

Frontière de séparation des classes, algorithme des kppv.

Exercice 1 *Apprentissage supervisé, frontière*

Question 1. Soit $\mathbf{w} = (w_1, \dots, w_p)$ un vecteur de poids à valeurs dans \mathbb{R}^p utilisé pour prendre une décision linéaire. Donner l'expression permettant de calculer le produit scalaire de chaque exemple de \mathbb{X} avec \mathbf{w} .

Question 2. On se place maintenant en langage Python, on note \mathbb{X} un `numpy.array` qui contient \mathbf{X} et \mathbf{w} un `numpy.array` qui contient \mathbf{w} . Donner les instructions python pour calculer le produit scalaire de tout vecteur de \mathbb{X} par \mathbf{w} , sans utiliser de boucle.

Question 3. On considère maintenant \mathbb{X} et un exemple \mathbf{z} , donner l'expression permettant de calculer la distance euclidienne entre \mathbf{z} et tous les exemples de \mathbb{X} , puis donner le code Python correspondant.

Question 4. Donner les valeurs de d et n correspondant à l'illustration de la figure 1, puis donner les valeurs de \mathbf{w} et b associés au tracé de la frontière [plusieurs valeurs sont possibles, mais l'une des solutions est plus facile à calculer].

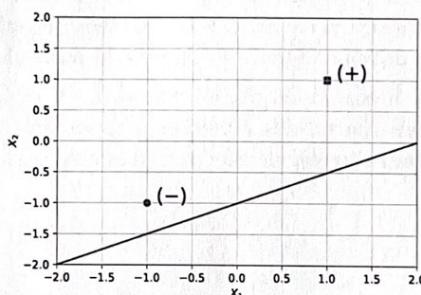


FIGURE 1 – Frontière et base d'apprentissage

Question 5. On considère jeu de données supervisé (\mathbb{X}, \mathbb{Y}) représenté sur la figure 1. Démontrer que les 2 points sont du même côté de la droite.

Exercice 2 *Classification supervisée*

Soit \mathbb{X} une base d'apprentissage contenant n exemples définis par d dimensions et un label associé (les labels sont pris dans $\mathbb{Y} = \{-1, +1\}$). Dans cet exercice, on utilise la distance euclidienne.

Question 1. On décide d'appliquer l'algorithme des k plus proches voisins (kppv) en utilisant \mathbb{X} pour classer un nouvel exemple \mathbf{x} , combien de calculs de distance sont nécessaires au plus ?

Question 2. On considère la base suivante : $\mathbb{X} = \{((1, 2), +1), ((1, 4), +1), ((2, 5), +1), ((4, 3), +1), ((3, 5), +1), ((2, 8), +1), ((3, 2), -1), ((4, 4), -1), ((5, 5), -1), ((4, 7), -1), ((6, 2), -1), ((5, 8), -1)\}$. En appliquant l'algorithme des kppv avec $k = 1$, et en détaillant les calculs réalisés, donner la classe de l'exemple $(3, 6)$.

Question 3. Représenter graphiquement la base de la question précédente et tracer la frontière de séparation des classes lors de l'application des kppv pour $k = 1$.

Question 4. Même question mais pour $k = 3$. En déduire la classe de l'exemple $(3, 6)$ dans ce cas.

Exercice 3 *k plus proches voisins et perceptron*

On considère la base d'apprentissage représentée dans la figure donnée ci-dessous. Cette base contient 20 exemples, dont la description est le couple représenté par leurs coordonnées (x, y) , et la classe est soit *rond* (notée R) soit *carré* (notée C).

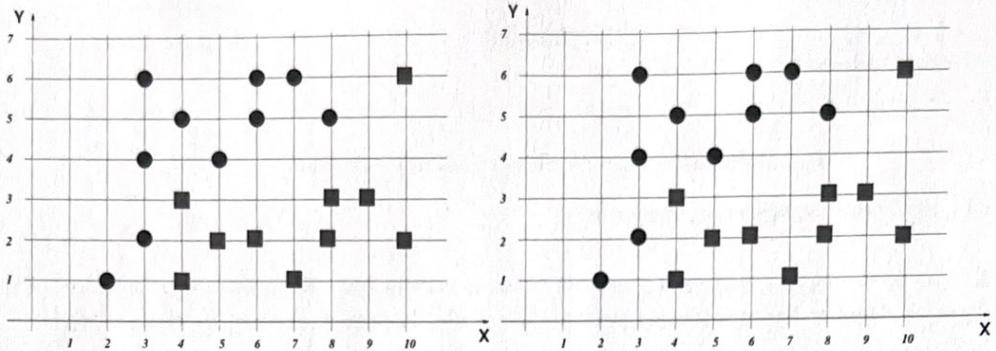


FIGURE 2 – Base d'apprentissage (à gauche pour question 1, à droite pour question 2)

Question 1. En utilisant l'algorithme des k plus proches voisins, avec $k = 3$ et la distance euclidienne, représenter graphiquement la frontière de séparation des classes et donner, en justifiant, la classe des 5 points suivants : le point L de coordonnées $(3, 3)$, M de coordonnées $(10, 3)$, N de coordonnées $(7, 4)$, P de coordonnées $(9, 5)$, et Q de coordonnées $(4, 2)$. En cas d'égalité de distances, les points de classe R seront considérés en priorité.

Question 2. On considère que la classe R correspond à la valeur $+1$ et la classe C correspond à la valeur -1 et on décide d'utiliser l'algorithme du perceptron. Sans dérouler l'algorithme, mais en justifiant votre réponse, tracer la frontière de décision obtenue. Quelle est la particularité de cette frontière ?

Question 3. En fait, les points M et P sont de la classe *carré* et L , N et Q sont de la classe *rond*. Donner la matrice de confusion pour chacun des 2 modèles appris dans les questions précédentes (k -ppv et perceptron). Quel est le taux d'erreur de chacun de ces modèles ? Lequel est préférable ?