

## Semaines 1 et 2 : Régression Linéaire Simple

### Grade Point Average

Le « Grade Point Average » (GPA) ou moyenne pondérée est un nombre représentant la valeur moyenne des notes finales accumulées dans les cours au fil du temps (i.e. Trimestre, semestre).

Plus couramment appelée **GPA**, la moyenne pondérée cumulative est calculée en additionnant toutes les notes finales accumulées et en divisant ce chiffre par le nombre de notes attribuées (Moyenne).

Ce calcul donne une moyenne de toutes les notes finales.

La forme la plus courante de moyenne pondérée cumulative est basée sur une *échelle de 0 à 4* (A = 4,0 ; B = 3,0 ; C = 2,0 ; D = 1,0 et F = 0), où une valeur de 4 représente une note parfaite.

Les écoles peuvent également attribuer des points partiels aux notes en utilisant des « plus » ou des « moins » après les lettres, telles que 3,7 pour un A- ou 3,3 pour un B+, etc. Les GPA peuvent être calculées à la fin d'un cours, d'un semestre ou d'un niveau, et une "GPA cumulative" représente la moyenne de toutes les notes finales obtenues par chaque élève depuis la première inscription dans une école jusqu'à la fin de ses études.

Pour un cours noté sur une échelle de 0 à 100%, le tableau ci-dessous résume le classement par lettre et GPA équivalent :

Classement par Lettre	Note en %	GPA
A	94-100	4.00
A-	90-93	3.67
B+	87-89	3.33
B	84-86	3.00
B-	80-83	2.67
C+	77-79	2.33
C	74-76	2.00
C-	70-73	1.67
D+	67-69	1.33
D	64-66	1.00
D-	60-63	0.67
F	0-65	0

Par exemple, si un étudiant a eu les notes suivantes pour ces **4 cours** :

- 68/100 en Maths (Cours 1) ➔ Classement « D+ » ➔ GPA = 1.33
- 87/100 en SVT (Cours 2) ➔ Classement « B+ » ➔ GPA = 3.33
- 54/100 en Physique (Cours 3) ➔ Classement « F » ➔ GPA = 0

- 95/100 en Chimie (Cours 4) → Classement « A » → GPA = 4.00

La moyenne pondérée cumulative (« Cumulative GPA ») sera  $(1.33 + 3.33 + 0 + 4.00)/4 = 2.165$ .

## SAT

Le SAT est un test standardisé largement utilisé pour les admissions dans les universités aux États-Unis. Le SAT est généralement pris par les lycéens. Le *College Board* indique que le SAT a pour objectif de mesurer les compétences en analyse de texte, en calcul et en rédaction nécessaires pour être admis en université.

Le test est réparti en 3 parties (Lecture, Calcul, Écriture) dont chacune est notée sur 800 points. Cela fait un total de 2400 points.

## Objectif

Il est très fréquent que les personnes ayant obtenu un score plus élevé en passant le SAT aient un GPA plus élevé. Étant fourni des échantillons du score SAT et le GPA correspondant de plusieurs étudiants, on s'intéresse à développer un modèle intelligent qui nous permet d'estimer le GPA des étudiants à partir de leur score SAT.

La base de données est disponible ici : <https://www.kaggle.com/luddarell/101-simple-linear-regressioncsv>

- I. Dans cette première partie, il faudra développer un modèle de régression linéaire qui prend en entrée un score SAT et nous retourne une prédiction du GPA en sortie. Voici les étapes à suivre :
  - a. Importer l'ensemble de données (le fichier .csv) et le répartir en 2 sets aléatoires :
    - i. 80% pour la « Training Set »
    - ii. 20% pour la « Test Set ».
  - b. Définir une fonction de Coût :
    - i. On se basera sur L'erreur quadratique moyenne (EQM).

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (\hat{y}_i - y_i)^2 = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x_i) - y_i)^2$$

- c. Implémenter l'algorithme d'apprentissage en se basant sur la descente de gradient.
    - d. Apprenez votre modèle avec la base d'apprentissage
    - e. Tester le modèle sur la base Test et Afficher les valeurs prédites contre les valeurs observées.
- II. Maintenant on va reprendre la 1<sup>ère</sup> partie mais en utilisant des bibliothèques comme Scikit-learn qui nous fournissent ces fonctions d'apprentissage. Pour cela, il faudra :
  - a. Importer le module Scikit-learn
  - b. Reprendre la 1<sup>ère</sup> partie mais en se basant sur les fonctions proposées par le module importé.