

# Les k plus proches voisins

## Objectifs

Pour ce TP nous allons utiliser l'algorithme des k plus proches voisins pour de la classification.

## Étape 1

Récupérer d'abord le jeu de données (grand classique) : Iris

Commencez par analyser les données.

- Combien d'exemples la base possède-t-elle ?
- Combien de caractéristiques y a-t-il ?
- Combien de classes ?
- De quel type est la caractéristique Species ?
- Combien y a-t-il d'occurrence de chaque type d'iris ?

## Étape 2

La fonction describe() permet d'avoir un résumé de l'objet passé en paramètre.

- Que constatez-vous sur les min/max de chaque caractéristique ?
- Standardiser les 4 caractéristiques afin qu'elles aient le même ordre de grandeur.  $\frac{x_{min}}{x_{max}-x_{min}}$   
Pourquoi est-ce important de faire cette étape pour l'algorithme des k plus proches voisins ?
- Stocker le résultat dans un nouveau data frame irisN

## Étape 3

- Créer les jeux de test et d'apprentissage en respectant les proportions des espèces d'Iris dans le jeu de données de départ.

## Étape 4

- Entraînez un modèle KNN sur le jeu d'apprentissage (valeurs des paramètres par défaut) et mesurer ses performances
- Etudier l'influence de k et jouer également sur la taille des échantillons d'apprentissage et de test.
- Tester le meilleur modèle sur le jeu de test