Projects in machine learning and AI
HW 2

Xiao Jin

September 2021

1 Task 1

We implement the decision tree classifier. We import the Decision Tree Classifier in the sklearn. We first use the default parameter in the package. Then we change some parameters and we have the following results

max depth	minimum samples leaf	testing accuracy
default (None)	default (1)	87.74%
10	1	87.56%
10	5	86.51%
20	10	87.66%
100	10	87.71%

Table 1: Different factors in decision tree

It seems that the 2 factors we choose don't affect the classification results much.

2 Task 2

For the bagging method, we choose random forest method. For the boosting method we choose Adaboost. We get the following results:

Generally, random forest outperforms Adaboost because its testing accuracy is higher. The cross validation score is close to the testing accuracy.

Algorithm	testing accuracy	cross validation score
Random Forest	97.04%	0.9679
Adaboost	72.99%	0.7255

Table 2: Algorithm of bagging and boosting

3 Task 3

We use the confusion matrix to compare the effectiveness of the models:

Class	0	1	2	3	4	5	6	7	8	9
0	920	0	8	9	3	10	12	4	8	6
1	1	1094	8	3	1	6	5	4	12	1
2	15	13	872	27	17	8	12	31	26	11
3	7	8	29	857	5	44	6	6	29	19
4	5	2	10	2	863	10	14	10	21	45
5	19	6	6	42	6	741	25	4	24	19
6	14	3	12	6	17	23	854	1	24	4
7	4	13	18	21	7	5	4	924	8	24
8	10	5	27	43	17	37	18	9	780	28
9	13	2	8	19	39	8	7	23	26	864

Table 3: confusion matrix of Decision Tree

It is clear that random forest performs best over all the algorithms. According to the confusion matrix, it is clear to demonstrate the classification accuracy and error of each digit. The decision tree performs better than Adaboosting generally

Class	0	1	2	3	4	5	6	7	8	9
0	971	1	0	0	0	3	1	1	2	1
1	0	1123	3	3	0	1	3	0	1	1
2	5	0	1002	4	3	0	3	8	7	0
3	0	0	10	975	0	5	0	10	6	4
4	1	0	1	0	955	0	4	0	3	18
5	2	0	1	12	2	860	6	2	5	2
6	7	3	1	0	5	3	937	0	2	0
7	1	3	20	1	1	0	0	990	1	11
8	3	0	5	8	5	6	4	4	931	8
9	4	5	2	10	12	5	1	5	3	962

Table 4: confusion matrix of Random Forest

Class	0	1	2	3	4	5	6	7	8	9
0	920	0	8	9	3	10	12	4	8	6
1	1	1094	8	3	1	6	5	4	12	1
2	15	13	872	27	17	8	12	31	26	11
3	7	8	29	857	5	44	6	6	29	19
4	5	2	10	2	863	10	14	10	21	45
5	19	6	6	42	6	741	25	4	24	19
6	14	3	12	6	17	23	854	1	24	4
7	4	13	18	21	7	5	4	924	8	24
8	10	5	27	43	17	37	18	9	780	28
9	13	2	8	19	39	8	7	23	26	864

Table 5: confusion matrix of Decision Tree

Class	0	1	2	3	4	5	6	7	8	9
0	883	0	25	3	4	28	23	3	3	8
1	0	1070	3	8	3	1	4	25	21	0
2	30	35	596	32	18	8	208	25	75	5
3	28	32	19	678	2	92	30	33	74	22
4	4	2	17	14	708	16	10	80	35	96
5	29	32	6	122	27	526	22	19	71	38
6	20	10	35	6	26	32	822	1	6	0
7	7	16	23	8	14	7	1	804	20	128
8	40	48	11	91	15	34	22	18	661	34
9	9	11	23	32	161	18	1	169	34	551

Table 6: confusion matrix of Adaboosting