

<p style="text-align: center;">Enseignants</p> <p>Cours : A. NAJJAR-I. BEN OTHMEN-M. FARHAT TP : F. JENHANI- S.BIROUZA – I. HAMROUNI</p>	<p style="text-align: center;">TP3 Machine Learning</p>	<p style="text-align: center;">Classe : 3ème GLSI</p>
--	---	--

Partie 1 : Régression Linéaire Simple

On souhaite effectuer de la **régression linéaire simple**. Dans cette partie du TP, on se basera sur la base "**Weather.csv**" contenant des données météorologiques.

- 1- Importer les données à partir du fichier.
- 2- Afficher la dimension de l'ensemble de données et en déduire le nombre des variables utilisées.
- 3- Afficher les noms de ces variables.

Pour faire de la régression linéaire simple, on se restreindra à deux variables : la valeur minimale de température (**MinTemp**) et sa valeur maximale (**MaxTemp**). On considère que "**MinTemp**" est la variable explicative et "**MaxTemp**" la variable expliquée. Par conséquent, on souhaite prédire la valeur de "**MaxTemp**" en fonction de la valeur enregistrée de "**MinTemp**".

- 4- Afficher toutes les observations de la base en considérant la valeur minimale de la température sur l'axe des abscisses et sa valeur maximale sur l'axe des ordonnées.
- 5- Construire le nouvel ensemble de données, composé seulement des deux variables "**MinTemp**" et "**MaxTemp**".
- 6- Diviser l'ensemble de données en un ensemble d'apprentissage qui contient 80% des observations et un ensemble de test (20%).
- 7- Déterminer les paramètres "**a**" et "**b**" de la droite de régression ($y=ax+b$) en utilisant l'ensemble d'apprentissage.
- 8- Afficher les paramètres du modèle : la pente "**a**" et l'ordonnée à l'origine "**b**".

- 8- Déterminer et afficher le **coefficient de détermination R^2** . Conclure.
- 9- En utilisant les températures minimales de l'ensemble de test et le modèle de régression calculé dans la question précédente, prédire les températures maximales.
- 10- Afficher, sur la même figure, les valeurs réelles observées et celles prédites des températures maximales en fonction des températures minimales, et ce à partir de l'ensemble de test.

Partie 2 : Régression Linéaire Multiple

On souhaite effectuer de la **régression linéaire multiple**. Dans cette partie du TP, on se basera sur la base " **FuelConsumptionCo2.csv** " contenant des données autour de l'émission CO2.

1. Faites les importations nécessaires pour charger une dataset, la manipuler, la visualiser et utiliser un modèle d'apprentissage.
2. Faites l'importation de la dataset **FuelConsumptionCo2.csv**.
3. Affichez la structure de la dataset.
4. Nous allons utiliser uniquement les colonnes: 'ENGINESIZE', 'CYLINDERS', 'FUELCONSUMPTION_CITY', 'FUELCONSUMPTION_HWY', 'FUELCONSUMPTION_COMB', 'CO2EMISSIONS'. Créer le dataframe correspondant.
5. Faites un affichage en nuage de points de l'émissionCO2 en fonction de ENGINESIZE.
6. Séparons notre ensemble de données en ensembles d'entraînement et de test. Environ 80 % de l'ensemble des données seront utilisés pour la formation et 20 % pour les tests. Nous créons un masque pour sélectionner des lignes aléatoires à l'aide de la fonction `np.random.rand()`.
7. Utiliser le modèle de la régression linéaire multiple pour prédire l'émission CO2 en fonction des variables explicatives: 'ENGINESIZE', 'CYLINDERS', 'FUELCONSUMPTION_COMB'. Pour ce faire, utiliser la méthode `LinearRegression()` de la bibliothèque `SicktLearn`.
8. Afficher les coefficients.
9. Maintenant, vous allez procéder au test moyennant la prédiction sur de nouvelles échantillons. Utiliser la méthode `predict()`.
10. Afficher le coefficient R^2 .
11. Essayez d'utiliser une régression linéaire multiple avec le même ensemble de données, mais cette fois, utilisez `FUELCONSUMPTION_CITY` et `FUELCONSUMPTION_HWY` au lieu de `FUELCONSUMPTION_COMB`. Auriez-vous une meilleure précision ?