

TD 4 de Statistiques Descriptives 2

---

On s'intéresse dans ce TD à la méthode des  $K$ -plus proches voisins.

**Exercice 1** Dans ce problème, vous allez examiner un petit ensemble de données pour mieux comprendre diverses propriétés de la méthode des  $K$ -NN.

Supposons qu'il existe un ensemble de points sur un plan bi-dimensionnel provenant de deux classes différentes.

Les coordonnées de tous les points sont les suivantes.

— Points de la classe Rouge :  $(0,1), (2,3), (4,4)$

— Points de la classe Bleu :  $(2,0), (5,2), (6,3)$

1. Dessinez la frontière de décision des  $K$ -plus proches voisins pour  $K = 1$ . Rappelez-vous que la frontière de décision est définie comme la ligne où la classification d'un point test change. Utilisez la distance euclidienne standard entre les points pour déterminer les plus proches voisins.

Le site suivant donne une approche systématique pour tracer les diagramme de Voronoi : <http://villemin.gerard.free.fr/Geometri/Voronoi.htm>.

2. Si la coordonnée  $y$  de chaque point était multipliée par 5, comment serait modifiée la frontière pour  $K = 1$  (dessinez une autre image) ? Expliquez en deux phrases au maximum comment cet effet peut poser des problèmes lorsque vous travaillez avec des données réelles.
3. La limite de décision  $K$ -NN pour  $K = 3$  est représentée dans la page suivante.

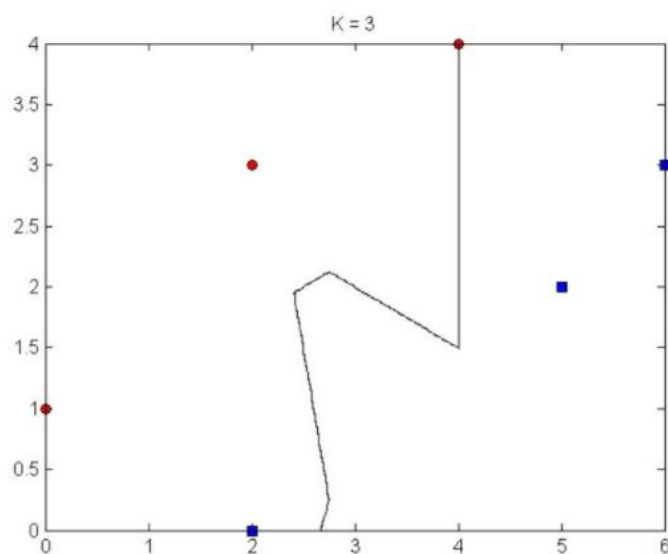


FIGURE 1 – Question 3. de l'Exercice 1.

Supposons maintenant que nous ayons un point test à  $(1, 2)$ . Comment le classer avec 3-NN ? Étant donné que vous pouvez modifier la limite de décision 3-NN en ajoutant des points à l'ensemble d'apprentissage dans le diagramme, quel est le nombre minimum de points que vous devez ajouter pour modifier la classification à  $(1, 2)$  ?