

APPRENTISSAGE ARTIFICIEL
M1 Informatique I2L
Examen - 3h

Fabien Teytaud

26 Mars 2021

Documents autorisés.

1 Objectifs

Vous avez été contacté par un client pour faire une étude sur la base `exam.csv`. Le client travaillant avec des données très sensibles, il ne peut nous donner aucune information. Tout ce que l'on sait est que l'on souhaite apprendre la colonne Z.

2 Analyse des données

Dans une première partie vous allez devoir effectuer une analyse des données et vous devez faire une fonction `analyzeData` qui :

- Affiche le nombre d'exemples dans la base et le nombre de caractéristiques.
- Affiche les différentes statistiques de la base.
- Affiche le nombre d'exemples de chaque classe.
- Affiche la matrice de corrélation.

3 Apprentissage

On souhaite maintenant être capable de prédire la colonne Z.

- Le client souhaite avoir la meilleure précision possible.
- Le client souhaite que vous compariez au moins 3 algorithmes d'apprentissage.
- Le client souhaite qu'au moins la moitié des données soit consacrée aux tests.
- Le client souhaite connaître la précision de votre algorithme.
- Le client souhaite également avoir les bonnes réponses par classe de votre algorithme (sous forme de matrice).

4 Conseils

- Les demandes des clients sont primordiales, il est important de les respecter.
- Plus vos résultats seront bons, plus le client sera content, il ne faut pas se contenter d'avoir juste 80%.
- La colonne 'Q' semble importante. Le client peut juste vous dire que la présence de cette caractéristique est très importante pour le problème.
- Comme toujours, le client est méfiant, il souhaite comprendre ce qu'il se passe. N'hésitez pas à proposer plusieurs résultats/améliorations et votre processus de réflexion
- Le client est très regardant de la concurrence, si un de vos concurrents fait mieux pour moins cher, il n'hésitera pas une seconde à vous laisser tomber.

5 Exemple de résultats attendus

Learning with knn

Train score: 0.9, Test score 0.9607843137254902

```
[[22  0  0  0  0  0  0]
 [ 0 11  0  0  0  0  0]
 [ 0  0  2  0  0  0  0]
 [ 0  0  0  6  0  0  0]
 [ 0  0  0  0  1  0  0]
 [ 0  0  0  0  0  4  0]
 [ 0  0  0  0  0  2  3]]
```

Learning with decision tree

Train score: 1.0, Test score 0.9803921568627451

```
[[22  0  0  0  0  0  0]
 [ 0 11  0  0  0  0  0]
 [ 0  0  2  0  0  0  0]
 [ 0  0  0  6  0  0  0]
 [ 0  0  0  0  1  0  0]
 [ 0  0  0  0  0  4  0]
 [ 0  0  0  0  0  1  4]]
```

Learning with neural network

Train score: 1.0, Test score 0.9803921568627451

```
[[22  0  0  0  0  0  0]
 [ 0 11  0  0  0  0  0]
 [ 0  0  2  0  0  0  0]
 [ 0  0  0  6  0  0  0]
 [ 0  0  0  0  1  0  0]
 [ 0  0  0  0  0  4  0]
 [ 0  0  0  0  0  1  4]]
```