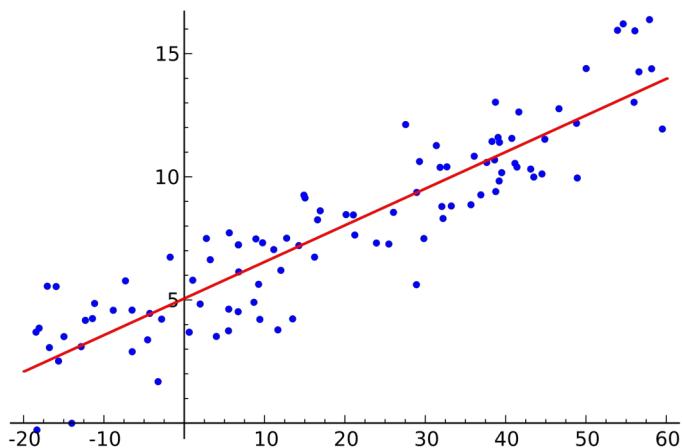


Atelier 1 « Régression »

Objective :

l'objectif principal de cet atelier est de pratiquer les deux concepts de la régression : la régression linéaire simple et la régression linéaire multiple, en traitant des données de plusieurs Data Sets.



Outils : Python , Pandas, Sklearn, matplotlib.

Data Sets :

Expérience et Salaire : <https://www.kaggle.com/rohankayan/years-of-experience-and-salary-dataset>

Assurance : <https://www.kaggle.com/sinaasappel/tutorial-multiple-regression/data>

China GDP :https://s3-api.us-geo.objectstorage.softlayer.net/cf-courses-data/CognitiveClass/ML0101ENv3/labs/china_gdp.csv

Partie 1 (Data Visualisation):

1. En utilisant pandas essayer d'explorer les données des deux Data sets.
2. Afficher le résumé statistique des deux Data Sets avec une interprétation des résultats obtenues.



3. Afficher la nuage des points du premier data set « Expérience / Salaire » en utilisant matplotlib et pandas , faire une conclusion.
4. Afficher les nuages des points du deuxième data set selon les propriétés « Features » en utilisant matplotlib et pandas « scatter_matrix ».

Partie 2 « Régression Simple cas Expérience Salaire »:

Dans cette partie on va utiliser le Data Set Expérience Salaire .

1. en utilisant l'API sklearn entraîner le modèle par intermédiaire de algorithme de la régression linéaire.
2. prédire les données d'un data set de test.
3. Visualiser le résultat de la régression sous forme d'un graphe.
4. Évaluer le modèle en utilisant ces trois méthodes :

Mean Squared Error (MSE)
Root Mean Squared Error (RMSE)
Mean Absolute Error (MAE)

Interpréter le résultat de l'évaluation.

partie 3 « Régression multiple cas d'assurance »:

Dans cette partie on va utiliser le Data Set Assurance .

1. Appliquer les techniques EDA puis donner des conclusions sur les données.
2. Appliquer des techniques pour sélectionner 3 propriétés selon leurs dégrée d'importance justifier le choix.
3. Appliquer des technique de standardisation ou bien de normalisation sur les propriétés choisit, justifier le choix.
4. en utilisant l'API sklearn entraîner le modèle par intermédiaire de algorithme de la régression linéaire.
5. prédire les données d'un data set de test.
6. Visualiser le résultat de la régression sous forme d'un graphe.
7. Évaluer le modèle en utilisant ces trois méthodes :

Mean Squared Error (MSE)
Root Mean Squared Error (RMSE)
Mean Absolute Error (MAE)

Interpréter le résultat de l'évaluation.



partie 4 « Régression linière polynomial multiple cas de china GDP»:

1. en utilisant l'API sklearn entraîner le modèle par intermédiaire de algorithme de la régression linière et puis la régression linière polynomiale .
 2. prédire les données d'un data set de test pour les deux modèles .
 3. Visualiser le résultat de la régression sous forme d'un graphe des deux modèles.
 4. Évaluer les deux modèles en utilisant ces trois méthodes :
Mean Squared Error (MSE)
Root Mean Squared Error (RMSE)
Mean Absolute Error (MAE)
- Interpréter le résultat de l'évaluation.

Note : un rapport bien détailler doit être rédige et imprimé.

Références :

- <https://becominghuman.ai/implementing-and-visualizing-linear-regression-in-python-with-scikit-learn-a073768dc688>
- <https://www.kaggle.com/sinaasappel/tutorial-multiple-regression/data#Tutorial---Multiple-Regression>
- <https://stackabuse.com/multiple-linear-regression-with-python/>
- <https://sweetcode.io/simple-multiple-linear-regression-python-scikit/>