Part I

1. (5%) Please compute the Entropy and Gini Index of the given array by the formula on the slides.

Gini of data is 0.4628099173553719

Entropy of data is 0.9456603046006401

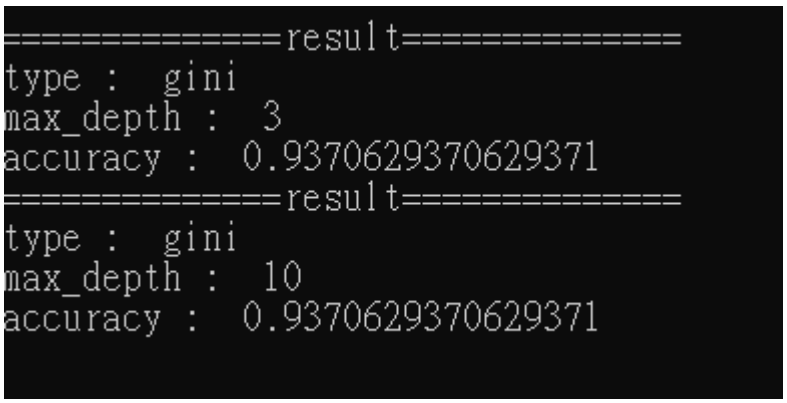
2.1. Using Criterion='gini' to train the model and show the accuracy score of test data by Max_depth=3 and Max_depth=10, respectively.

Max_depth = 3 :

Accuracy : 0.9370629370629371

Max_depth = 10 :

Accuracy : 0.9370629370629371



2.2. Using Max_depth=3 to train the model and show the accuracy score of test data by Criterion='gini' and Criterion='entropy', respectively.

Gini :

accuracy : 0.9370629370629371

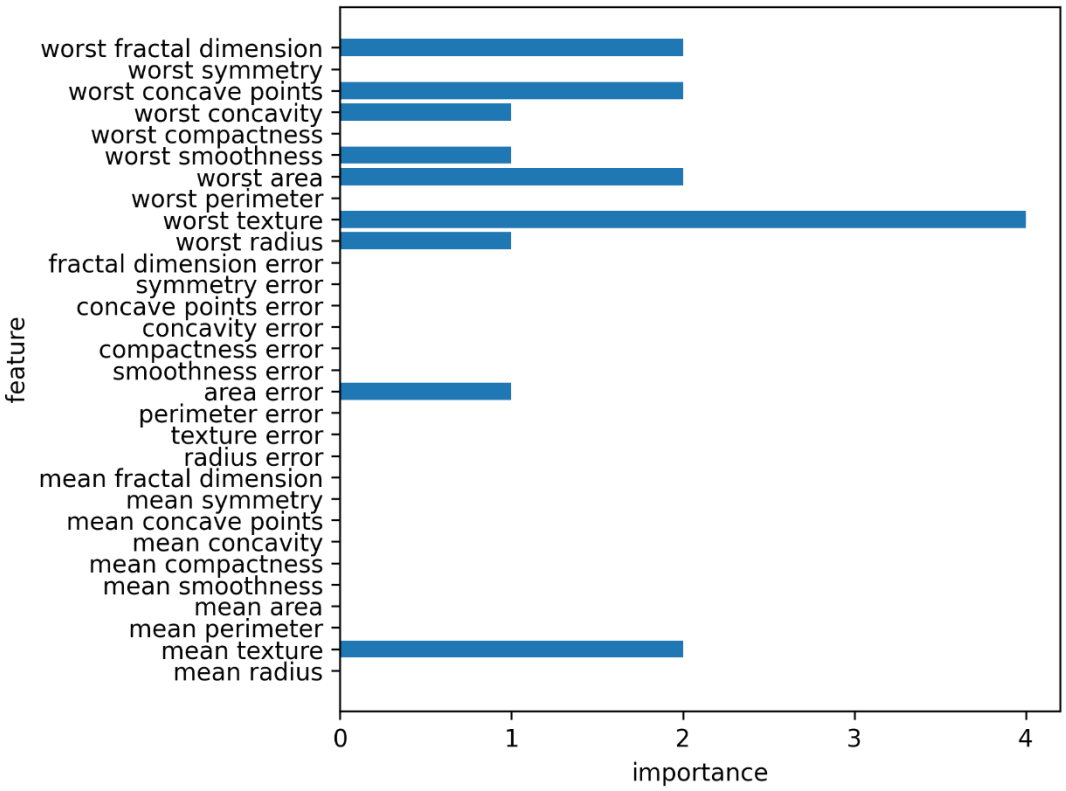
Entropy :

accuracy : 0.958041958041958

```
==============result==============
type :  gini
max_depth :  3
accuracy :  0.9370629370629371
==============result==============
type :  entropy
max_depth :  3
accuracy :  0.958041958041958
```

(15%) Plot the feature importance of your Decision Tree model. You can use the model for Question 2.1, max_depth=10. (You can simply count the number of a feature used in the tree, instead of the formula in the reference. Find more details on the sample code. (matplotlib is allowed to use)

Y 軸是 feature 的種類，X 軸是該 feature 在樹中被使用的次數。



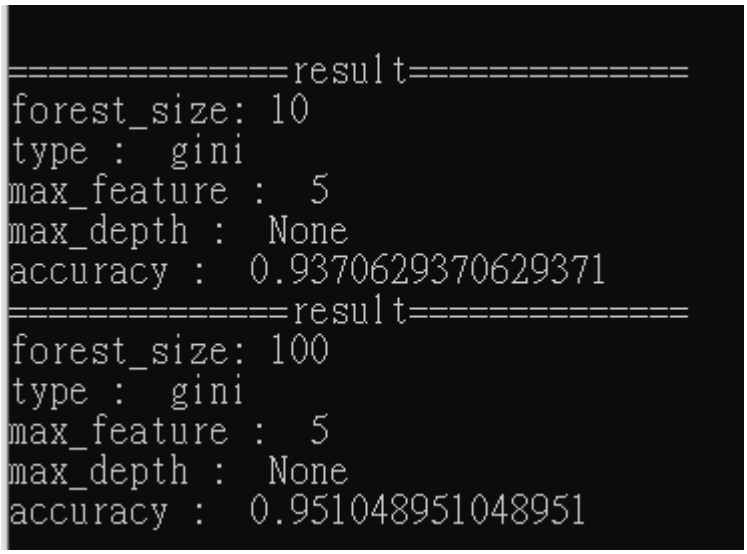
4.1. Using Criterion='gini', Max_depth=None, Max_features=sqrt(n_features), Bootstrap=True to train the model and show the accuracy score of test data by n_estimators=10 and n_estimators=100, respectively.

N = 10:

accuracy : 0.9370629370629371

N=100 :

accuracy : 0.951048951048951



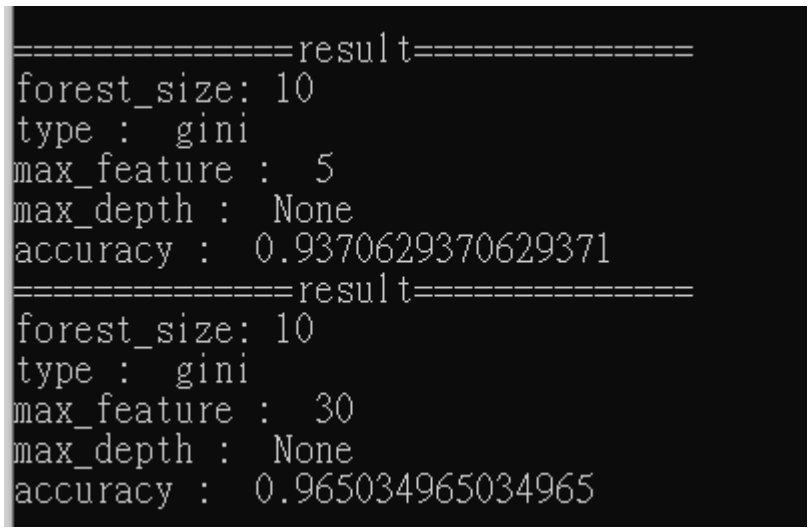
4.2. Using Criterion='gini', Max_depth=None, N_estimators=10, Bootstrap=True, to train the model and show the accuracy score of test data by Max_features=sqrt(n_features) and Max_features=n_features, respectivel

max_feature : 5 sqrt(n_features)

accuracy : 0.9370629370629371

max_feature : 30 (n)

accuracy : 0.965034965034965


==========result==========
forest_size: 10
type : gini
max_feature : 5
max_depth : None
accuracy : 0.9370629370629371
==========result==========
forest_size: 10
type : gini
max_feature : 30
max_depth : None
accuracy : 0.965034965034965

Part II

1.

$$E = \left(e^{\frac{\alpha_m}{2}} - e^{\frac{-\alpha_m}{2}}\right) \sum w^{(m)} I\left(y_m(x_n) \neq t_n\right) + e^{-\frac{\alpha_m}{2}} \sum_{n=1}^{N} w_n^{(m)}$$

$$E' = \frac{1}{2}\left(e^{\frac{\alpha_m}{2}} + e^{\frac{-\alpha_m}{2}}\right) \sum w^{(m)} I\left(y_m(x_n) \neq t_n\right) + \left(-\frac{1}{2}\right) e^{-\frac{\alpha_m}{2}} \sum_{n=1}^{N} w_n^{(m)} = 0$$

$$\Rightarrow \frac{1}{2}\left(e^{\frac{\alpha_m}{2}} + e^{\frac{-\alpha_m}{2}}\right) \sum w^{(m)} I\left(y_m(x_n) \neq t\right) = \frac{1}{2} e^{-\frac{\alpha_m}{2}} \sum_{n=1}^{N} w_n^{(m)}$$

$$\Rightarrow \frac{\sum w^{(m)} I\left(y_m(x_n) \neq t\right)}{\sum_{n=1}^{N} w_n^{(m)}} = \frac{\frac{1}{2} e^{\frac{-\alpha_m}{2}}}{\frac{1}{2}\left(e^{\frac{\alpha_m}{2}} + e^{\frac{-\alpha_m}{2}}\right)}$$

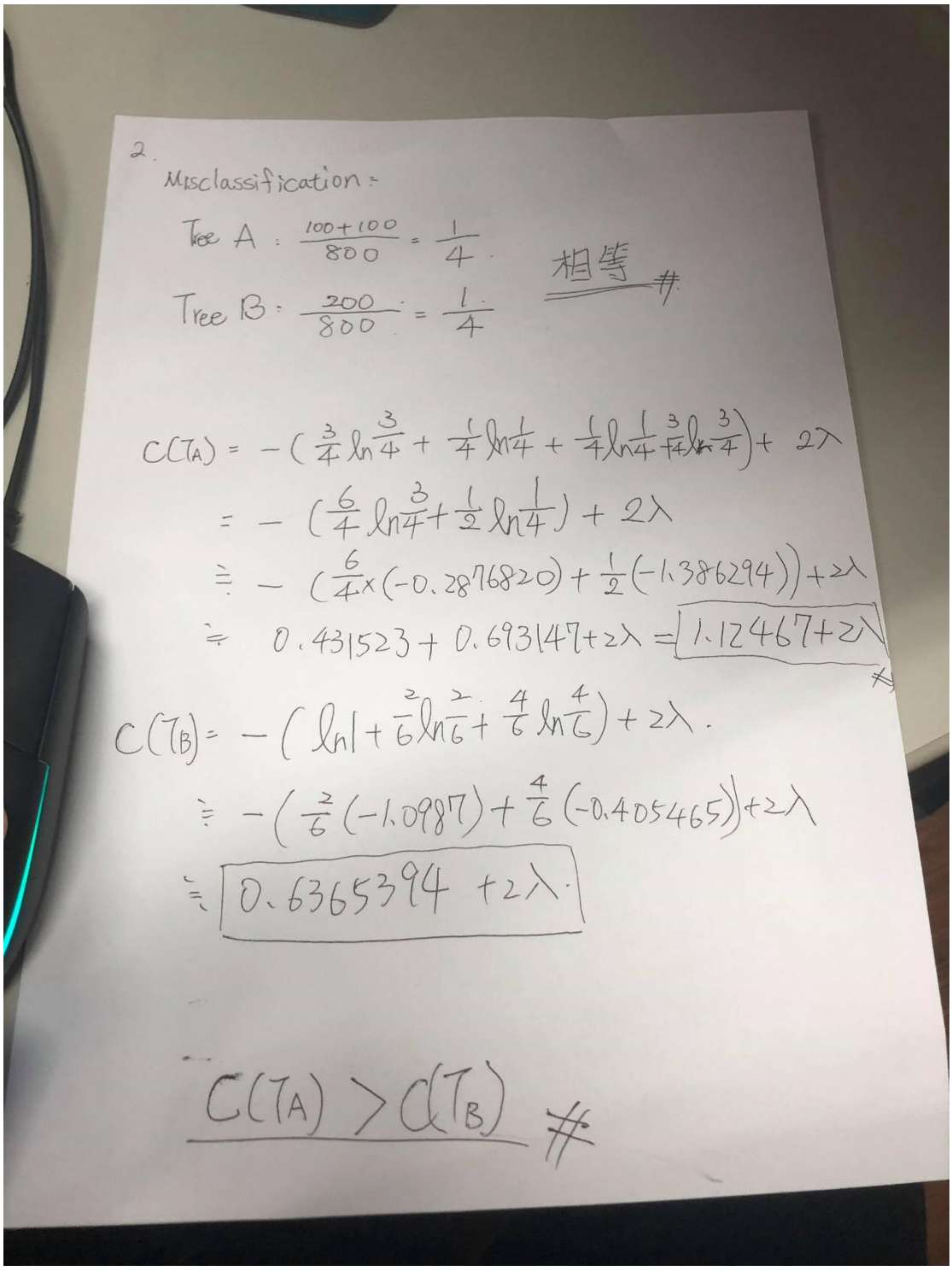
$$\Rightarrow \epsilon_m = \frac{e^{\frac{-\alpha_m}{2}}}{\frac{e^{\alpha_m} + 1}{e^{\frac{\alpha_m}{2}}}} = \frac{1}{e^{\alpha_m} + 1}$$

$$\Rightarrow e^{\alpha_m} + 1 = \frac{1}{\epsilon_m}$$

$$\Rightarrow e^{\alpha_m} = \frac{1}{\epsilon_m} - 1 = \frac{1 - \epsilon_m}{\epsilon_m}$$

$$\Rightarrow \alpha_m = \ln\left(\frac{1 - \epsilon_m}{\epsilon_m}\right)$$

2.



2.

Misclassification:

Tree A: $\frac{100+100}{800} = \frac{1}{4}$.

Tree B: $\frac{200}{800} = \frac{1}{4}$

相等 #

$$C(T_A) = -\left(\frac{3}{4}\ln\frac{3}{4} + \frac{1}{4}\ln\frac{1}{4} + \frac{1}{4}\ln\frac{1}{4} + \frac{3}{4}\ln\frac{3}{4}\right) + 2\lambda$$

$$= -\left(\frac{6}{4}\ln\frac{3}{4} + \frac{1}{2}\ln\frac{1}{4}\right) + 2\lambda$$

$$\approx -\left(\frac{6}{4}\times(-0.287820) + \frac{1}{2}(-1.386294)\right) + 2\lambda$$

$$\approx 0.431523 + 0.693147 + 2\lambda = \boxed{1.12467 + 2\lambda}$$ #

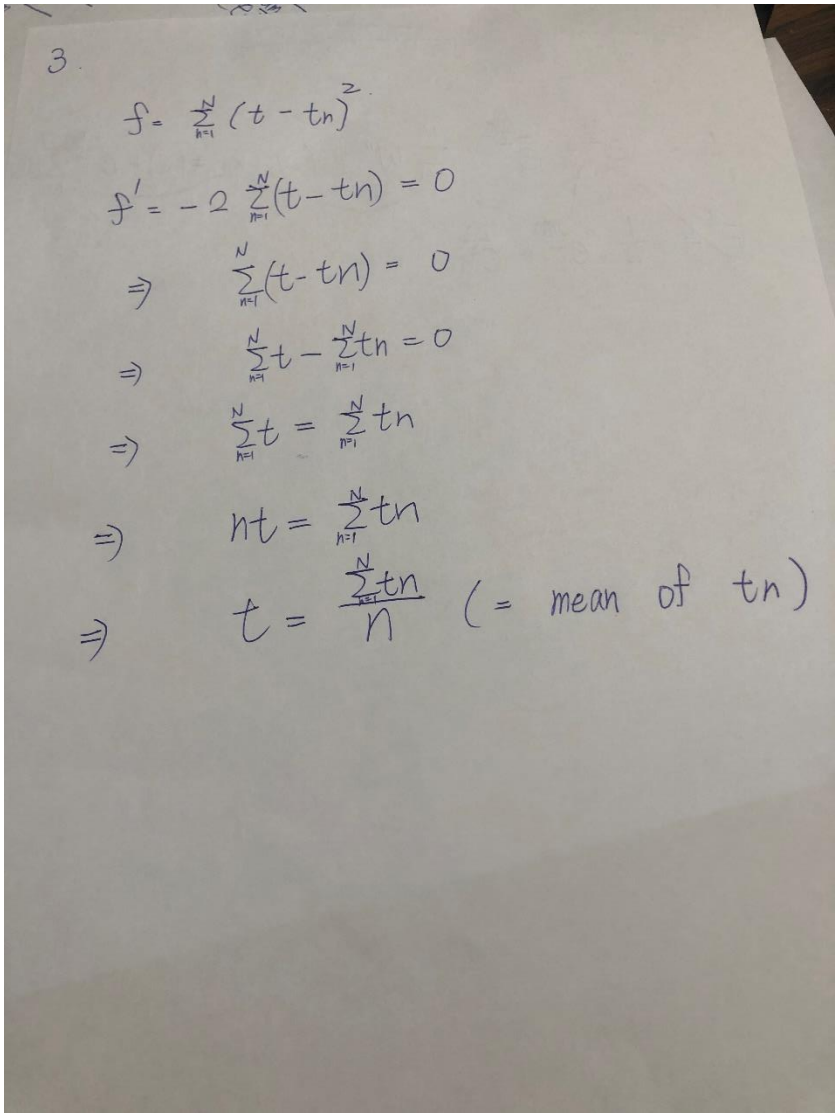
$$C(T_B) = -\left(\ln 1 + \frac{2}{6}\ln\frac{2}{6} + \frac{4}{6}\ln\frac{4}{6}\right) + 2\lambda.$$

$$\approx -\left(\frac{2}{6}(-1.0987) + \frac{4}{6}(-0.405465)\right) + 2\lambda$$

$$\approx \boxed{0.6365394 + 2\lambda}.$$

$C(T_A) > C(T_B)$ #

3



3.

$$f = \sum_{n=1}^{N} (t - t_n)^2$$

$$f' = -2 \sum_{n=1}^{N} (t - t_n) = 0$$

$$\Rightarrow \quad \sum_{n=1}^{N} (t - t_n) = 0$$

$$\Rightarrow \quad \sum_{n=1}^{N} t - \sum_{n=1}^{N} t_n = 0$$

$$\Rightarrow \quad \sum_{n=1}^{N} t = \sum_{n=1}^{N} t_n$$

$$\Rightarrow \quad Nt = \sum_{n=1}^{N} t_n$$

$$\Rightarrow \quad t = \frac{\sum_{n=1}^{N} t_n}{N} \quad (= \text{mean of } t_n)$$