PROJECT PHASE – 3

CS – 480
Artificial Intelligence

Fall -2015

submitted To

Mustafa Bilgic

by

Darpan Patel
(A20345898)

Detail Description of the choose_the_best_classifier method.

There are so many parameters in the sklearn.metrics to choose the best classifiers among the BournolliNB , Logistic Regression and SVC.

1.) Accuracy
2.) Precision
3.) Recall
4.) Log-Loss
5.) jaccard similarity matrix etc...

- I have tried to find out all the parameters starting from Accuracy. I have tried for all the datasets given by the professor and found the accuracy
- The trial code is in the All-Trial.txt for the reference. But, I couldn't make decision on the basis of accuracy statistics from the accuracies obtained by all the three classifiers. The result gives the classifier which has highest accuracy but less wealth.

- If the wealth is not greater than all the other classifiers, then it is not the best classifier chosen by my Agent_dpatel96. Wealth is the main criteria to be the best classifier so my agent has to choose the classifier which gives the highest wealth. So I don't consider accuracy as decision criteria. I tried the other options too from the above list.

- Then I tried to calculate the predicted value of validation dataset from training dataset and then compare it to the y_val. This is also in All_trial.txt. This method gives me the count of product being "Excellent" and "Trash" for predicted value and given value(y_val) for all the classifiers and all datasets(3). This method gives the nearest predicted value(X_val) of the given value of(y_val). But it gives various results for every dataset so the agent couldn't take proper decision to choose the best classifier.

- I tried all other parameters from sklearn.metrics mentioned above using their methods but none of the parameters give the accurate output to make proper decision. These all trials are mentioned in the file named All_Trial.txt

Main Criteria for choosing a best classifier from BournolliNB , Logistic Regression and SVC is "wealth". If the wealth is higher of chosen classifier, then it is best classifier so I selected the decision parameter as "Wealth". Agent finds the wealth of each classifier based on the validation dataset.

Description:

1.) Train each classifier using X_train and y_train.
2.) Calculate the wealth on validation dataset (X-val and y_val) using simulate_agents.

NOTES : simulate_agents function is directly used from given file simulate_agents_phase3.py(REFRENCE). I have tried to import this method from simulate_agents_phase3.py but it gives me ERROR(Import Error) so, I used this method in my agent Agent_dpatel96.py.

3.) I got dictionary of wealths value returned from simulate_agents method, calculated maximum wealth from it and returned the untrained object of chosen classifier.
4.) Here chosen classifier is best classifier because it gives maximum wealth on validation dataset (X_val and y_val). If it gives maximum wealth on validation dataset then definitely it gives maximum wealth on test dataset.

NOTE : I can choose the parameters like accuracy, precision if the given classifier is other than “SVC”. Because of “SVC ” I need to choose wealth as decision parameter.

Output :

```
--------------------------------------------------
SIMULATION RESULTS ON dataset1
--------------------------------------------------
```

Wealth (the larger the better)

Agent_bnb:	$1,775,950.00
Agent_lr:	$1,638,100.00
Agent_svc:	$1,596,800.00
Agent_dpatel96:	**$1,775,950.00**

Log-loss (the smaller the better)

Agent_bnb:	332.25
Agent_lr:	373.89
Agent_svc:	408.75
Agent_dpatel96:	**332.25**

0/1 Loss (the smaller the better)

Agent_bnb:	88
Agent_lr:	115
Agent_svc:	147
Agent_dpatel96:	**88**

SIMULATION RESULTS ON dataset2

Wealth (the larger the better)

Agent_bnb:	$1,507,950.00
Agent_lr:	$1,717,100.00
Agent_svc:	$1,540,800.00
Agent_dpatel96:	**$1,717,100.00**

Log-loss (the smaller the better)

Agent_bnb:	553.13
Agent_lr:	487.71
Agent_svc:	564.25
Agent_dpatel96:	**487.71**

0/1 Loss (the smaller the better)

Agent_bnb:	250
Agent_lr:	223
Agent_svc:	293
Agent_dpatel96:	**223**

SIMULATION RESULTS ON dataset3

Wealth (the larger the better)

Agent_bnb:	$795,950.00
Agent_lr:	$810,100.00
Agent_svc:	$1,102,800.00
Agent_dpatel96:	**$1,102,800.00**

Log-loss (the smaller the better)

Agent_bnb:	571.94
Agent_lr:	566.91
Agent_svc:	413.10
Agent_dpatel96:	**413.10**

0/1 Loss (the smaller the better)

Agent_bnb:	250
Agent_lr:	255
Agent_svc:	165
Agent_dpatel96:	**165**

REFRENCE :

1.) “sklearn” some classes like metrics , svm, linear_model.
2.) simulate_agents_phase3.py for simulate_agents method