SchoolNova Computer Science 202

Homework 23

Due 4/10/2021 on Google Classroom

Task 1

Using PDR's .get_data_tiingo(), download the stock data for DJIA companies (you can get the tickers from https://www.cnbc.com/dow-30/). Feel free to choose your own start date. Hint: creating a list of tickers in advance is probably the easiest approach.

Task 2

For this task, you may want to build on the ideas and code that we developed in class.

Let's say your goal is to find an optimal test_size and an optimal lag value for each company: that is, the values that give you the highest percentage of correct predictions.

How can you automate this exploration process? (You certainly do not want to run each model manually). Moreover, you may want to repeat each specification at least 100 times and use the average result as a measure of success (again, see the class code).

Task 3

Based on your results from Task 2, which DJIA company is most predictable and which one is most unpredictable, given your machine learning models?

Sort the DJIA companies, from most to least predictable.

Task 4

For a quick description of the key differences between regression and classification see, optionally: https://www.geeksforgeeks.org/ml-classification-vs-regression/

Task 5

Repeat Tasks 2-3 using a SVC classifier instead of the linear regression model. Make sure to import SVC:

from sklearn.svm import SVC

To replace our previous regression model with a new classification model, all you need to do is replace reg with clf:

```
clf.fit(X_train, y_train)
predicted = clf.predict(X_test)
```

Make sure to define clf in advance:

```
clf = SVC()
```