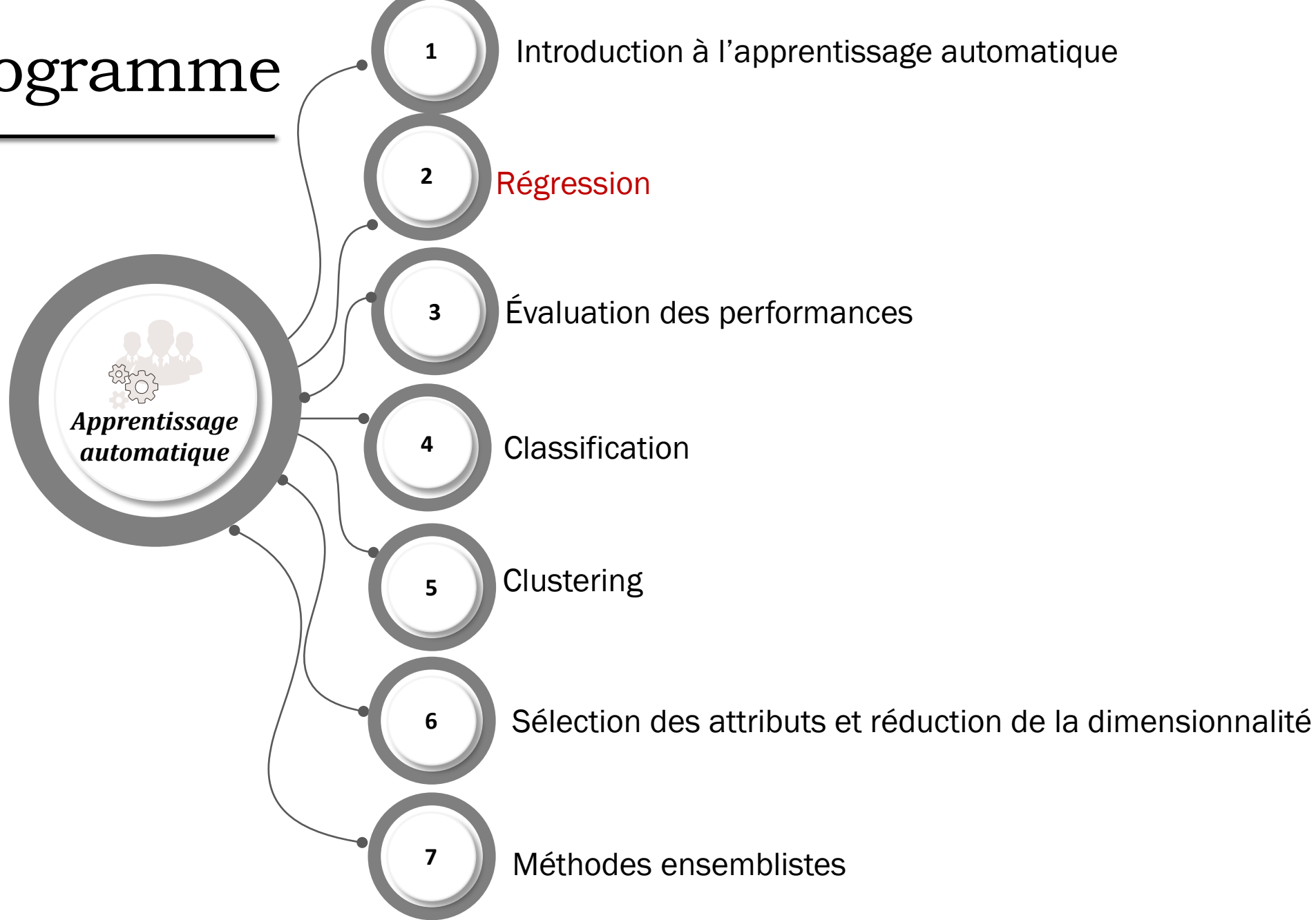


Apprentissage Automatique

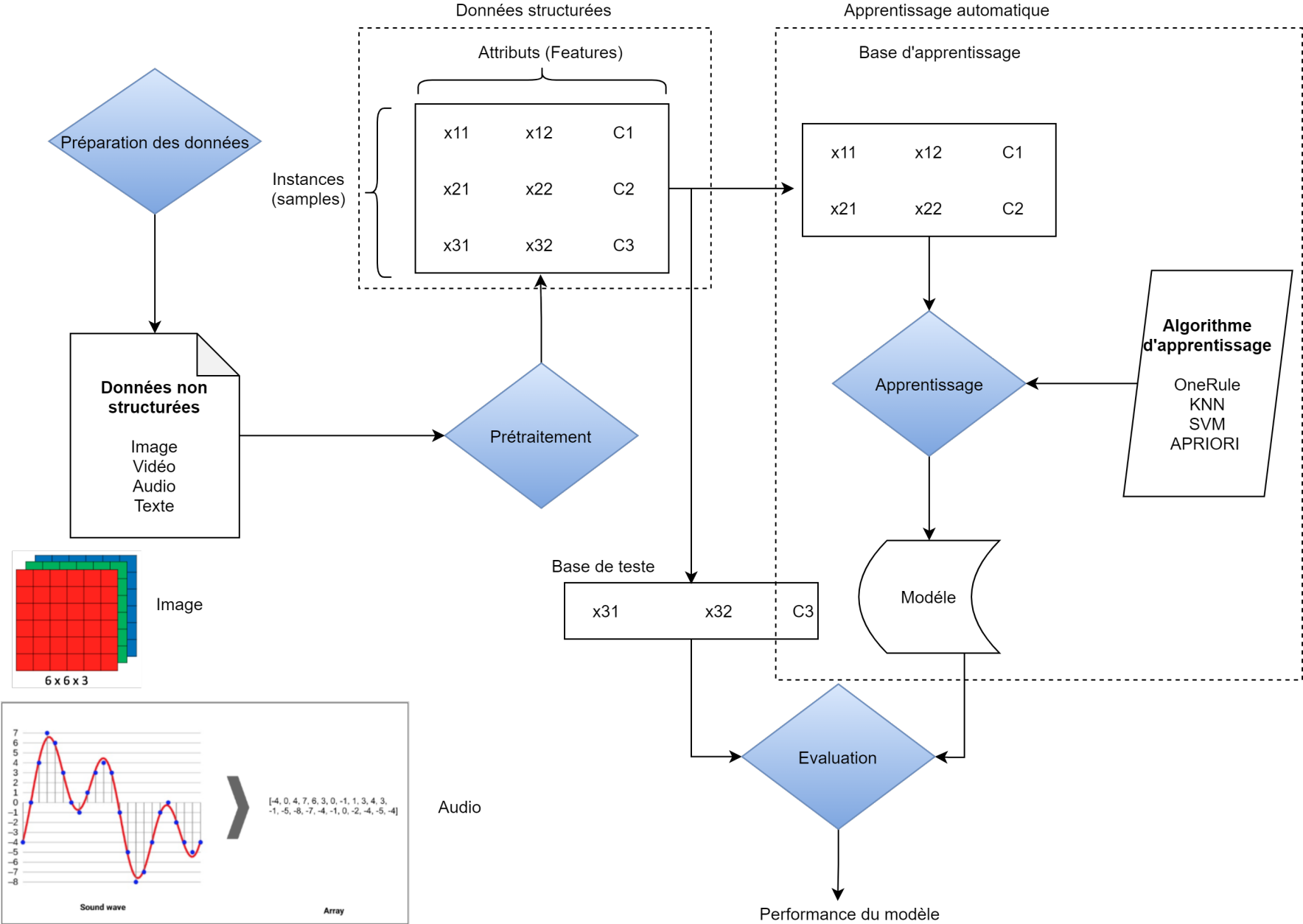
*Intelligence Artificielle et
Sciences de Données
(IASD)*

DR N. DIF

Programme



Rappel



1. LA RÉGRESSION LINÉAIRE ET NON LINÉAIRE



Let's see first a real-world application of regression

Véhicules autonomes



Deviner (prédire) l'angle approprié à chaque instant t



$$f\left(\text{image_road}\right) = -116.60^\circ$$

Pour plus de détails, voir :

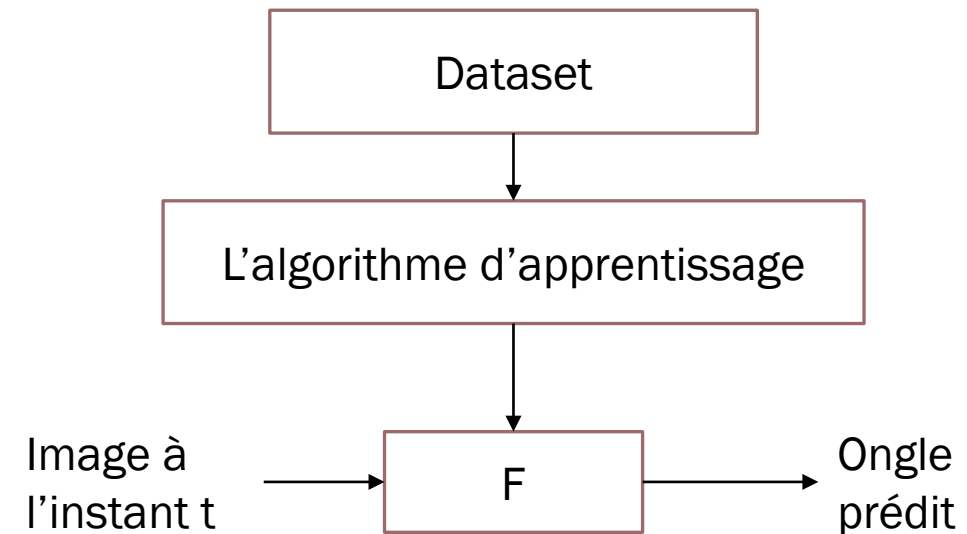
<https://www.youtube.com/watch?v=H0igiP6Hg1k>

1. LA RÉGRESSION LINÉAIRE ET NON LINÉAIRE



Let's see first a real-world application of regression

DATA	
X (INPUT DATA)	y (TARGET OUTPUT DATA)
	0.86°
	-99.82°
	-144.20°



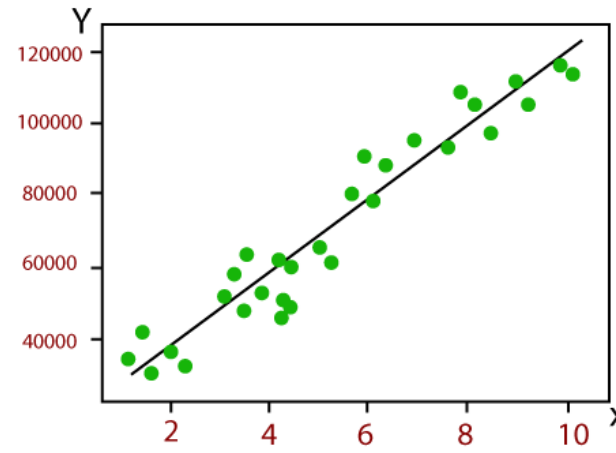
1. LA RÉGRESSION LINÉAIRE ET NON LINÉAIRE



Comment peut on représenter la fonction F qui représente le modèle dans notre cas ? Comment peut on tirer une connaissance à partir des instances d'apprentissage?

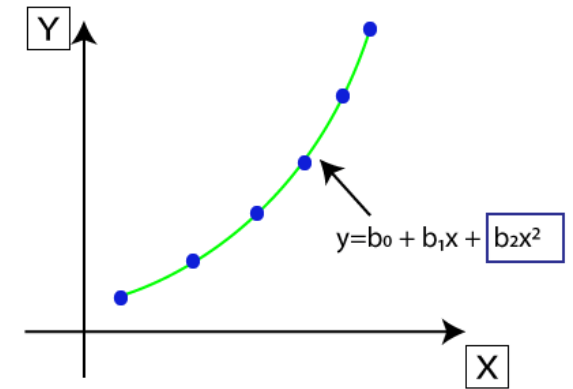
1. LA RÉGRESSION LINÉAIRE ET NON LINÉAIRE

- Méthode statistique d'apprentissage supervisé
- Les modèles de régression sont conçus spécialement pour les valeurs numériques.
- La classe est représentée sous une forme numérique continue (exemple : température, âge, salaire, prix)
- Elle permet d'expliquer la relation entre une variable dépendante (cible) Y et une ou plusieurs variables indépendantes (prédicteur) X_j ($j=1, \dots, q$) de type numérique. La relation exprimée entre les deux variables peut être linéaire ou non linéaire.



Régression linéaire

Une ligne droite qui sépare entre les différents points de données.



Régression non linéaire

Un graphe séparant entre les différents points de données.

1. LA RÉGRESSION LINÉAIRE ET NON LINÉAIRE

- Le choix du type de régression est effectué en fonction de la nature de la dataset, dans certaines situations, impossible de séparer entre les points de données de dataset par une simple ligne droite, dans ce cas, la régression non linéaire est utilisée.
 1. **Régression linéaire** : estimation du loyer en fonction de la surface d'un appartement, estimation du salaire.
 2. **Régression non linéaire** : estimation de la croissance de la population dans le temps.
- La régression linéaire est catégorisée en régression linéaire simple et multiple.
 1. **Régression linéaire multiple** : utilise plusieurs variables indépendantes ($Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \varepsilon_t, t = 1, 2, \dots, m$)
 2. **Régression linéaire simple** : utilise une seule variable indépendantes ($Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, t = 1, 2, \dots, m$),
 Y_t est la variable dépendante à la date t,
 X_t est la variable indépendante à la date t, β_i sont les paramètres inconnus du modèle, ε_t est l'erreur aléatoire du modèle, m est le nombre d'observations, et k est le nombre de variables indépendantes.

2. LA RÉGRESSION LINÉAIRE SIMPLE

- Le modèle de régression linéaire simple est une variable endogène (dépendante) expliquée par une seule variable exogène (indépendante).
- L'estimation des paramètres β_0 et β_1 est faite par la méthode des moindres carrés ordinaires (MCO).
- L'estimation est obtenue en minimisant la somme des carrés des erreurs suivant :

$$\text{Min} \sum_{t=1}^m \varepsilon_t^2 = \text{Min} \sum_{t=1}^m (Y_t - \beta_0 - \beta_1 X_t)^2$$

Pour avoir un minimum, il faut les dérivées par rapport à β_0 et β_1 soient nuls

2. LA RÉGRESSION LINÉAIRE SIMPLE

$$\frac{\partial S}{\partial \beta_0} = 0 \Leftrightarrow 2 \sum_{t=1}^m (Y_t - \beta_0 - \beta_1 X_t)(-1) = 0 \rightarrow \sum_{t=1}^m Y_t = m\beta_0 + \beta_1 \sum_{t=1}^m X_t \quad \dots (1)$$

$$\frac{\partial S}{\partial \beta_1} = 0 \Leftrightarrow 2 \sum_{t=1}^m (Y_t - \beta_0 - \beta_1 X_t)(-X_t) = 0 \rightarrow \sum_{t=1}^m Y_t X_t = \beta_0 \sum_{t=1}^m X_t + \beta_1 \sum_{t=1}^m X_t^2 \quad \dots (2)$$

De (1) on obtient

$$\begin{cases} \widehat{\beta_0} = \frac{\sum_{t=1}^m Y_t - \widehat{\beta_1} \sum_{t=1}^m X_t}{m} \\ \widehat{\beta_0} = \bar{Y} - \beta_1 \bar{X} \end{cases}$$

En remplaçant β_0 dans l'équation (2) on obtient :

$$\begin{cases} \widehat{\beta_1} = \frac{\sum_{t=1}^m X_t Y_t - \bar{Y} \sum_{t=1}^m X_t}{\sum_{t=1}^m X_t^2 - \bar{X} \sum_{t=1}^m X_t} \\ \widehat{\beta_1} = \frac{\sum_{t=1}^m X_t Y_t - \bar{Y} \sum_{t=1}^m X_t}{\sum_{t=1}^m X_t^2 - m \bar{X}^2} \\ \widehat{\beta_1} = \frac{\sum_{t=1}^m X_t Y_t - m \bar{X} \bar{Y}}{\sum_{t=1}^m X_t^2 - m \bar{X}^2} \\ \widehat{\beta_1} = \frac{\sum_{t=1}^m (Y_t - \bar{Y})(X_t - \bar{X})}{\sum_{t=1}^m (X_t - \bar{X})^2} \end{cases}$$

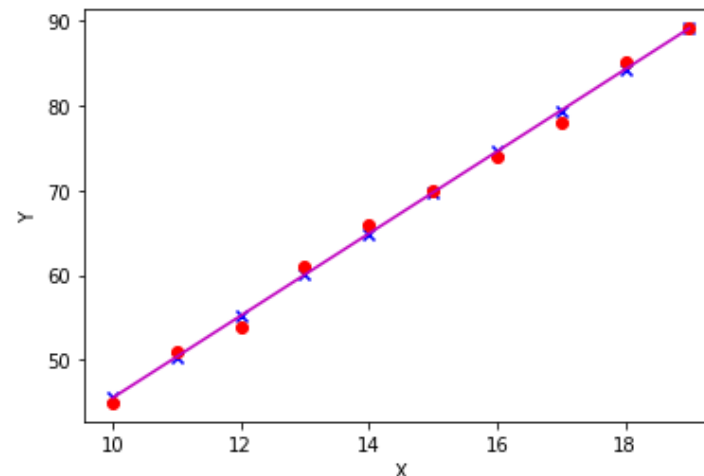
2. LA RÉGRESSION LINÉAIRE SIMPLE

Pour l'ensemble des données suivantes, déterminez quelle valeur espérée pour $X=21$.

X	10	11	12	13	14	15	16	17	18	19	145
Y	45	51	54	61	66	70	74	78	85	89	-
XY	450	561	648	793	924	1050	1184	1326	1530	1691	10157
X ²	100	121	144	169	196	225	256	289	324	361	2185
\bar{X}	14.5				\bar{Y}	67.3					

$$\begin{cases} \widehat{\beta}_1 = \frac{10157 - 67.3 * 145}{2185 - 10 * (14.5)^2} = 4.8303 \\ \widehat{\beta}_0 = 67.3 - 4.8303 * 14.5 = -2.7394 \end{cases}$$

Le modèle prédit est caractérisé par une bonne performance, car la distance entre les valeurs prédites et réelles est minimale.



Modèle de régression : mauve.
Les valeurs prédites : croix bleue.
Les valeurs réelles : bulle rouge.

3. LA RÉGRESSION LINÉAIRE MULTIPLE

La régression linéaire simple permet de résoudre seulement les problèmes linéairement séparable et qui sont caractérisé seulement par un seul attribut ou caractéristique, par exemple: prédire le prix d'une maison en fonction d'une superficie. Mais généralement, une dataset est caractérisé par plusieurs attributs, par exemple dans le cas de prédiction des prix de maison en fonction de la superficie et la nature de la maison, dans ce cas, on fait appel aux algorithmes de régression linéaire multiple.

Le numéro de la maison est t'il nécessaire pour la prédiction? Pourquoi?

Sr. No.	Details	Price	Bedrooms	Bathrooms
1	23534368	221456	3	2
2	89756456	321234	4	3
3	45767857	134000	2	2
4	25756756	214679	3	1
5	23445466	213245	3	1

3. LA RÉGRESSION LINÉAIRE MULTIPLE

Le problème cité précédemment est caractérisé par 3 caractéristiques (details ou superficie (X_1 , Bathrooms X_2 , and Bedrooms X_3 , price Y):

- $F(x) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3$ pour généraliser $F(x) = \theta^T X = \theta_0 + \theta_1 X_1 + \dots + \theta_n X_n$, où

n est le nombre de caractéristiques.

m est le nombre d'instances.

Les paramètres à fixer : $\theta_0, \theta_1, \dots, \theta_n$

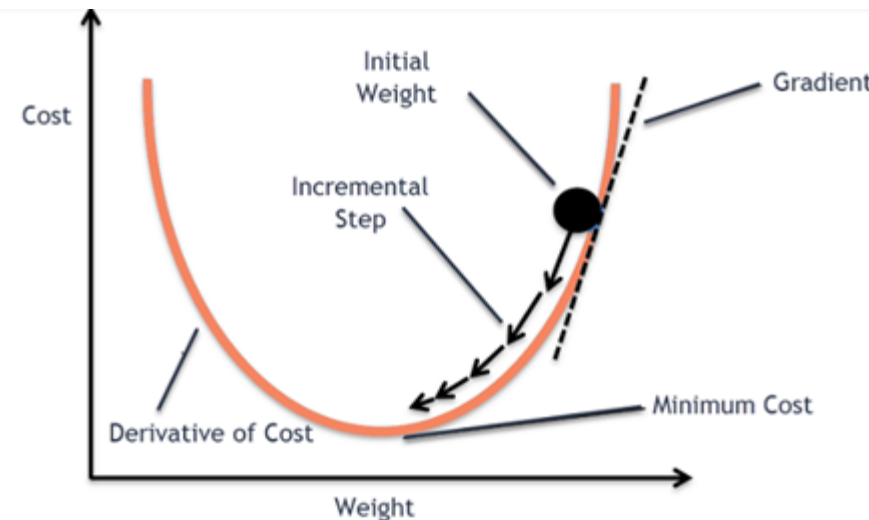
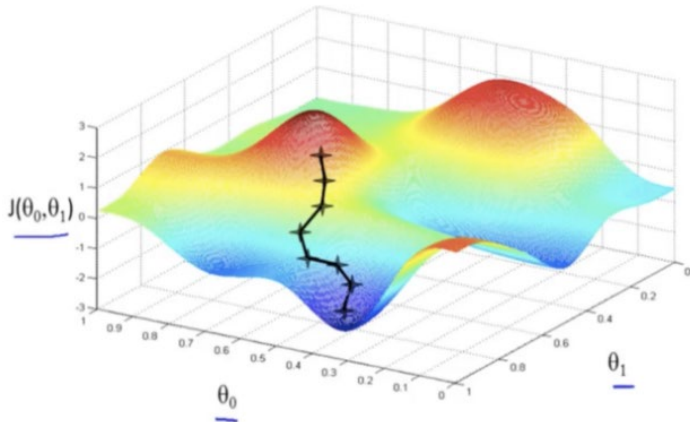
- Le but : prédire les paramètres $\theta_0, \theta_1, \dots, \theta_n$ en minimisant la **fonction de cout** :
- $J(\theta_0, \theta_1, \dots, \theta_n) = \frac{1}{2m} \sum_{i=1}^m (f(X^{(i)}) - Y^{(i)})^2$
- Parmi les méthodes utilisées pour prédire l'ensemble des paramètres $\theta_0, \theta_1, \dots, \theta_n$ est la méthode de **descente de gradient**.

3. LA RÉGRESSION LINÉAIRE MULTIPLE

La descente de gradient est un algorithme d'optimisation itératif stochastique qui permet de trouver le minimum d'une fonction.

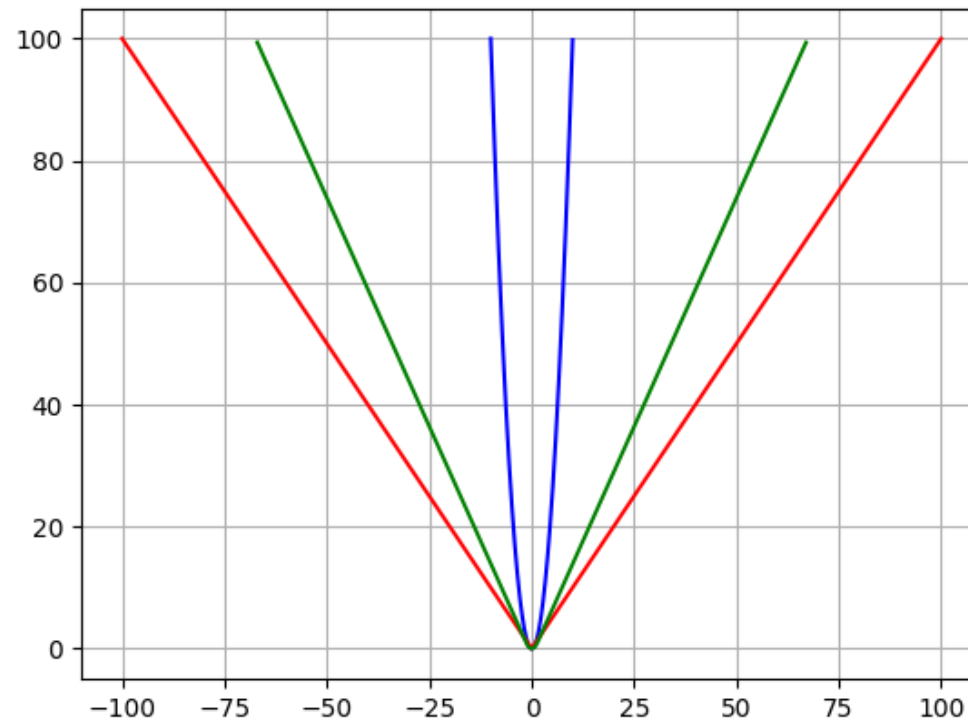


Quelle est l'inspiration de la descente de gradient? Pourquoi l'utilisation de la dérivée partielle ? C'est quoi un minimum local et global?



3. LA RÉGRESSION LINÉAIRE MULTIPLE

Loss functions



MAE (red), MSE (blue), and Huber (green) loss functions

3. LA RÉGRESSION LINÉAIRE MULTIPLE

Dans la méthode de descente de gradient : la mise à jour des paramètres s'effectue de la manière suivante :

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta), j \in (1, 2, \dots, n)$$

$$\theta_j = \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (F(x^{(i)}) - y^{(i)}) X_j^{(i)}, \theta_0 = \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (F(x^{(i)}) - y^{(i)})$$

J est la fonction de coût et n est $\frac{\partial}{\partial \theta_j} J(\theta)$ est la dérivée partielle sur θ_j , α est le taux d'apprentissage.

Algorithme : descente de gradient pour la régression linéaire multiple

Initialiser aléatoirement θ_j

Répéter {

Mettre à jour simultanément θ_j pour $j \in (1, 2, \dots, n)$

}

**Quand faut t'il s'arrêter ?
Comment paramétrer le taux
d'apprentissage ? C'est quoi
son intérêt**

3. LA RÉGRESSION LINÉAIRE MULTIPLE



Cette méthode de descente de gradient est elle toujours efficace ?

Descente de gradient

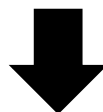
Dans chaque itération, cette méthode considère la moyenne des gradients de tous les instance de la dataset.



Couteuse en terme de temps d'exécution et de stockage

Descente de gradient stochastique

Dans chaque itération, cette méthode considère seulement un exemple sélectionné aléatoirement de la dataset.



Peut causer un problème d'oscillation

Descente de gradient par lot

Dans chaque itération, cette méthode utilise un lot de données choisis de la base d'apprentissage.

Descente de gradient stochastique par lot

3. LES MESURES D'ÉVALUATION EN REGRESSION

- **Mean Absolute Error (MAE):** l'erreur absolue moyenne est la moyenne de la différence entre les valeurs réelles et les valeurs prédites. Cette mesure est à minimiser.
- L'équation suivante illustre le calcul de la MAE, où, n est le nombre des observations, y_j est la valeur réelle de l'observation j et \hat{y}_j est la valeur prédite.

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

- **Root Mean Squared Error (RMSE):** l'erreur quadratique moyenne est la mesure d'évaluation la plus utilisée en régression. L'équation suivante présente le calcul de cette mesure. Contrairement à la MAE, cette mesure donne plus de poids aux grandes erreurs. RMSE a l'avantage de pénaliser davantage les grosses erreurs par rapport à la MAE.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

3. LES MESURES D'ÉVALUATION EN REGRESSION

Exemple : Le tableau suivant présente les erreurs RMSE et MAE d'un système d'estimation de la distance de freinage d'une voiture.

	Erreur 1	Erreur 2	MAE	RMSE
Modèle A	10	0	5	7
Modèle B	6	5	5.5	5.52

- En termes de MAE le modèle A est plus performant, alors qu'il a fait une erreur de 10 mètres, et cela peut engendrer des dégâts dans ce genre de systèmes, par contre le modèle B est plus performant en termes de RMSE.
- Dans certaines situations les grandes erreurs sont provoquées par des valeurs aberrantes (outliers), dans ce cas, il vaut mieux tourner vers la MAE car elle donne une meilleure représentation.

Bibliographie

- Adrew Ng. Learning: Regression and Classification, Coursera. <https://fr.coursera.org/learn/machine-learning>
- Welch Labs. Self Driving Cars [S1E2: ALVINN]. <https://www.youtube.com/watch?v=H0igiP6Hg1k>
- BOUKRIF Nouara. Polycopié : *Régression Linéaire simple et multiple*.
- Ritchie Ng. Linear Regression with Multiple Variables. <https://www.ritchieng.com/multi-variable-linear-regression/>
- Matthew MacFarquhar. Gradient Descent: The Magic Behind ML. <https://blog.devgenius.io/gradient-descent-the-magic-behind-ml-6dc17668d0af>