

Travail pratique 01 - Introduction aux librairies et notebooks Python

Mathieu Monteverde, Sathiya Kirushnapillai

2018

Questions - Réponses

Concernant la base de données wine, en observant les boîtes à moustache générées à la section 6 du travail, quel(s) composant(s) semblent le plus distinct? Pourquoi?

Il est difficile de désigner un trait de vin pour lequel les classes sont absolument distinctes. Une boîte à moustache permet de visualiser, pour une distribution de valeurs donnée, si une valeur peut être considérée comme aberrante ou non (respectivement en dehors ou en dedans des deux moustaches). Le cas idéal que l'on rechercherait serait donc un trait de vin pour lequel les moustaches des boîtes à moustaches de chaque classe forment des plages qui ne se chevauchent pas.

Il est évident que nous n'avons pas trouvé un tel trait, nous allons donc considérer les traits dont les boîtes de la boîte à moustache ne se chevauchent pas (ou peu).

On peut donc retenir par exemple les traits de vin suivants comme relativement distincts entre les différentes classes :

- Flavanoids
- Proline
- Color intensity

Pouvez-vous estimer la performance de la méthode de classification (Single-rule) comme celle présentée à la section 7?

Définissez une règle qui utilise le composant le plus distinct pour classer les observations de vin.

```
1  pred = []
2
3  for row in df['flavanoids']:
4      if row >= 2.5:
5          pred.append(1);
6      elif row > 1.8 and row < 2.5 :
7          pred.append(2);
8      else:
9          pred.append(3)
10
11 # A new column is added to the dataframe
12 df['prediction'] = pred
```

lol

lol