Algo Depth DataSet

1. Data Summary

From the dataset.csv file, we totally have 83 features(with word feature in the column name), 3 label(with word Label in the column name), 1 time information(Date), 1 assets' symbol column, and 3 assets' future return.

Here are explanations of some columns:

(1) **Date**, this column denotes date of the corresponding asset's information. At each date, we are given a portfolio with total 50 assets.

(2) **Symbol**, this column give a symbolic presentation of an asset, please note that this is not the true symbol of an asset.

(3) ***Feature***, those columns denotes features of assets, totally have 83 columns, columns with letter d means daily feature, w means weekly feature, m means month feature.

(4) **Ret_mtp1, Ret_wtp1, Ret_dtp1**, these three columns represent an asset's next month, next week, next day's return separately. These columns can be used to calculate portfolio return.

(5) **mLabel, wLabel, dLabel**, these three columns represent the classification of an assets, which you can train you algorithm based on features and label

2. Project Summary

The object of this project is to train the data with machine learning algorithms and maximize the out of sample accuracy. You can choose any algorithm with different features and corresponding label to train and optimize the algorithm.

For example:

features : {dFeature1, dFeature2, dFeature3, dFeature4, dFeature5}
label. : {dLabel}

Date	Symbol	dFeature1	dFeature2	dFeature3	dFeature4	dFeature5	dLabel
1/2/01	S8	-0.0046	-0.0072	-0.0034	-0.0012	0.0148	0
1/2/01	S9	-0.0194	0.0135	-0.0105	-0.0167	0.0439	0
1/2/01	S11	0.0335	-0.0179	-0.0438	-0.0214	0.0133	1
1/2/01	S1	0.0146	-0.0810	-0.0180	0.0179	-0.0217	1
1/2/01	S28	0.0679	-0.0229	-0.0066	0.0291	-0.0123	1
...	...	...	...	...	...	...	...
1/2/01	S47	0.0335	0.0130	-0.0034	-0.0035	-0.0170	1
1/2/01	S15	0.0288	0.0088	0.0091	-0.0244	0.0039	0
1/2/01	S20	0.0169	-0.0288	0.0080	-0.0425	0.0189	1
1/2/01	S2	0.0014	0.0063	0.0376	0.0129	0.0213	0
...	...	...	...	...	...	...	...
1/3/01	S27	0.0083	-0.0480	0.0077	-0.0057	0.0240	0
...	...	...	...	...	...	...	...

Table 1, Sample data with 50 stocks in one day, details are in dataSet.csv in the same folder
dLabel column shows the class of the stock should be belong to.