

MSBD 6000B Deep Learning

Project 1

CHANG, YaJie 20459996

1. Pre-processing

We have totally 57 features. For first 53 features, there are a lot of samples are '0's, and only a few of them have some non-zero numbers. I use these 53 features directly because I could not distinguish the noises or outliers with such a few information. However, most of the samples are non-zero for the last three features. I choose to scale them with mean 0 and variance 1:

```
# Pre-process data  
train[:, 54:57] = preprocessing.scale(train[:, 54:57], axis = 0)  
test[:, 54:57] = preprocessing.scale(test[:, 54:57], axis = 0)
```

2. Algorithm

2.1 Key idea:

Build three best-performed models and combine them together to predict the test data. I choose the majority classification result among the classification results by these three models as the final result.

2.2 Detailed algorithm:

- (1) Load data from the .csv files and save them as matrix.
- (2) Pre-process data.
- (3) Split initial train data as new train data and validation data.
- (4) Use AdaBoost based on Decision Tree to build the model1 and compute the accuracy for model1.
- (5) Use Random Forest Classifier to build the model2 and compute the accuracy for model2.
- (6) Use Multilayer Perceptron Classifier to build model3 and compute the accuracy

for model3.

(7) Ensemble model1, model2 and model3 to predict the validation data and compute the ensembled accuracy

(8) Ensemble model1, model2 and model3 to predict the test data

(9) Save the prediction of the test data into a .csv file

3. Result

3.1 Console output:

```
In [99]: runfile('E:/HKUST/Deep Learning/Project 1/  
project1.py', wdir='E:/HKUST/Deep Learning/Project 1')  
AdaBoost Classifier: 0.939130434783  
RandomForest Classifier: 0.942857142857  
MLPClassifier Classifier: 0.946583850932  
Ensemble previous models: 0.9515527950310559
```

3.2 File output:

	A	B	C	D	E	F	G	H
1	1							
2	0							
3	1							
4	1							
5	0							
6	0							
7	0							
8	0							
9	1							
10	1							
11	0							
12	0							
13	0							
14	0							
15	0							
16	1							
17	1							
18	0							
19	1							