



Notion de classification supervisée

Mots clefs : Jeu de données, entrées, sorties, classes

Exercice 1

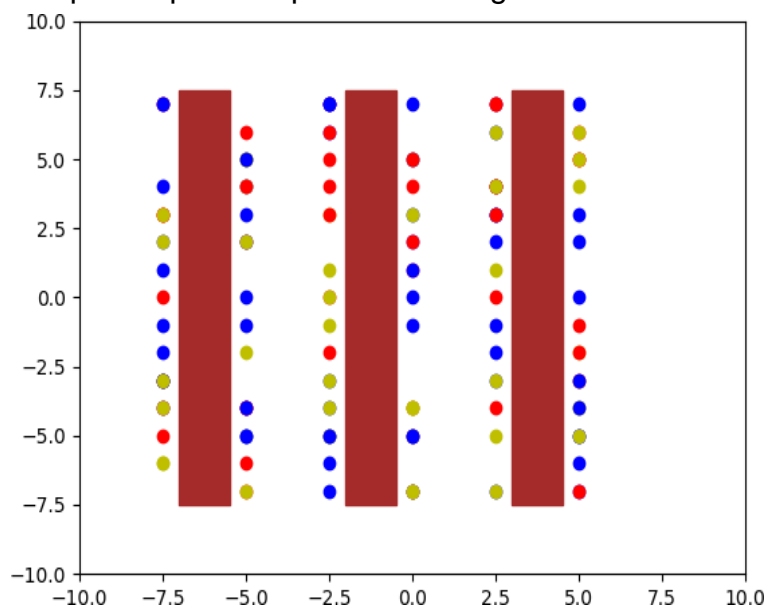
Dans une cantine d'un lycée, les élèves se regroupent en fonction de leur niveau, c'est-à-dire un élève de seconde s'assoit proche d'élèves de seconde, un élève de première s'assoit proche d'élèves de première et ainsi de suite. Un nouvel arrivant s'assoit où il peut mais en essayant d'être proche d'élèves de son niveau. La position des élèves assis est supposée connue. *Dans ce cas, on dit que la classification est supervisée, c'est-à-dire, qu'on connaît la classification des données déjà placée.*



Objectif : prédire la classe, c'est-à-dire le niveau de ce nouvel élève.

Jeu de données : la figure ci-dessous représente trois grandes tables de la cantine, on représente les élèves par des petites disques de couleurs différentes, les verts représentent des élèves de seconde, les rouges ceux de premières les bleus ceux de terminale.

- 1) Placer un nouvel élève dans la figure, représenté par un petit disque noir. En regardant les voisins, prédire le niveau de l'élève.
- 2) En élargissant le cercle des voisins, peut-on affirmer la prédiction faite à la question précédente ?
- 3) **Les entrées :** d'un point de vue mathématique, comment situer les élèves dans le restaurant ?
- 4) Donner les coordonnées du nouvel élève.
- 5) Calculer ou donner la mesure des 5 plus proches voisins.
- 6) **En sortie :** prédire la classe du nouvel élève en regardant les 2 plus proches voisins ? les 5 plus proches voisins ? les 10 plus proches voisins ? Peut-on prédire avec exactitude la classe du nouvel élève ?



Exercice 2

À l'aide d'un logiciel de reconnaissance de forme, un robot arrive à faire une classification des fruits et légumes qu'il "voit". Une fois la reconnaissance terminée, le robot affiche l'image du fruit ou légume observée.

Objectif : comprendre comment la reconnaissance est-elle effectuée ?

Un premier travail est de constituer un jeu de données et de prédéfinir des classes de fruit et légume en précisant quelques attributs, par exemple, la couleur, la forme, la régularité...etc

- 1) On considère cinq fruits-légumes : tomate, concombre, banane, salade et litchi. Ces fruits légumes seront considérés comme des classes. Quels sont les attributs qu'il faut considérer ?

On considère le tableau

	couleurs			forme		régularité	
	rouge	vert	jaune	rond	allongé	lisse	non lisse
Banane							
Concombre							
Litchi							
Salade							
Tomate							

- 2) Compléter le tableau par 0 ou 1

- « 1 » si la caractéristique est présente et « 0 » si elle est absente.

On souhaite mesurer la ressemblance entre deux aliments, on considère la fonction :

$$S(A, B) = \frac{a}{n - b}$$

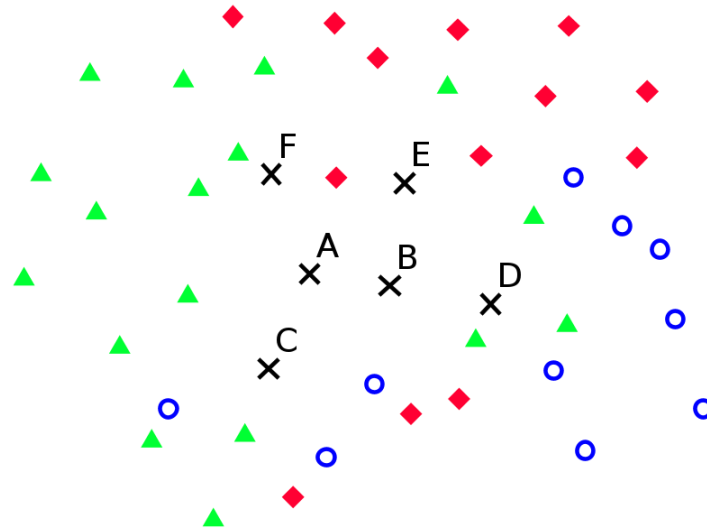
où

- **A et B** : fruit ou légume.
- **a** : le nombre de couples 11 dans le tableau
- **b** : le nombre de couples 00 dans le tableau
- On prend $n = 7$.

Plus le nombre $S(A, B)$ est grand plus les aliments A et B se ressemblent.

- 3) À partir du tableau, **calculer** $S(\text{tomate}, \text{concombre})$, $S(\text{tomate}, \text{salade})$.

Exercice 3 (Extension de la base de données d'apprentissage) On considère une base de données d'apprentissage de points représentée ci-dessous. La classe associée à chaque point est sa forme. Les points à classer sont indiqués par une croix et leur nom.



1. En utilisant l'algorithme du plus proche voisin, déterminer la forme associée à chaque point.

2. Refaire le même travail en utilisant l'algorithme des 5 plus proches voisins.

3. On cherche à faire évoluer la base de données d'apprentissage en ajoutant à celle-ci les points dont on a déjà prédit la classe.

- (a) Appliquer l'algorithme des 5 plus proches voisins en choisissant les lettres selon l'ordre alphabétique.
- (b) Appliquer l'algorithme des 5 plus proches voisins en choisissant les lettres selon l'ordre anti- alphabétique.
- (c) Que peut-on dire des deux nuages de points ?
- (d) Proposer un ordre de choix qui permet de rajouter uniquement des formes triangulaires.

4. Lister les failles de l'algorithme des k plus proches voisins mises en lumière par cet exercice.