作業3-1 Report Discussion

姓名:鄭至善 系級:資訊109 學號: F74051043

* How did you preprocess this dataset ?

Ans:

1. Adding more features:

Including yesterday’s information (E.g. 漲幅, 震盪幅度)

Using yesterday’s information & today’s information to predict tomorrow's stock movement.

1. Data normalization.

* Which classifier reaches the highest classification accuracy in this dataset ?
  + Why?
  + Can this result remain if the dataset is different?

Ans:

1. Accuracy: NN > SVM > Logistic regression.

Logistic regression:

Training accuracy: 54.5%

Testing accuracy: 50%

In other words, can’t learn well on this model.

With the same method, more features should be considered.

SVM:

Training accuracy: 55.2%

Testing accuracy: 50.8%

Get almost same result with Logistic regression.

However, there are more hyperparameters can be optimized.

Utilizing different kernel function for features mapping in higher dimension may be an alternative solution.

NN:

Training accuracy: 54.5%

Testing accuracy: 53.2%

Can get higher accuracy than previous models due to the more complexity.

Nevertheless, 53% of accuracy is still a low performance output.

Taking other NN model(e.g. LSTM) rather than fully-connected NN

may perform better.

1. Logistic Regression and SVM definitely not.

NN maybe.

* How did you improve your classifiers ?

Ans:

1. Using Grid search for hyperparameter tuning.
2. Using 5-Fold cv to estimate the skill of ML models.
3. Manually experiment with different combination of features for training.
4. As for NN, I tried fully-connected model for prediction. And experiment with different numbers of hidden layer. Finally pick the best performance one.

(2 hidden layer)