NYCU Pattern Recognition, Homework 3

[Student ID], [Name]

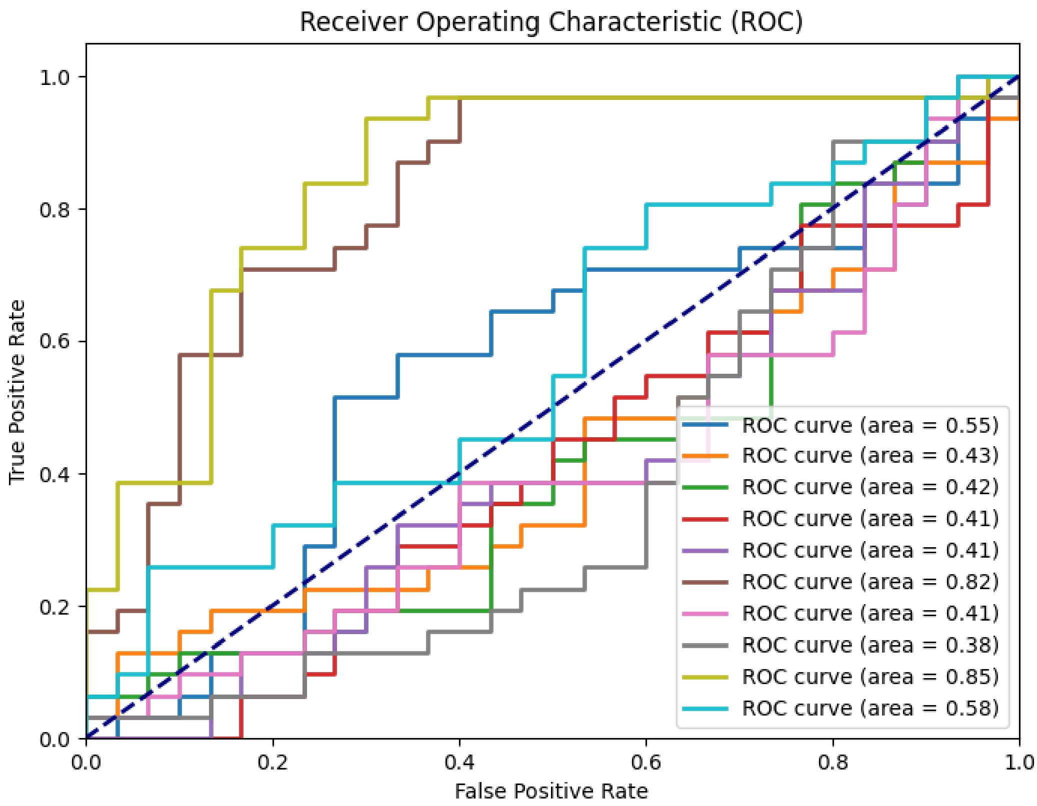
Part. 1, Coding (60%):

(20%) Adaboost

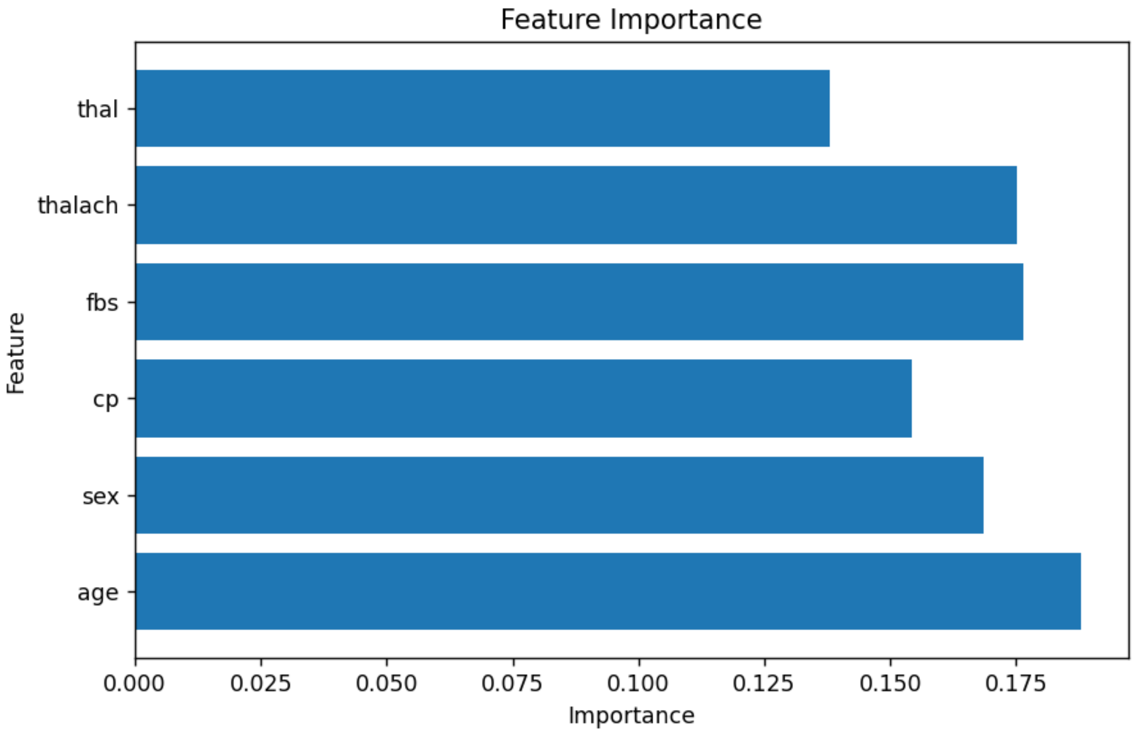
1. (10%) Show your accuracy of the testing data (n_estimators = 10)

```
AdaBoost - Accuracy: 0.7705
```

2. (5%) Plot the AUC curves of each weak classifier.



3. (5%) Plot the feature importance of the AdaBoost method. Also, you should snapshot the implementation to calculate the feature importance.

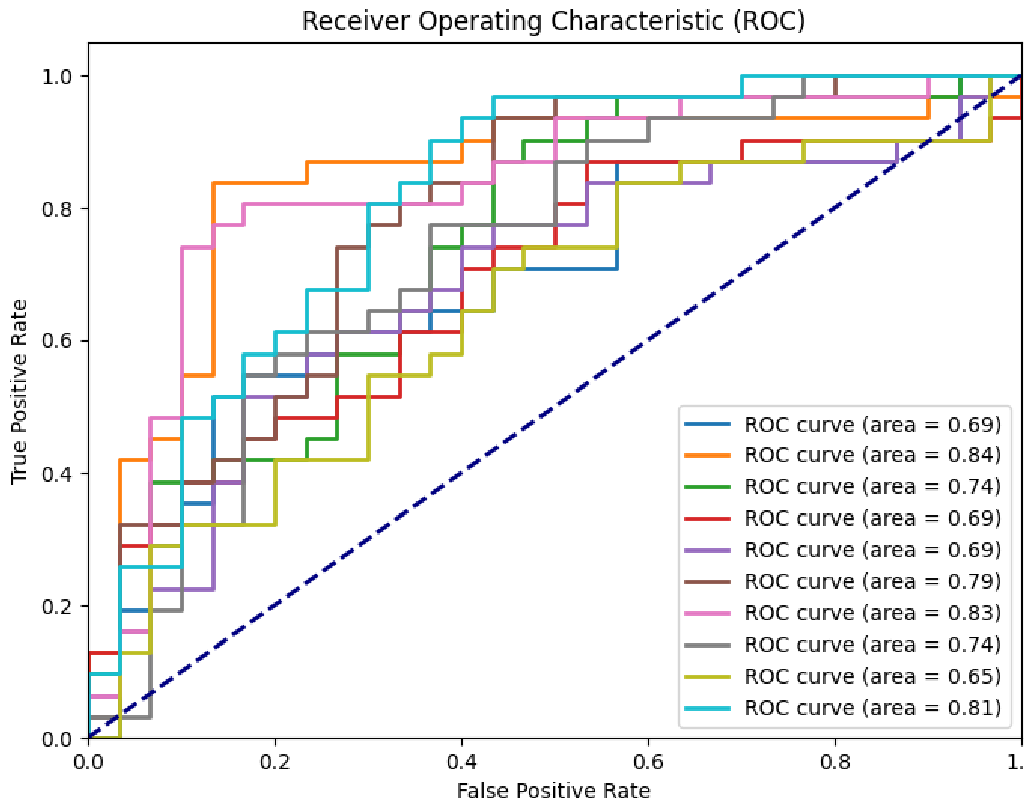


(20%) Bagging

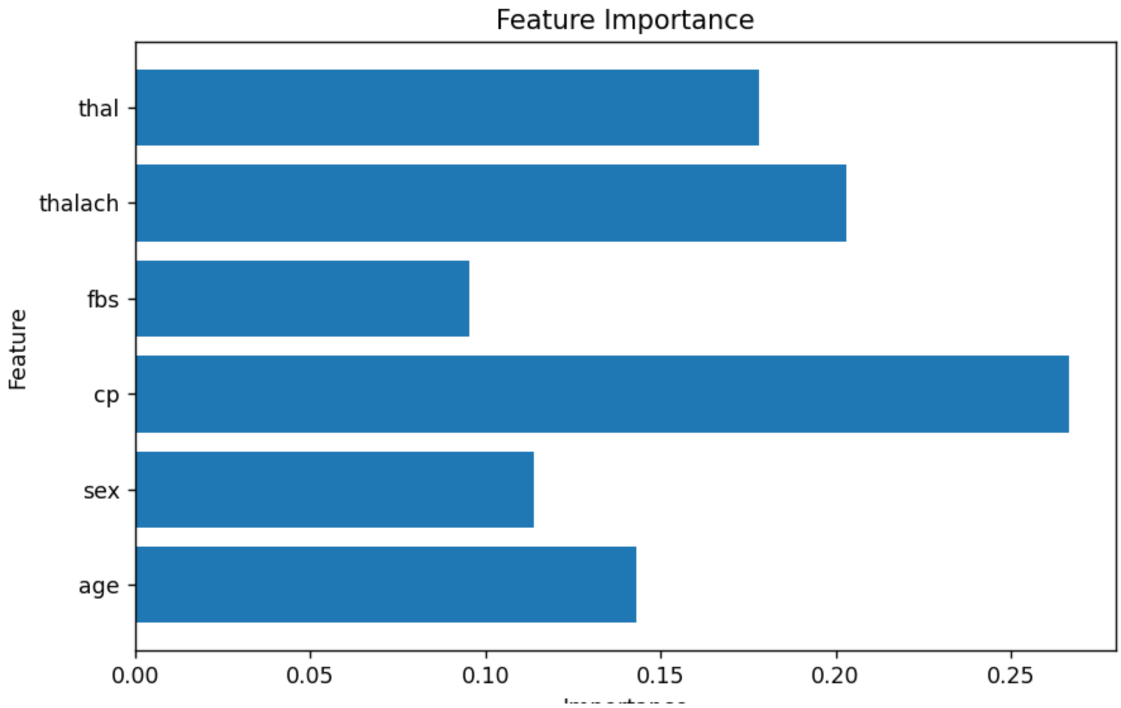
4. (10%) Show your accuracy of the testing data with 10 estimators. (n_estimators=10)

```
- Bagging - Accuracy: 0.8197
```

5. (5%) Plot the AUC curves of each weak classifier.



6. (5%) Plot the feature importance of the Bagging method. Also, you should snapshot the implementation to calculate the feature importance.



(15%) Decision Tree

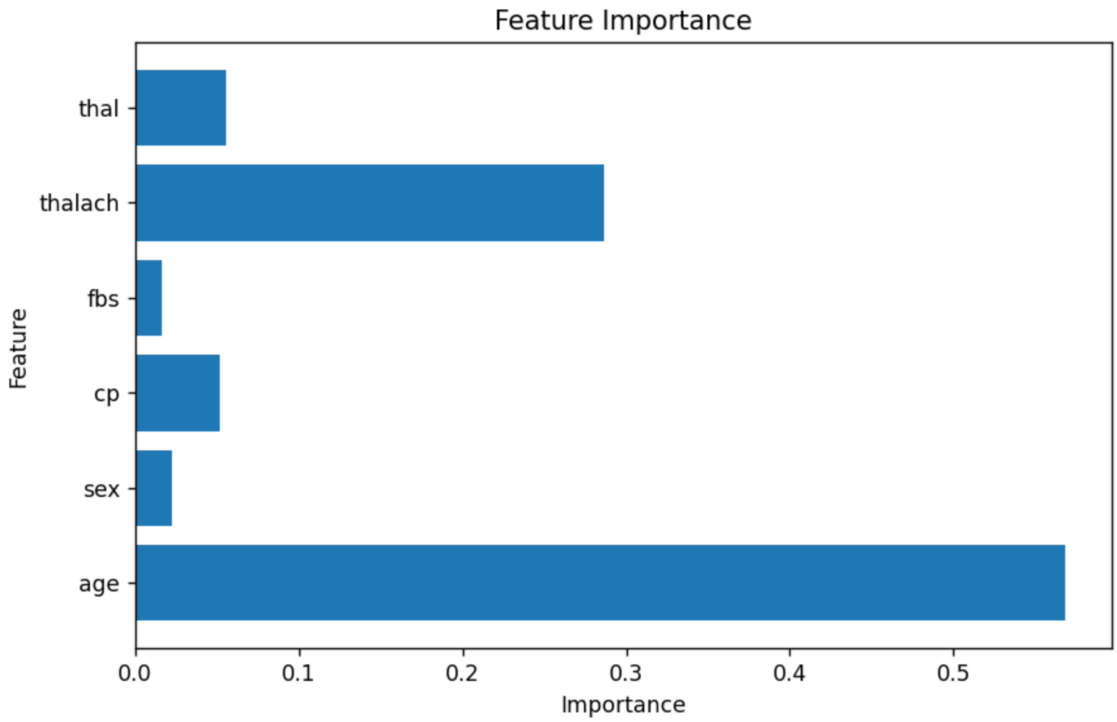
7. (5%) Compute the gini index and the entropy of the array [0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1].

```
Gini of [0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1] is : 0.4628
Entropy of [0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1] is : 0.9457
```

8. (5%) Show your accuracy of the testing data with a max-depth = 7

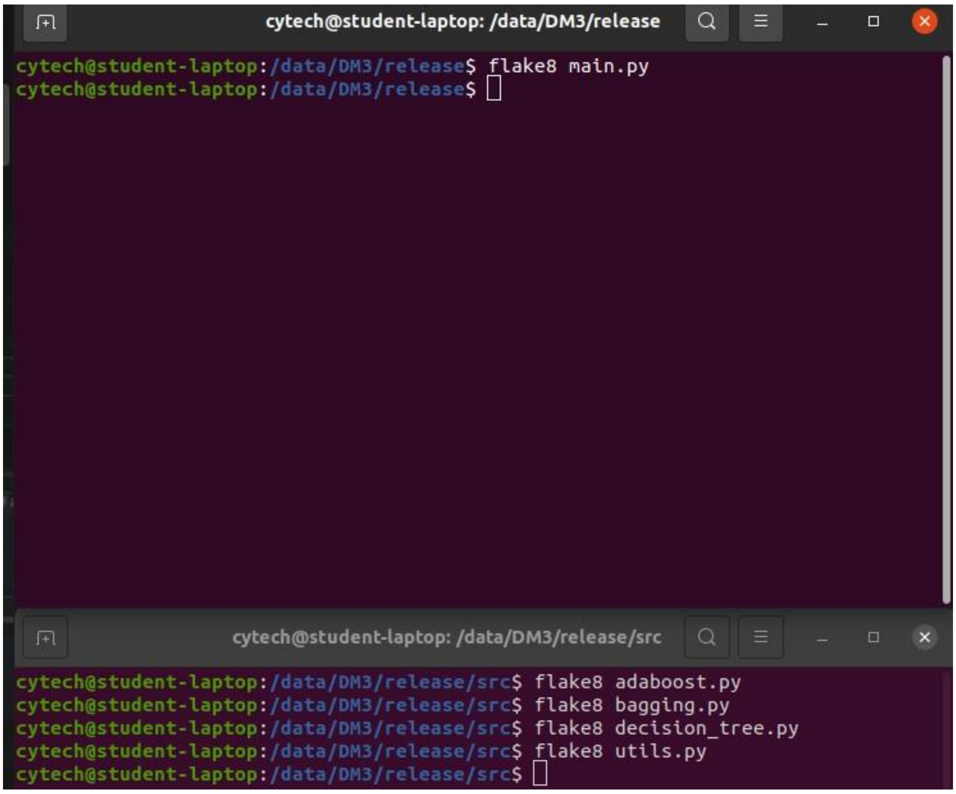
```
DecisionTree - Accuracy: 0.7213
```

9. (5%) Plot the feature importance of the decision tree.



(5%) Code Linting

10. Show the snapshot of the flake8 linting result.



Part. 2, Questions (40%):

1. (10%) We have three distinct binary classifiers, and our goal is to leverage them in creating an ensemble classifier through the majority voting strategy to make decisions.
   Assuming each individual binary classifier operates independently of the others with an accuracy of 60%, what would be the accuracy of the ensemble classifier?

Knowing that each classifier has an accuracy of 60%, the accuracy of the classifier set is equal to the probability that exactly two classifiers predict the right class plus the probability that three classifiers predict the right class.

— The probability P(A) that exactly two classifiers predict the right class is :
$P(A) = \binom{3}{2} * p^2 * (1-p)^{3-2}$
$P(A) = \binom{3}{2} * 0.6^2 * 0.4^1$
$P(A) = 0.432$

— The probability P(B) that all three classifiers predict the right class is :
$P(B) = p * p * p$
$P(B) = 0.6^3$
$P(B) = 0.216$

Finally the $P(2 \leq X) = 0.216 + 0.432 = 0.648$. Thus the accuracy of the classifier ensemble is 64.8%.

2. (15%) For the decision tree algorithm, we can use the "pruning" technique to avoid overfitting. Does the random forest algorithm also need pruning?

To avoir overfitting, we can prune the decision tree, but random forests don't need to be pruned for two reasons.

First, the random forest uses bagging (bootstrap aggregation), which means that each tree is trained on different data. This tends to reduce overfitting.

Also, the features selected at each partition are completely random; by randomizing the feature selection, the correlation between each tree is reduced, which reduces overfitting.

Thus, pruning a random forest could reduce the performance of our model.

3. (15%) Activation functions are core components of neural networks. They need to be differentiable to ensure backpropagation works correctly. Please calculate the derivatives of the following commonly used activation functions.

(For questions 1. and 2., consider the cases where $x > 0$ and $x \leq 0$)

1. f(x) = relu(x),	df(x)/dx = ?

2. f(x) = leaky_relu(x) with negative_slope=0.01,	df(x)/dx = ?
3. f(x) = sigmoid(x),	df(x)/dx = ?
4. f(x) = silu(x),	df(x)/dx = ?
5. f(x) = tanh(x),	df(x)/dx = ?

Write or type your answer here.

For f(x) = relu(x) we have :

— f(x)= $\begin{cases} 0 & if x \leq 0 \\ \text{x} & if x > 0 \end{cases}$

— f'(x)= $\begin{cases} 0 & if x < 0 \\ undefined & if x = 0 \\ 1 & if x > 0 \end{cases}$

For f(x) = leaky_relu(x) with negative_slope=0.01 we have :

— f(x)= $\begin{cases} 0.01\text{x} & if x < 0 \\ \text{x} & if x \geq 0 \end{cases}$

— f'(x) = $\begin{cases} 0.01 & if x < 0 \\ 1 & if x \geq 0 \end{cases}$

For f(x) = sigmoid(x), we have :

— $f(x) = \frac{1}{1+e^{-x}}$

— $f'(x) = \frac{1}{1+e^{-x}} * (1 - \frac{1}{1+e^{-x}})$
$f'(x) = f(x) - (1 - f(x))$

For f(x) = silu(x), we have :

— $f(x) = \frac{x}{1+e^{-x}}$

— $f'(x) = \frac{1+e^{-x}+x*e^{-x}}{(1+e^{-x})^2}$

For f(x) = f(x) = tanh(x), we have :

— $f(x) = \frac{e^x-e^{-x}}{e^x+e^{-x}}$

— $f'(x) = 1 - \frac{(e^x-e^{-x})^2}{(e^x+e^{-x})^2}$
$f'(x) = 1 - f(x)^2$