

Machine Learning

Exercice 1

Dans cet exercice, il fallait générer un dataset Artificiel avec 6 colonnes ayant chacune une spécificité unique avec à chaque fois 300 valeurs dans chacune d'entre elle et un écart type différent.

Colonne 1 : Nombres Entier

Il fallait générer des nombres entier, nous avons donc utilisé la fonction **`np.random.randint(0, 9999, 300)`** afin d'avoir des valeurs entre 0 et 9999 pour 300 entités

Colonne 2 : Nombres Décimaux

Il fallait générer des nombres décimaux, nous avons donc utilisé la fonction **`np.random.randn(7.9, 2, 300)`** afin d'avoir des valeurs avec une moyenne de 7.9 et un écart type de 2 et 300 entités

Colonne 3 : Moyenne de 2.5

Il fallait générer des nombres décimaux avec une moyenne de 2.5, nous avons donc utilisé la fonction **`np.random.randn(2.5, 1, 300)`** pour une moyenne à 2.5 et un écart type de 1 afin d'avoir le résultat le plus proche de la moyenne

Colonne 4 : Corrélation proche de 0

Ici, nous utilisons la fonction **`np.random.normal(50, 9, 300)`** pour une moyenne à 50 et un écart type de 9, la corrélation est proche de 0 car les nombres sont générés aléatoirement, ce qui résulte une impossibilité de linéarité dans les résultats.

Colonne 5 : Corrélation positive

Nous venons ***additionner la première et la deuxième colonne*** pour donner une corrélation de $r = 1$, c'est-à-dire que si un résultat augmente, le résultat corrélé aussi.

Colonne 5 : Corrélation négative

Nous venons ***soustraire la première et la deuxième colonne*** pour donner une corrélation de $r = -1$, c'est-à-dire que si un résultat diminue, le résultat corrélé aussi.

Exemple de DataSet généré :

Int	Float	Mean	Zero_correla	Positive_corr	Negative_cor
818	11.15	3.31	51.62	829.15	806.85
8892	4.93	1.84	43.04	8896.93	8887.07
8722	8.42	1.73	48.34	8730.42	8713.58
9173	10.66	2.69	64.27	9183.66	9162.34
8897	6.21	1.6	44	8903.21	8890.79

Source pour la corrélation : <https://blog.nalo.fr/lexique/coefficient-de-correlation/>