

# Fondamentaux théoriques en Machine Learning

Matthieu Archambault

16 juillet 2020

## Exercice 1 :

a) Pour le risque empirique de l'estimation 1 :

Nous avons 8 faux positifs et 1 faux négatifs

Donc un total de :  $8 + x$

Et une moyenne de :  $(8 + x) / 17$

Pour le risque empirique de l'estimation 2 :

Nous avons 1 faux positifs et 3 faux négatifs

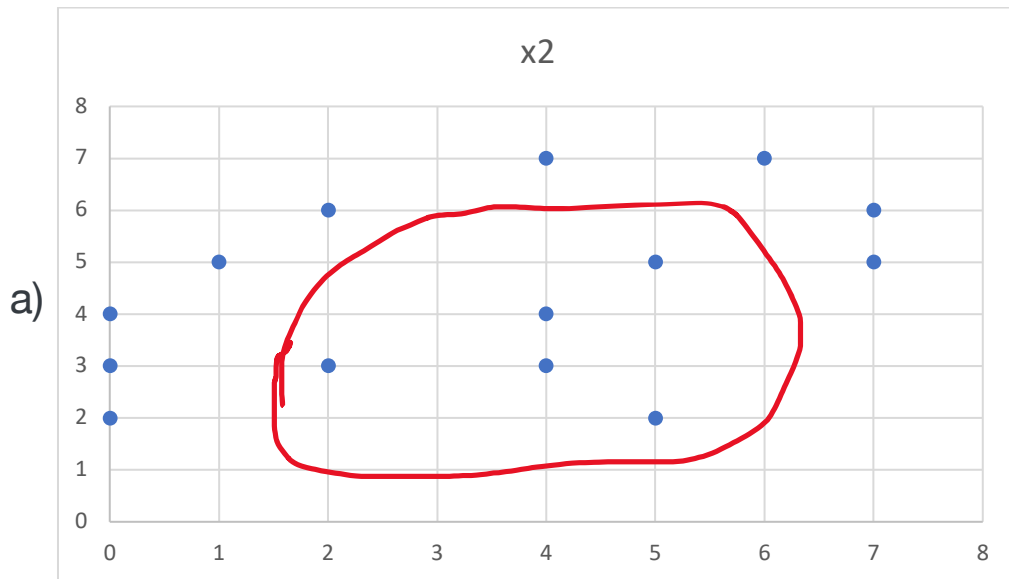
Donc un total de :  $1 + 3x$

Et une moyenne de :  $(1 + 3x) / 17$

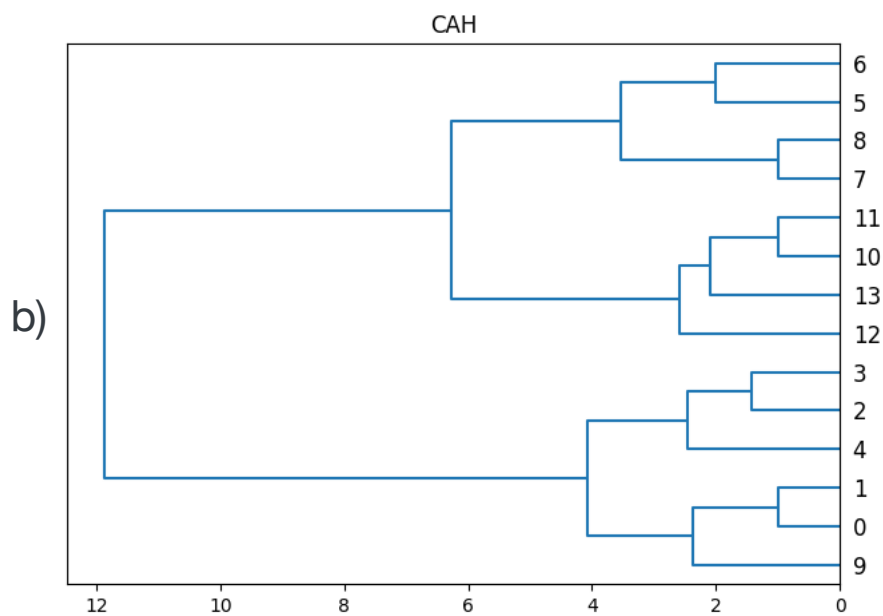
Pour  $x < 3$ , l'estimation 1 à un meilleur score, mais sinon, l'estimation 2 à un meilleur score

b)

## Exercice 2 :



L'algorithme des k-means me paraît le plus à même de distinguer les clusters.

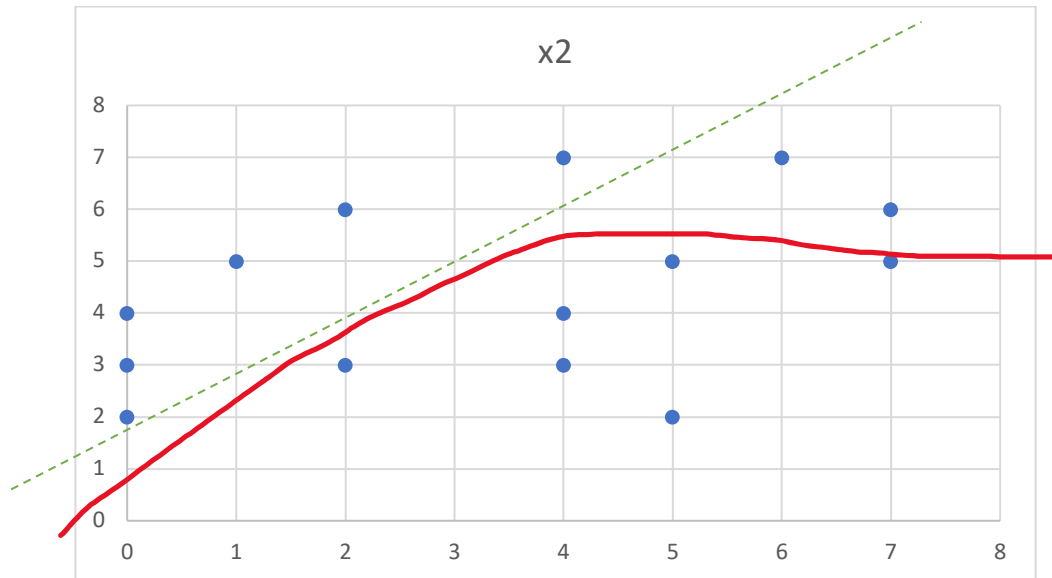


La méthode utilisée est la méthode de Ward.

C)

Ceci est un cas non-séparable en coordonnées cartésiennes, mais si on reconsidère le problème dans un espace de dimension supérieure, il est alors probable que l'on trouve un séparateur linéaire.

D)



### Exercice 3 :

- Pour déterminer l'indépendance des variables concernées, il faudrait que chacun des totaux de ces variables soit égale à 100%.
- Pour trouver le nombre maximal de points pulvérisable à la surface d'une sphère, j'utiliserai le principe de VC-dimension que j'appliquerai à la formule de la surface de la sphère.
- La différence entre le risque et l'ambiguïté concerne les compétences, la perception et les sentiments du décideur quant à savoir s'il est qualifié pour attribuer des probabilités subjectives aux paramètres de la situation.

