

## **TD7 : Algorithmes Régression Linéaire simple et Kmeans**

**Pour que le TD soit pris en considération, rendez-le dans les délais.**

### **Questions de cours :**

- Principe de la régression en général et de la régression linéaire en particulier.
- Principe du clustering en général, ses types, puis celui du K-means en particulier.
- Donner des exemples d'application des 2 algorithmes.

**Exercice 1 :** En utilisant la régression linéaire simple (le critère des moindres carrés) sur la table de données suivante, donner l'équation  $Y = AX + B$ . Quelle valeur est prédite pour  $X = 180$  ?

Valeurs de X (attribut)	155	170	160	185	195	190	175
Valeurs correspondantes observées de Y (classe)	45	70	65	80	85	95	60

**Exercice 2 :** En utilisant l'algorithme K-Means, générer 2 clusters en prenant E7 et E8 comme centres initiaux et la distance de Manhattan, pour les données suivantes :

N°exempl e	ATT 1	ATT 2
E1	6	16
E2	5	17
E3	7	11
E4	5	14
E5	9	18
E6	6	17
E7	11	14
E8	4	10

**Exercice 3 :** En utilisant la régression linéaire simple (le critère des moindres carrés) sur la table de données suivante, donner la fonction  $Y = AX + B$ .

Quelle valeur est prédite pour  $X = 35$  ? **5 points**

Valeur de X (attribut)	11	14	24	21	38	55	42
Valeurs correspondantes observées de Y (classe)	25	40	55	50	79	116	86

**Exercice 4 :** En utilisant la régression linéaire simple (le critère des moindres carrés) sur la table de données suivante, donner l'équation  $Y = AX + B$ . Quelle valeur est prédite pour  $X = 85$  ?

Valeurs de X (attribut)	62	72	68	81	92	94	78
Valeurs correspondantes observées de Y (classe)	11.7	7.8	10.9	9	10.8	13.2	11.6

**Exercice 5 :** Utilisez l'algorithme k-means et la distance euclidienne pour regrouper les 8 exemples suivants en 3 clusters : A1(2,10) A2(2,5) A3(8,4) A4(5,8) A5(7,5) A6(6,4) A7(1,2) A8(4,9).

On considère comme centre de classes à l'initialisation les points A1, A4 et A7.

**Exercice 6 :** Utilisez l'algorithme k-means et la distance manhattan pour regrouper les 8 exemples suivants en 3 clusters : A1(2,10,5) A2(2,5,3) A3(8,4,1) A4(5,8,7) A5(7,5,3) A6(6,4,7) A7(1,2,4) A8(4,9,5).

On considère comme centre de classes à l'initialisation les points A1, A4 et A7.