

Data Science HW3 Report

學號：b08502141

姓名：石旻翰

系級：電機四

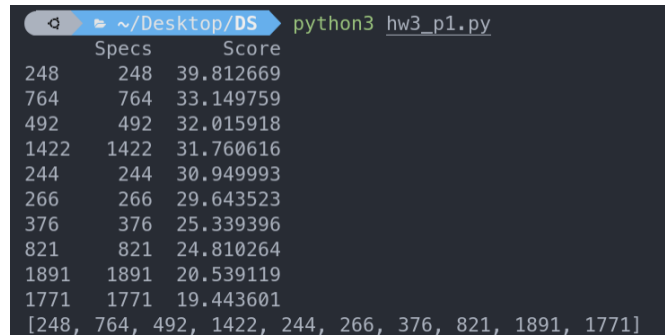
1 Problem1

1.1 My feature selection methods

I use the package SelectKBest in sklearn.feature_selection library, and use the function of f_regression to calculate the scores, and select the top 10 features by sorting the scores.

1.2 Result and my code

Code is "hw3_p1.py" in the zip file. In Fig.1, we can see the score and its corresponding index, and the last line is the features I select. For more details about the selected features, please run the code and refer to the "p1.txt" in the repo.



A terminal window showing the output of a Python script. The output is a table with three columns: 'Index', 'Specs', and 'Score'. The 'Index' column contains values from 248 to 1771. The 'Specs' column contains values from 248 to 1771. The 'Score' column contains values from 39.812669 to 19.443601. The last line of the output is a list of the selected feature indices: [248, 764, 492, 1422, 244, 266, 376, 821, 1891, 1771].

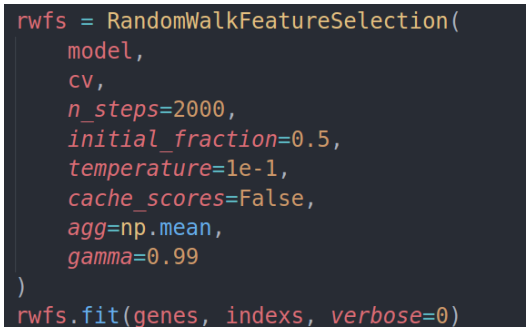
	Specs	Score
248	248	39.812669
764	764	33.149759
492	492	32.015918
1422	1422	31.760616
244	244	30.949993
266	266	29.643523
376	376	25.339396
821	821	24.810264
1891	1891	20.539119
1771	1771	19.443601
[248, 764, 492, 1422, 244, 266, 376, 821, 1891, 1771]		

Figure 1: Result of feature selection

2 Problem2

2.1 My algorithm

I choose SA as my algorithm[1], and my objective function is KNN(K Nearest neighbors, K=2). The score is the mean accuracy on the given test data and labels.



```
rwfs = RandomWalkFeatureSelection(  
    model,  
    cv,  
    n_steps=2000,  
    initial_fraction=0.5,  
    temperature=1e-1,  
    cache_scores=False,  
    agg=np.mean,  
    gamma=0.99  
)  
rwfs.fit(genes, indexs, verbose=0)
```

Figure 2: Module and my tunable parameters setting

2.2 Result and my code

Code is "hw3_p2.py" in the zip file, and for the result is too long, I print it in the "p2.txt", please run the code and check the output file to see my selected features in this part.

3 Problem3

3.1 ARIMA parameters and MSE

My code is "hw3_p3.py" in the zip file, and this code is refer to the github repo provided by TAs[2][3]. My parameters setting is $(p, d, q, P, D, Q, s) = (1, 2, 1, 3, 1, 5, 12)$, and the MSE = 104611.186.

3.2 Plot of forecast data

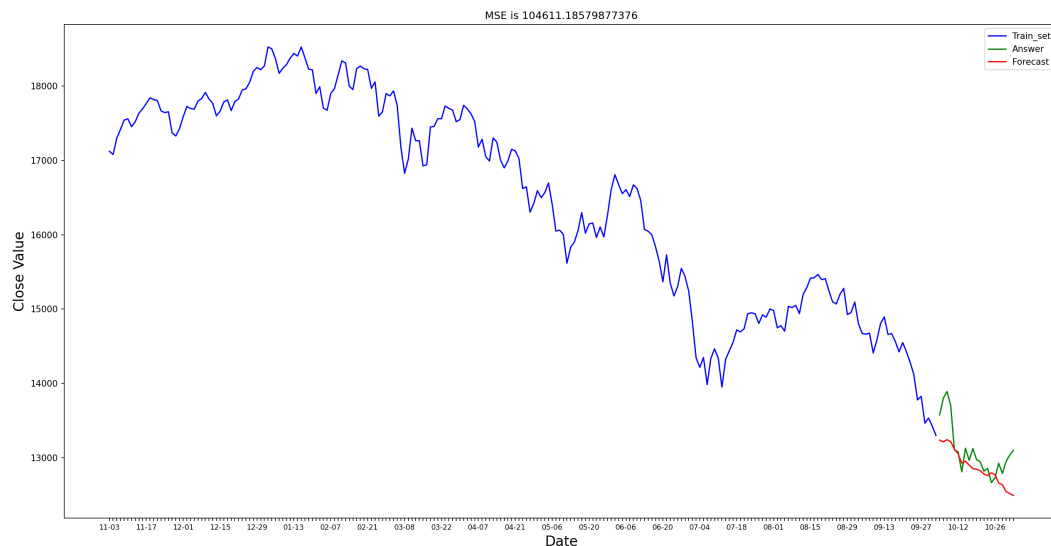


Figure 3: Forecast data

References

- [1] <https://github.com/chrka/rwfs>
- [2] <https://alkaline-ml.com/pmdarima/>
- [3] <https://github.com/alkaline-ml/pmdarima>