Pattern Recognition, Homework 3

0516003 李智嘉

Part. 1, Coding

1. Gini Index or Entropy

```
Gini of data is  0.4628099173553719
Entropy of data is  0.9456603046006401
```

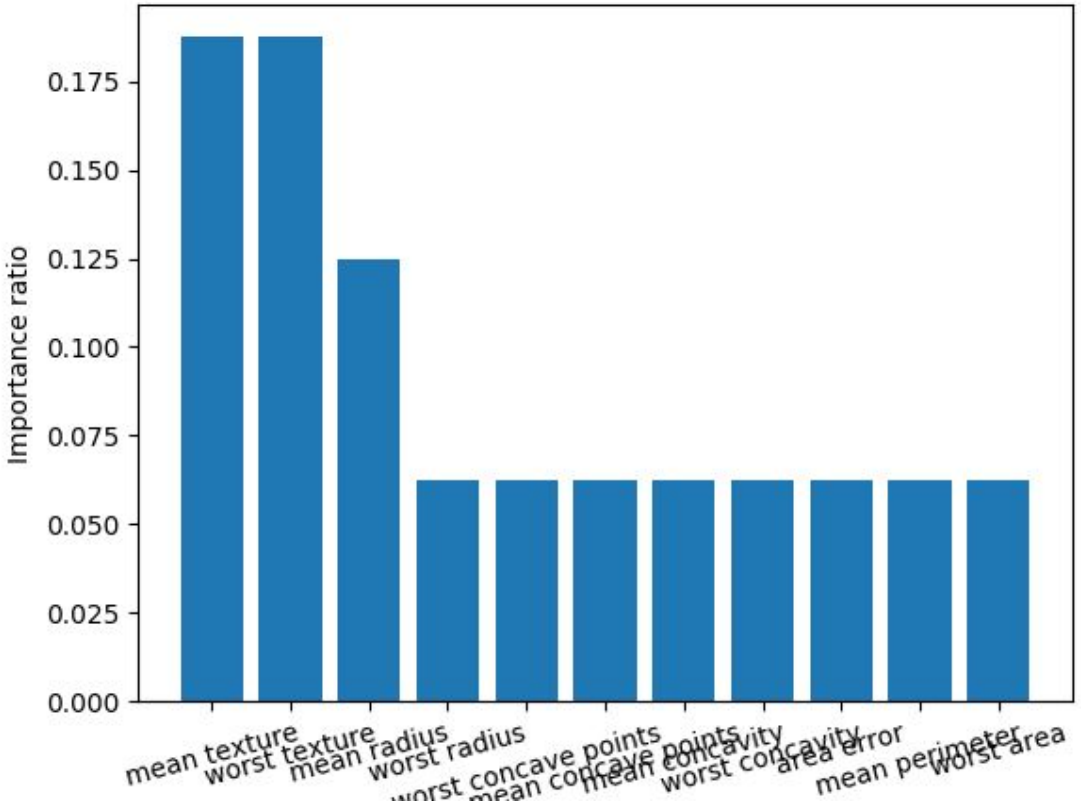
2. Implement the Decision Tree algorithm
   i. Using Criterion='gini', showing the accuracy score of test data by Max_depth=3 and Max_depth=10, respectively.

```
gini accuracy depth 3: 0.9370629370629371
gini accuracy depth 10: 0.916083916083916
```

   ii. Using Max_depth=3, showing the accuracy score of test data by Criterion='gini' and Criterion='entropy', respectively.

```
gini accuracy depth 3: 0.9370629370629371
entropy accuracy depth 3: 0.951048951048951
```

3. Plot the feature importance of your Decision Tree model.



4. Implement the Random Forest algorithm by using the CART
   i. Using Criterion='gini', Max_depth=None, Max_features=sqrt(n_features), Bootstrap=True, showing the accuracy score of test data by n_estimators=10 and n_estimators=100, respectively.

```
clf_10tree accuracy: 0.9440559440559441
clf_100tree accuracy: 0.958041958041958
```

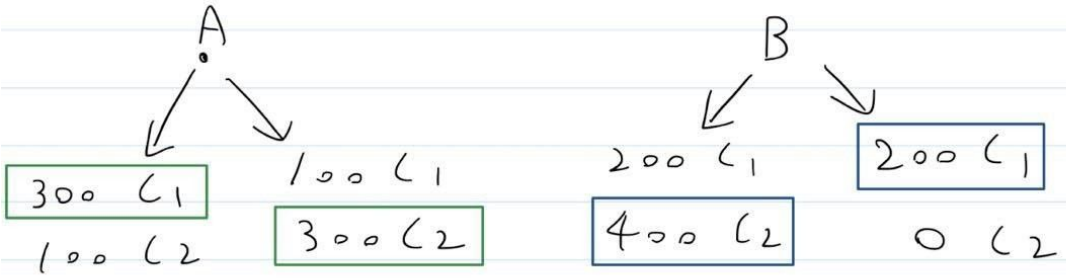
   ii. Using Criterion='gini', Max_depth=None, N_estimators=10, Bootstrap=True, showing the accuracy score of test data by

Max_features=sqrt(n_features) and Max_features=n_features, respectively.

```
clf_random_features accuracy: 0.9440559440559441
clf_all_features accuracy: 0.951048951048951
```

Part. 2, Questions

1. Consider a data set comprising 400 data points from class C_1 and 400 data points from class C_2



Misclassification ratio :

$$\begin{cases} A \Rightarrow (100+100) \div 800 = \frac{1}{4} \\ B \Rightarrow (200+0) \div 800 = \frac{1}{4} \end{cases}$$

$\Rightarrow$ A and B are equal in these data #

Entropy of A: $\frac{4}{8}\left(-\frac{3}{4}\log\frac{3}{4}-\frac{1}{4}\log\frac{1}{4}\right) + \frac{4}{8}\left(-\frac{1}{4}\log\frac{1}{4}-\frac{3}{4}\log\frac{3}{4}\right) \simeq 0.8113$

Entropy of B: $\frac{6}{8}\left(-\frac{2}{6}\log\frac{2}{6}-\frac{4}{6}\log\frac{4}{6}\right) + \frac{2}{8}\left(-\frac{1}{2}\log\frac{2}{2}\right) \simeq 0.6887$

$\Rightarrow$ Entropy of B is less than Entropy of A #

Gini of A: $\frac{4}{8}\left(1-\left(\left(\frac{3}{4}\right)^2+\left(\frac{1}{4}\right)^2\right)\right) + \frac{4}{8}\left(1-\left(\left(\frac{1}{4}\right)^2+\left(\frac{3}{4}\right)^2\right)\right) \simeq 0.375$

Gini of B: $\frac{6}{8}\left(1-\left(\left(\frac{2}{6}\right)^2+\left(\frac{4}{6}\right)^2\right)\right) + \frac{2}{8}\left(1-\left(\frac{1}{2}\right)^2\right) \simeq 0.333$

$\Rightarrow$ Gini of B is less than Gini of A #