n (y_i - f(x_i))^2 = \sum_{i=1}^n (\varepsilon_i)^2$$



# R-squared

- Sur cet exemple, la droite de régression linéaire ou la droite des moindres carrés de  $Y$  en  $X$  représente la droite d'ajustement linéaire qui résume le mieux la structure du nuage de points pendant la phase d'apprentissage. Elle rend minimale la somme des carrés des erreurs d'ajustement.
- **$R^2$  (R-squared)** représente le coefficient de corrélation de Bravais-Pearson au carré. Ce coefficient mesure l'intensité de la relation linéaire entre  $Y$  et  $X$ .
- **$R^2$  (R-squared)** est appelé aussi **coefficient de détermination**.



# R-squared

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{Y})^2}$$

- Le **coefficient de détermination** (1- (la somme des erreurs / la somme de la variance de la moyenne ))est un nombre toujours compris entre 0 et 1.
- Si proche de 1 : **Bonne modélisation** (on capture presque toute la variance)
- Si proche de 0 : **Mauvaise modélisation** (la variance n'est pas capturée)

# La régression linéaire

- La régression linéaire simple consiste à prendre pour hypothèse que la relation  $f$  est un **polynôme** de degré 1 de  $X$  :  $f(X) = aX + b$
- Ce qui nous donne :  $RSS(f) = RSS(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$
- $\mathbb{P} = \{a, b\}$  est l'ensemble des paramètres du modèle et on cherche les estimations  $\hat{a}$  et  $\hat{b}$  qui minimisent **RSS**.
- Minimiser cette expression revient à résoudre un problème d'optimisation, on obtient une solution analytique pour les estimateurs  $\hat{a}$  et  $\hat{b}$ :

$$\hat{a} = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{X})^2} \quad \text{et} \quad \hat{b} = \bar{Y} - \hat{a}\bar{X}$$

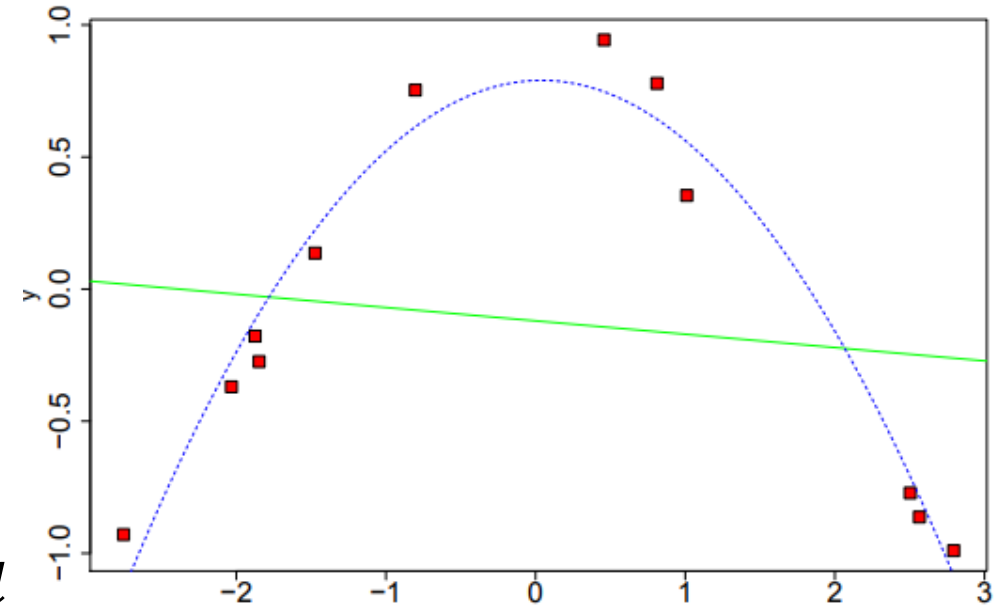
- Le **modèle de prédiction** est alors donné par :  $\hat{f}(x) = \hat{a}x + \hat{b}$

# Polynomial Regression

---

# Pourquoi se limiter à des lignes droites?

- Les relations ne sont pas toutes linéaires.
- Formule linéaire:  $Y = ax + b$  (un polynôme du premier ordre)
- Polynôme de second ordre:  $Y = ax^2 + bx + c$
- Polynôme de 3ème ordre:  $Y = ax^3 + bx^2 + cx + d$
- Des ordres plus élevés produisent des courbes plus complexes.



# La régression polynomiale

- On peut considérer la régression polynomiale comme un cas généralisé de régression linéaire.
- **Remarques:**
  - Il ne faut pas un degré élevé si vous n'en avez pas besoin.
  - Visualisez d'abord vos données pour voir à quel point la courbe est complexe
  - Visualisez l'ajustement (fitting)– La courbe se met-elle hors place pour s'adapter aux Outliers?
  - Un R-squared élevé signifie simplement que votre courbe correspond bien à vos données d'entraînement; mais ce n'est peut-être pas un bon prédicteur.
  - Plus tard, on donnera autres mesures de performances pour évaluer un modèle.



# Multiple regression

---

# Et si plus d'une variable influence celle qui nous intéresse?

## Exemple:

- Prédire un prix pour une voiture en fonction de ses nombreux attributs (style de carrosserie, marque, kilométrage, etc.)
- On se retrouve juste avec des coefficients pour chaque facteur.

$$Prix = \alpha + \beta_1 \text{ kilométrage} + \beta_2 \text{ age} + \beta_3 \text{ Nombre\_Portes}$$



# La régression multiple

- Étant donné un échantillon  $(Y_i, X_{i1}, \dots, X_{ip})$ , on cherche à expliquer, avec le plus de précision possible, les valeurs prises par  $Y_i$ , à partir d'une série de variables explicatives  $X_{i1}, \dots, X_{ip}$ . Le modèle théorique, formulé en termes de variables aléatoires, prend la forme:  $Y_i = a_0 + a_1 X_{i1} + \dots + a_p X_{ip} + \varepsilon_i$
- où  $\varepsilon_i$  est l'erreur du modèle qui résume, l'information manquante dans l'explication linéaire des valeurs de  $Y_i$  à partir des  $X_{i1}, \dots, X_{ip}$ . Les coefficients  $a_0, a_1, \dots, a_p$  sont les paramètres à estimer.
- Remarques:
  - Ces coefficients impliquent l'importance de chaque facteur (si les données sont toutes normalisées!). Il faut enlever ceux qui ne comptent pas!
  - On peut mesurer aussi l'ajustement (fitting) avec le R-squared