

FTML - Partiel

Lucas Mouminoux

16 Juillet 2020

1 Exercice 1

1.1 a)

Estimateur 1: On a 8 faux-positif et 1 vrai-négatif pour 17 éléments. Donc :

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

Estimateur 2: On a 1 faux-positif et 3 vrai-négatif pour 17 éléments. Donc :

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

Plus la valeur de x est grande plus l'erreur en cas de faux négatif sera grande.

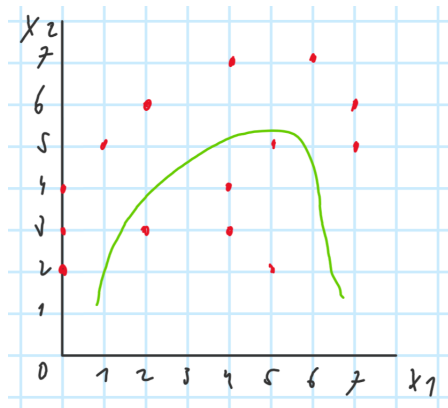
Lorsque $x=3.5$, les deux estimateurs sont égaux.

Lorsque $x > 3.5$, on pénalise plus fortement les faux-positifs.

Lorsque $x < 3.5$, on pénalise plus fortement les vrai-négatifs.

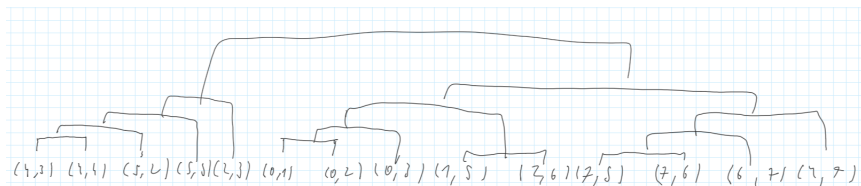
2 Exercice 2

2.1 a)



Des classifieurs non linéaire comme des classifications hiérarchiques permettraient de les distinguer.

2.2 b)



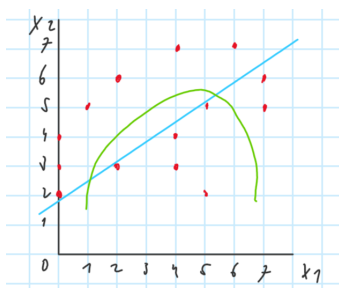
J'ai utilisé une fonction qui calcule la norme 2 de la distance entre 2 points.

2.3 c)

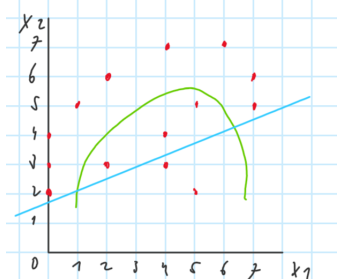
Cela est possible, la séparation ne pourra pas être parfaite mais seulement quelques points seront dans la mauvaise classe en fonction des coefficients de pénalisations.

Une séparation parfaite est impossible avec un SVM linéaire.

2.4 d)



Dans le graphique ci-dessus, on pénalise les erreurs des points appartenant à la classe 1 plus fortement que les erreurs de la classe -1.



Dans le graphique ci-dessus, on pénalise les erreurs des points appartenant à la classe -1 plus fortement que les erreurs de la classe 1.

3 Exercice 3

3.1 a)

Pour cela il faut faire un chi2. Si la valeur obtenu est inférieure à 0.05 on considère que les variables sont indépendantes.

On calcule le chi2 grâce à cette formule :

$$\chi^2 = \sum_{i,j} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

où :

$$E_{i,j} = \frac{\sum_{j=1}^J O_{i,j} * \sum_{i=1}^I O_{i,j}}{N}$$

Avec O l'effectif et N le nombre de données.

Il existe aussi une fonction dans la librairie Pandas en python.

3.2 b)

Je testerai de pulvériser les points un à un jusqu'à ne plus pouvoir. Pour cela, j'isolerais X points et si il est possible de les regrouper grâce à l'ensemble donnée je recommencerais avec X+1 points. Lorsque je ne pourrai plus pulvériser, j'aurai trouvé le nombre maximum de points pulvérisable.