

FTML

adrien.paul

July 2020

1 Exercice 1

1.1 Question a :

EST_1	Positif	Négatif
Positif	6	1
Négatif	8	2

EST_2	Positif	Négatif
Positif	4	3
Négatif	1	9

Pour EST_1, le risque empirique est :

$$\frac{8+x}{17}$$

Pour EST_2 le risque empirique est :

$$\frac{1+3x}{17}$$

Comparons les deux formules en fonction de la valeur de x.

$$\frac{8+x}{17} = \frac{1+3x}{17} \quad (1)$$

$$8+x = 1+3x \quad (2)$$

$$2x = 7 \quad (3)$$

$$x = 3.5 \quad (4)$$

Donc pour x supérieur à 3.5, le risque empirique le plus élevé est celui de EST_2 sinon, s'est celui de EST_1.

2 Exercice 2

2.1 Question a :

J'utiliserais le regroupement hiérarchique pour pouvoir distinguer les points au moyen d'un dendrogramme.

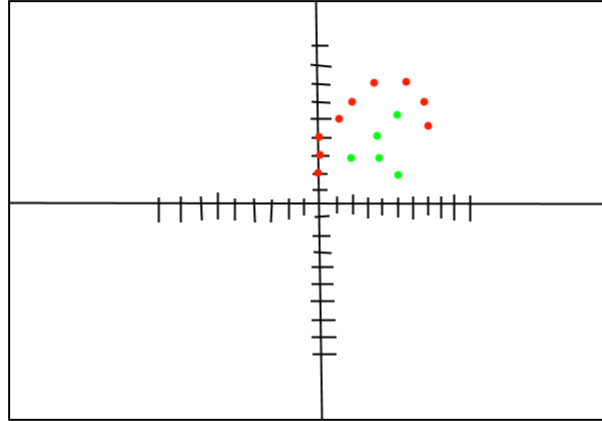


Figure 1: Points visualisation

2.1.1 Question b :

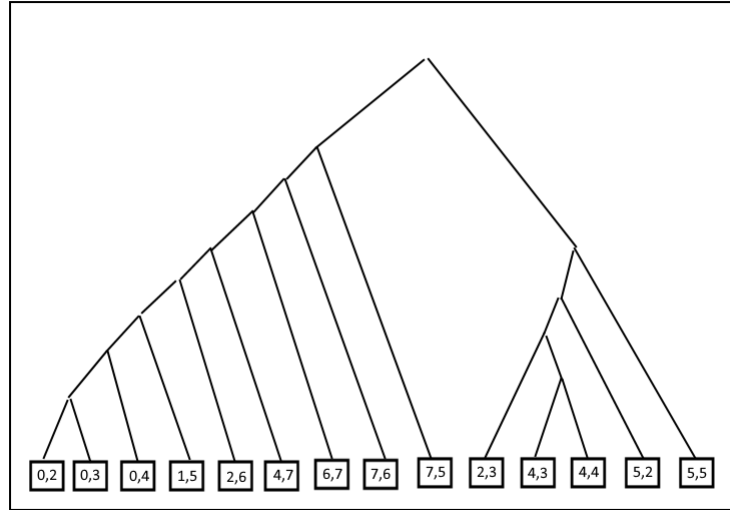


Figure 2: Dandogramme

Pour évaluer le coût de l'intégration, j'ai choisis d'utiliser la distance entre deux points.

2.2 Question c :

Il est possible de les séparer mais avec une fonction de perte très permissive.

2.3 Question d :

Voici les deux SVM linéaires avec les niveaux de pénalisation fixés au préalable. L'un des niveaux de pénalisation est stricte et l'autre non.

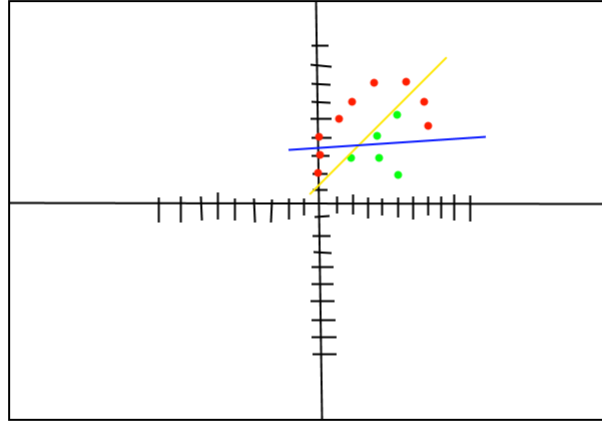


Figure 3: SVM Linéaires

3 Exercice 3

3.1 Question a :

Pour résoudre ce problème, j'utiliserais la méthode du test du khi-deux. Qu'est-ce que ce test ? C'est un test statistique où la statistique de test suit une loi du χ^2 sous l'hypothèse nulle. Dans notre cas, quand il faut trouver la valeur p (probabilité pour un modèle statistique donné sous l'hypothèse nulle d'obtenir la même valeur ou une valeur encore plus extrême que celle observée), on rejette l'hypothèse nulle lorsque $p < 0,05$.

3.2 Question b :