Project 4 - CS378

Isac Simkin

March 2019

1 Results

With my code I was able to analyze the data from the grades.csv file and with those results, I generated 4 separate tables demonstrating the accuracy of each model taking into account the number of quizzes used.

In the image inserted in the next page you can see how all 4 classification algorithms compare to each other. Each different shape represents a different score used to analyze the performance of each method. The 'Blue Square' or is used for accuracy score, the 'Red Triangle' is used for F-Score, the 'Green Triangle' is used for precision score, and the 'Magenta Hexagon' is used for recall score.

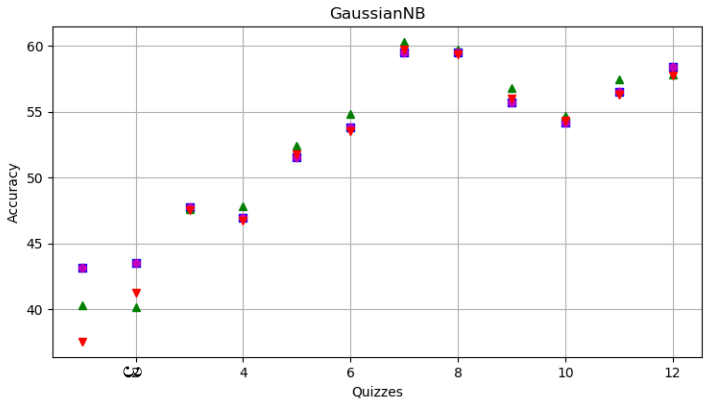
My algorithm started by importing all necessary packages. Then, open the csv file and filled all the missing quiz scores with '0' to avoid a Null/None error. After that, I wrote the function which converts the final scores (number) and creates a new column according to the cuts given in the homework handout. Following that, there is another function that is in charge of plotting all of the scores used in each set of quizzes using the shapes described above.

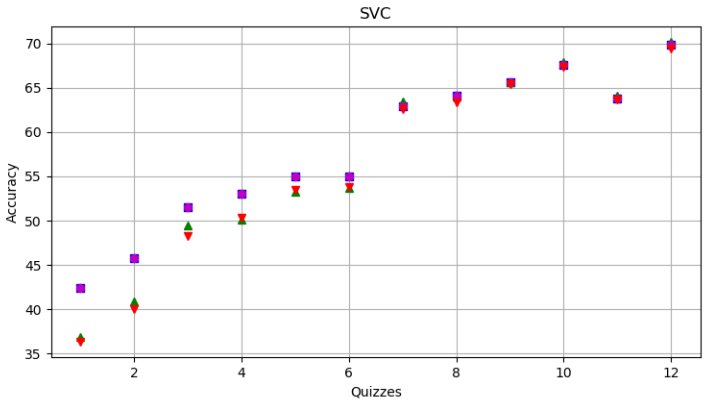
After the 'plotter' function, there is the core of my algorithm. It takes a range of 12-24 and plugs it into the 'X' values, this is done to add a quiz for every loop pass. I defined each model, fitted the split data, and then predicted using the built in functions. For each of the scores I used the 'weighted' average parameter because it supports label imbalance which was a main concern while analyzing the results. Data splits may have different labels and the 'weighted' option is the only one that takes it into consideration.

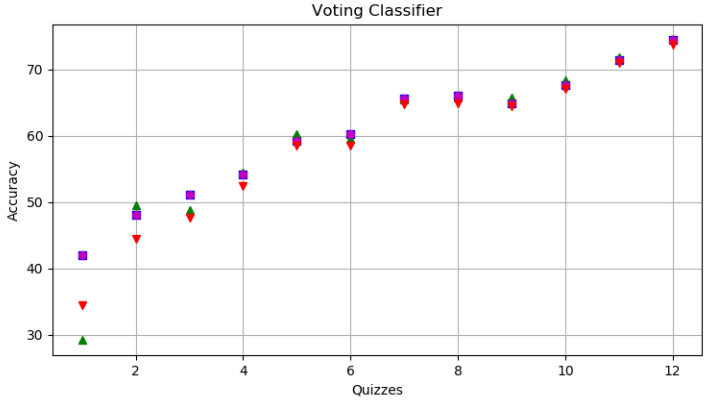
Even though I was not able to graph the cross-validation scores, they are printed in the console when the program is ran using the instructions in the README.txt file. Nonetheless, in the graph below you

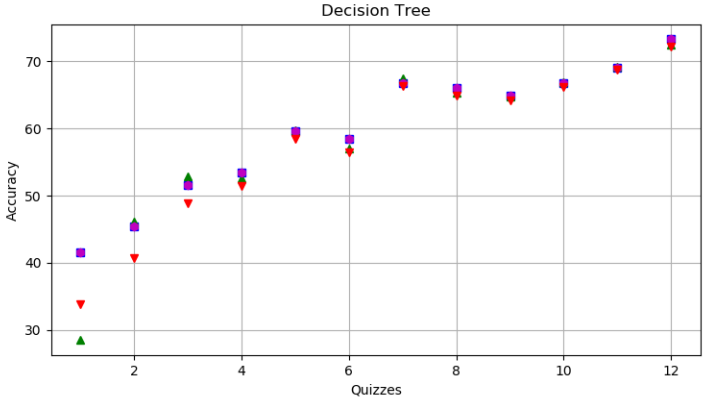
can see there is a similar trend with all the models. The four have an upwards trend that correlates to the number of quizzes included in the set.

Moreover, there is a slight downward trend when it reaches the 7 quizzes in all but SCV, where it occurs after the 10 quiz set. However, accuracy increases right after that point overwriting the maximum accuracy in all except in the GaussianNB. The general upwards trend in the graphs show that as there is more data to consider, there is are more accurate predictions. Nonetheless, because of the nature of the quizzes, meaning the material evaluated in them, there might be students who do not end up performing as expected because they fall from the projections after Quiz 07 or Quiz 08 are added to the grade.


GaussianNB
Accuracy
Quizzes


SVC
Accuracy
Quizzes


Voting Classifier
Accuracy
Quizzes


Decision Tree
Accuracy
Quizzes