Report

Task 1.

```
### Step 2(c): Compute weighted error rate ###
Error_rate =0.334

### Step 2(d): Compute coefficient ###
Coefficient =0.3450743387818443

### Step 2(e): Updated weights ###
Updated weights = [0.05098807 0.05098807 0.05098807 0.05098807 0.05098807 0.05098807
 0.11826401 0.23581985 0.23581985 0.05098807]

### Step 2(f): Normalize weights to sum to 1 ###
Normalized updated weights = [0.05385191 0.05385191 0.05385191 0.05385191 0.05385191
0.05385191 0.12490651 0.24906507 0.24906507 0.05385191]
```

Index	x	y	weights	$\hat{y}$	Updated weights
1	1.0	1	0.072	1	0.05385191
2	2.0	1	0.072	1	0.05385191
3	3.0	1	0.072	1	0.05385191
4	4.0	-1	0.072	-1	0.05385191
5	5.0	-1	0.072	-1	0.05385191
6	6.0	-1	0.072	-1	0.12490651
7	7.0	1	0.167	1	0.24906507
8	8.0	1	0.167	-1	0.24906507
9	9.0	1	0.167	-1	0.05385191
10	10.0	-1	0.072	-1	0.05385191

Detailed procedure is shown in Task 1.py.

Task 4.1. Comparison of ensemble methods with baseline method

Classifiers	Train accuracy	Test accuracy
Perceptron	0.979	0.933
Bagging	1	0.946
Random_forest	1	0.959
Adaboost	0.905	0.894

The selected dataset is digits provided in the sklearn library, the baseline classifier (Perceptron) is utilized to compare the accuracy with three ensemble methods (Bagging, Random_forest, and Adaboost). The train and test accuracies of Perceptron classifier are 0.979 and 0.933, respectively. Due to the ensemble methods, the Bagging and Random_forest classifiers achieve a higher accuracy than the Perceptron. However, the accuracy of Adaboost is lower than that of Perceptron. Another difference can be easily observed between train and test accuracies. For the Perceptron, the train accuracy is about 4% higher than the test accuracy, which is termed as overfitting. In contrast, the difference between train and test accuracies in the ensemble methods of three classifiers can be ignored. This improvement is achieved by the implementation of ensemble methods.

Task 4.2. Comparison of ensemble methods train accuracy with different values

Based on the parameter tuning processes for each classifier, the corresponding parameters are optimized through the comparison among three selected values. The tuned parameters are shown in the last column in each table. The train accuracies are listed in the Train accuracy column accordingly.

Table 1. Parameter tuning of Bagging (Train accuracy)

No.	Parameters				Train accuracy	Tuned parameter
	n_estimators	min_samples	max_features	bootstrap		
1	10	10	10	True	1	10
2	50	10	10	True	1	
3	100	10	10	True	1	
4	10	1	10	True	1	1
5	10	10	10	True	1	
6	10	20	10	True	1	
7	10	1	1	True	1	1
8	10	1	5	True	1	
9	10	1	10	True	1	
10	10	1	1	True	1	True
11	10	1	1	False	1	

Table 2. Parameter tuning of Random_forest (Train accuracy)

No.	Parameters					Train accuracy	Tuned parameter
	n_estimators	criterion	Max_depth	Min_samples_split	Min_samples_leaf		
1	10	gini	2	2	2	0.715	50
2	50	gini	2	2	2	0.822	
3	100	gini	2	2	2	0.832	
4	50	gini	2	2	2	0.822	entropy
5	50	entropy	2	2	2	0.846	
6	50	entropy	2	2	2	0.846	10
7	50	entropy	10	2	2	0.997	
8	50	entropy	20	2	2	0.998	
9	50	entropy	10	2	2	0.997	5
10	50	entropy	10	5	2	1	
11	50	entropy	10	10	2	0.963	
12	50	entropy	10	5	2	1	2
13	50	entropy	10	5	5	0.996	
14	50	entropy	10	5	10	0.982	

Table 3. Parameter tuning of Adaboost (Train accuracy)

No.	Parameters		Train accuracy	Tuned parameter
	n_estimators	Learning_rate		
1	10	0.1	0.686	100
2	50	0.1	0.637	
3	100	0.1	0.747	
4	100	0.001	0.460	0.5
5	100	0.01	0.733	
6	100	0.5	0.859	

Task 4.3. Comparison of ensemble methods test accuracy with different values

With the same procedure in problem 2, the test accuracies are also provided in the Test accuracy column, which are all slight less than the train accuracy. The overall accuracy is reasonable based on the tuning process.

Table 4. Parameter tuning of Bagging (Test accuracy)

No.	Parameters				Test accuracy	Tuned parameter
	n_estimators	min_samples	max_features	bootstrap		
1	10	10	10	True	0.946	10
2	50	10	10	True	0.957	
3	100	10	10	True	0.959	
4	10	1	10	True	0.946	1
5	10	10	10	True	0.946	
6	10	20	10	True	0.946	
7	10	1	1	True	0.946	1
8	10	1	5	True	0.946	
9	10	1	10	True	0.946	
10	10	1	1	True	0.946	True
11	10	1	1	False	0.946	

Table 5. Parameter tuning of Random_forest (Test accuracy)

No.	Parameters					Test accuracy	Tuned parameter
	n_estimators	criterion	Max_depth	Min_samples_split	Min_samples_leaf		
1	10	gini	2	2	2	0.704	50
2	50	gini	2	2	2	0.809	
3	100	gini	2	2	2	0.819	
4	50	gini	2	2	2	0.809	entropy
5	50	entropy	2	2	2	0.831	
6	50	entropy	2	2	2	0.831	10
7	50	entropy	10	2	2	0.941	
8	50	entropy	20	2	2	0.948	
9	50	entropy	10	2	2	0.941	5
10	50	entropy	10	5	2	0.959	
11	50	entropy	10	10	2	0.963	
12	50	entropy	10	5	2	0.959	2
13	50	entropy	10	5	5	0.957	
14	50	entropy	10	5	10	0.950	

Table 6. Parameter tuning of Adaboost (Test accuracy)

No.	Parameters		Test accuracy	Tuned parameter
	n_estimators	Learning_rate		
1	10	0.1	0.685	100
2	50	0.1	0.622	
3	100	0.1	0.730	
4	10	1	0.452	0.5
5	10	10	0.704	
6	10	20	0.831	