

Contrôle 2019 - data mining

A) ACP

- 1) on peut retenir les 4 premières composantes car leurs valeurs propres sont > 1
- 2) intérêt : évaluer les corrélations des variables avec les différents axes
- 3) l'analyse factorielle peut diviser la customer experience en 4 axes:
 - 1) qualité ~~de la~~ ~~manchandise~~ magasin :
(coordonnées haute pour les variables 1, 2, 3, 7)
 - 2) qualité de la marchandise
(coord. haute pour les variables 4 et 8 et 9)
 - 3) degré de confiance / sécurité
(variables 5, 6, 7)
 - 4) qualité du service (variables 3)

B) arbres

- 4) critère : impureté
- 5) prune (arbre full, $cp = 0,03$)
 cp : choisi à partir de la cp table.
 $=$ valeur cp ~~correspondant~~ entre les 2 valeurs cp correspondant à l'erreur x error minimale et

1^{ème} méthode : Scores

score $\leq 0 \Rightarrow$ classe A (plus proche de $\varnothing A$)
score $> 0 \Rightarrow$ classe B (" " " $\varnothing B$)

$\Rightarrow P_1, P_2, P_3, P_4 \in A$

$P_5, P_7, P_8 \in B$

pour l'individu P_6 avec un score 0

$\Rightarrow P_6$ est équidistant aux 2 classes

donc on procède comme précédemment

B) Arbre

6) $rpart =$ fct de construction de l'arbre
 $y \sim . \Rightarrow$ l'arbre dépend de toutes les variables

$data = form \Rightarrow$ base de données

$minsplit = 3 \Rightarrow$ nombre minimal d'individus
dans une feuille de l'arbre

$method = "class" \Rightarrow$ arbre de classification

7)

