

---

**TD n° 3**  
*Clustering*

---

**Exercice 1** Soit  $X$  le tableau de données suivant :

X	1	2	9	12	20
---	---	---	---	----	----

1. Appliquez l'algorithme K-means en utilisant la distance euclidienne, avec les valeurs de  $k$  et les points de départ suivants :
  - (a)  $k = 2, \mu_1 = 1, \mu_2 = 20$
  - (b)  $k = 3, \mu_1 = 1, \mu_2 = 12, \mu_3 = 20$
  - (c)  $k = 4, \mu_1 = 1, \mu_2 = 9, \mu_3 = 12, \mu_4 = 20$
2. On aimerait maintenant comparer la qualité de ces regroupements. Pour cela, on recommence par regarder l'inertie intra-cluster  $J_w = \sum_{k=1}^K \sum_{i \in C_k} D^2(x_i, \mu_k)$ .
  - (a) Calculer cette valeur pour les 3 regroupements précédents.
  - (b) En utilisant ce critère, quel serait le meilleur regroupement possible ? est-ce que cela vous paraît réaliste ?