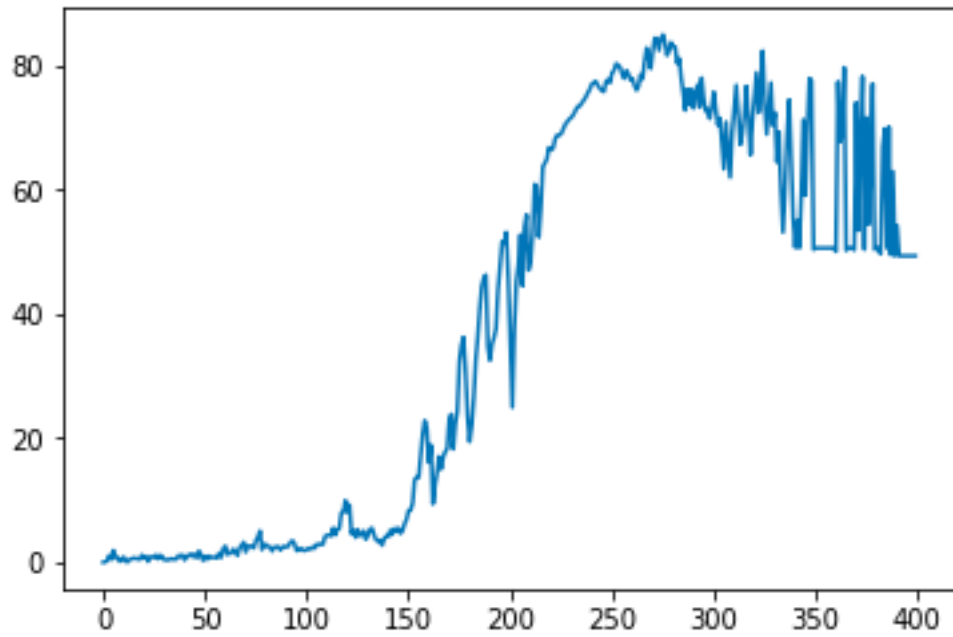


Machine Learning Project 2 Report

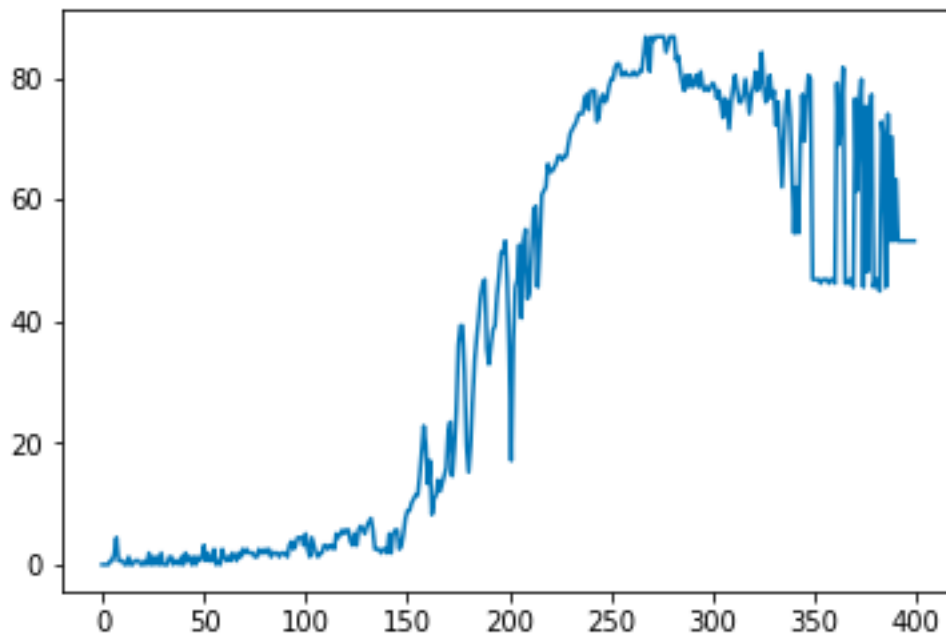
Yun Bao
CSE 474
2018/11/01

1. Linear Regression of Human observed concatenate data set

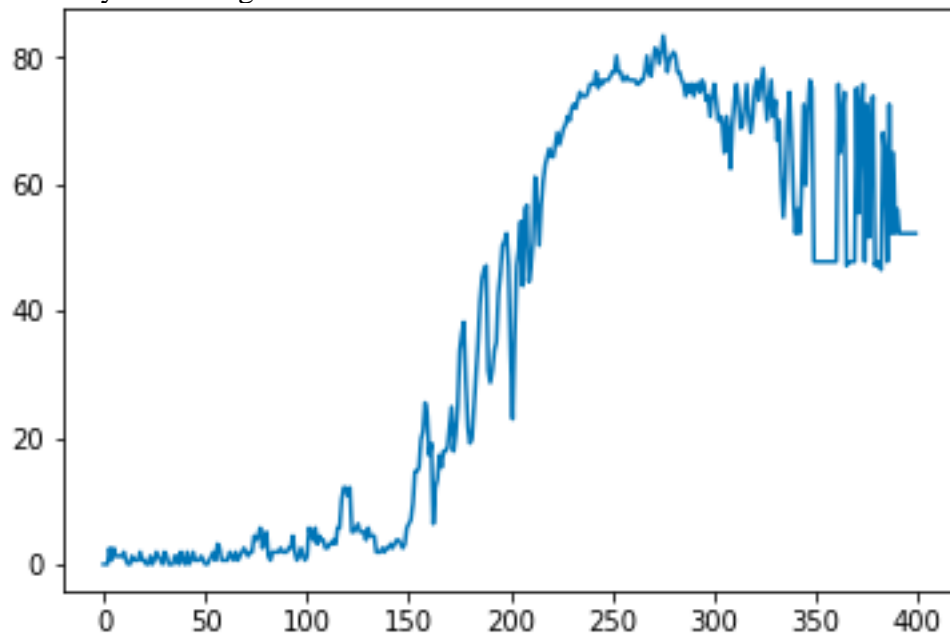
Accuracy for training data:



Accuracy for validation data:



Accuracy for testing data:

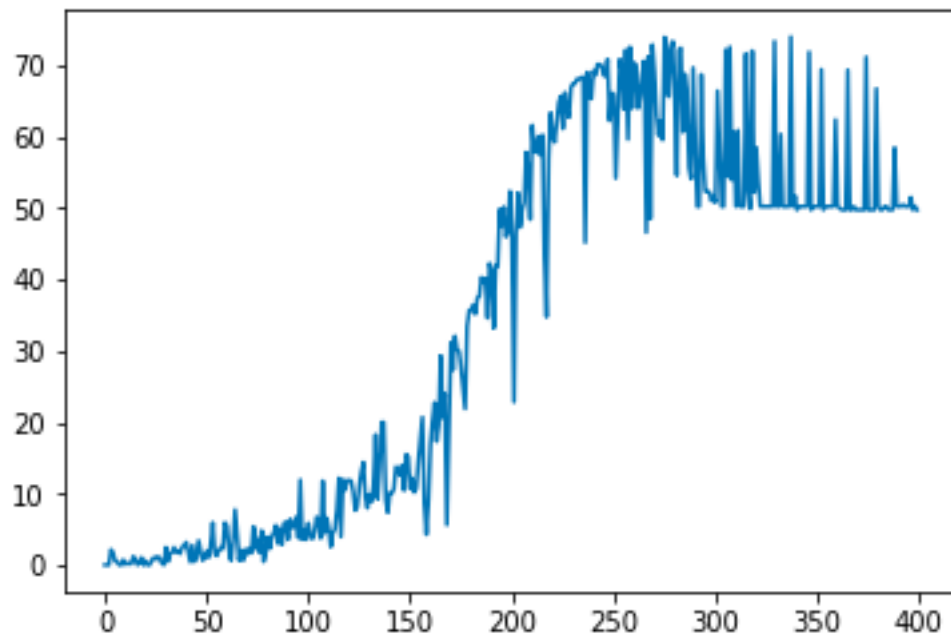


```
E_rms Training    = 0.3851
E_rms Validation  = 0.40037
E_rms Testing     = 0.4284
```

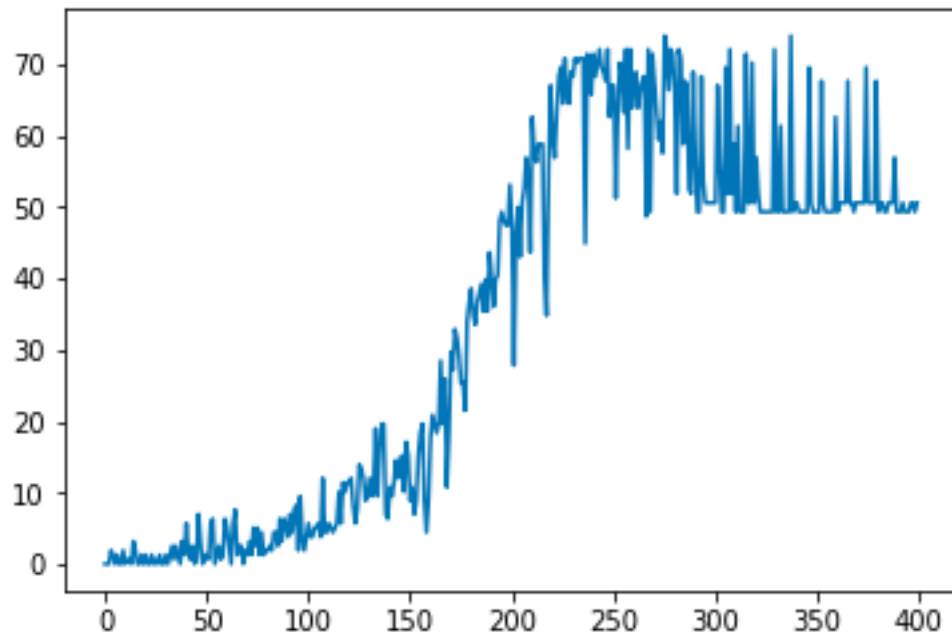
For concatenation, I concatenate the same pair features together. Therefore, for each writer has 18 features, and give them the target, either 0 or 1 from the same pair file or the different pair file. I stored them in a matrix, and split them into two files which are data file and target file. Then I used project 1.2 linear regression solutions to split 80% data for training, 10% data for validation, and 10% data for testing. Then using linear approach to modelling the relationship between a scalar response and one or more explanatory variables. The accuracy I got is around 50%-60%. According to the graph, the accuracy went up above 80% quickly, then went down around 50% constantly.

2. Linear Regression of Human observed subtract data set

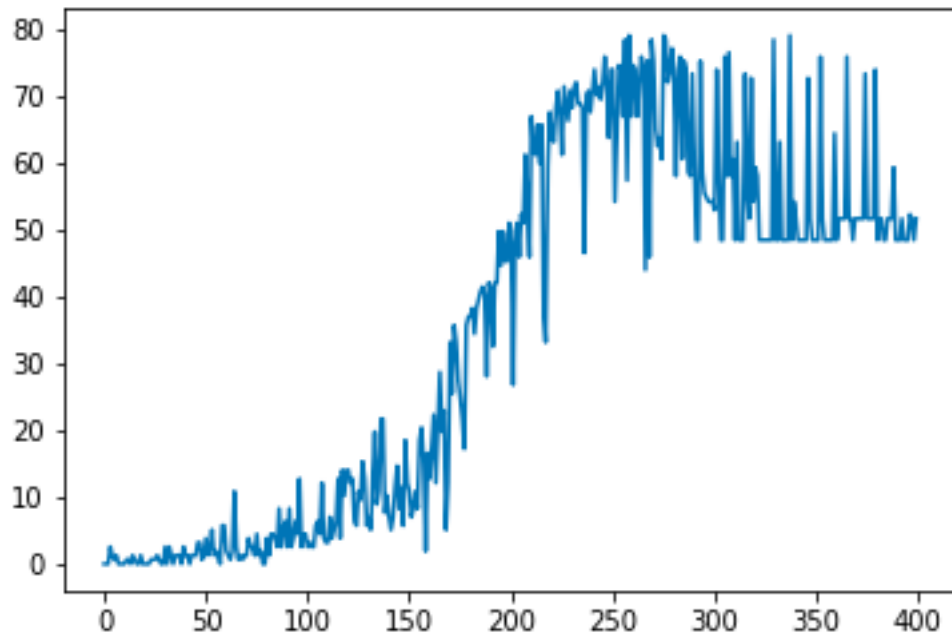
Accuracy for training data:



Accuracy for validation data:



Accuracy for testing data:

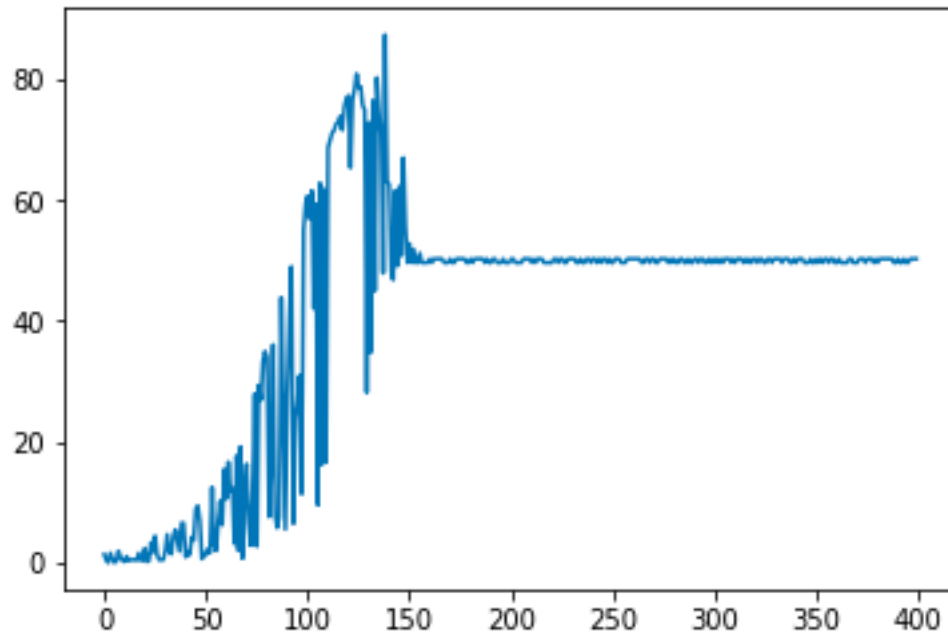


```
E_rms Training    = 0.42123
E_rms Validation  = 0.44555
E_rms Testing     = 0.43797
```

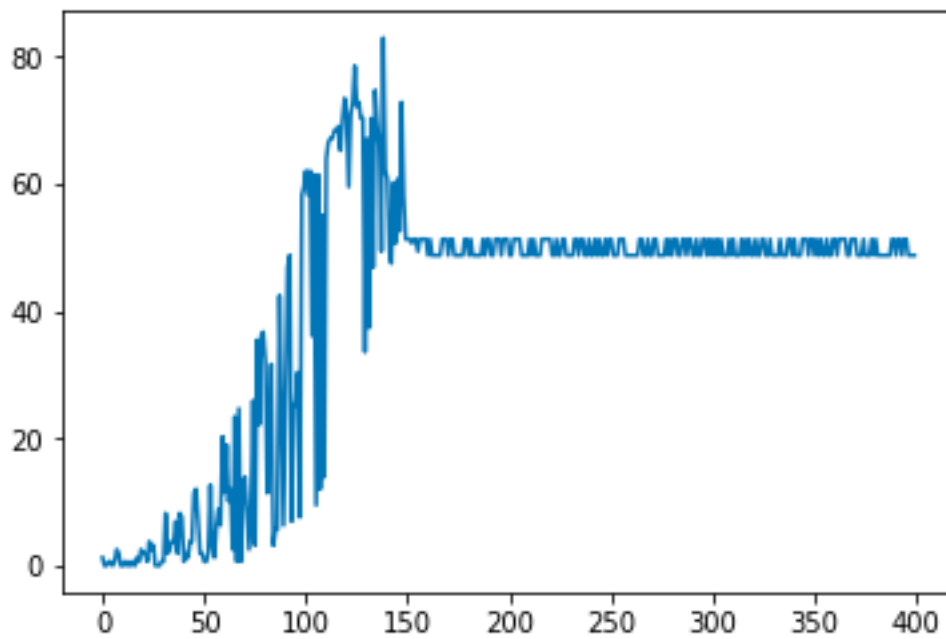
For subtraction, what I did is similar to concatenation, but instead of concatenate two images' 18 features together, I subtracted them and give them the target value corresponding to the same pair and different pair files. Therefore, each writer would still have 9 features. Then I also applied project 1.2 linear regression to train the data set. The accuracy is similar to the concatenation, but for subtraction, the graph has obvious fluctuations, and it still went up above 75%, then came down to 50% eventually.

3. Linear Regression of GSC concatenate data set

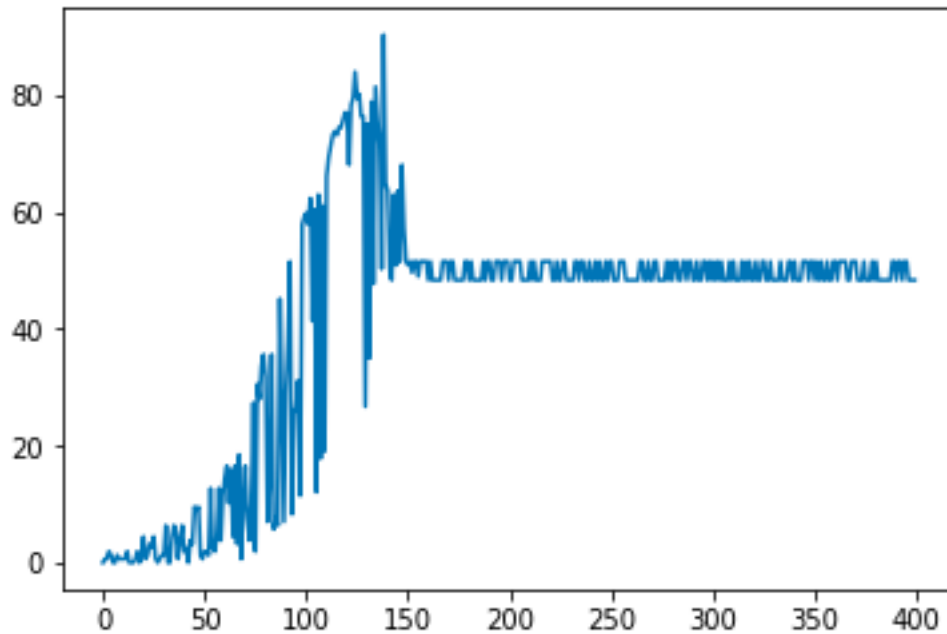
Accuracy for training data:



Accuracy for validation data:



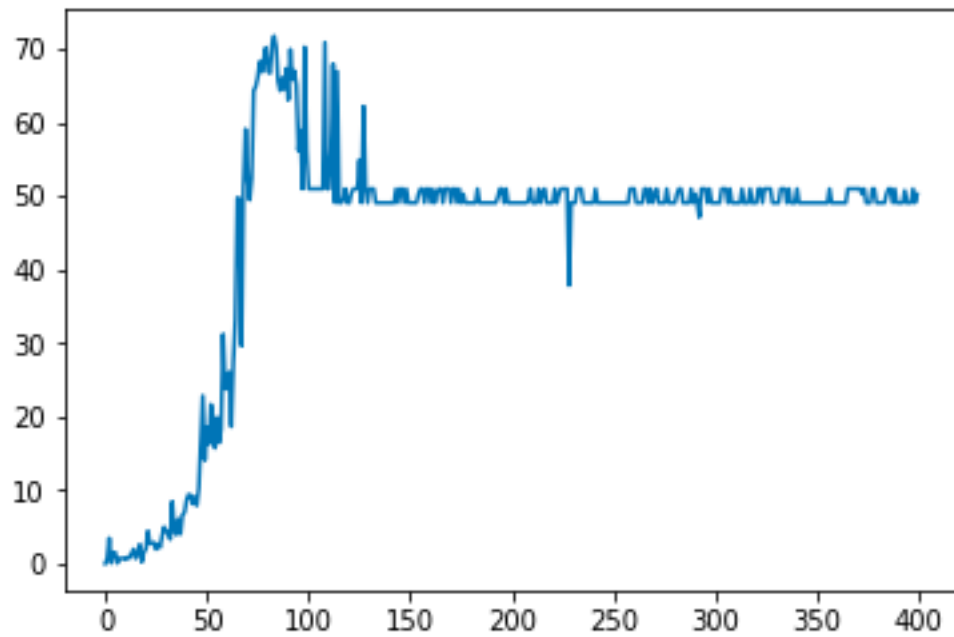
Accuracy for testing data:



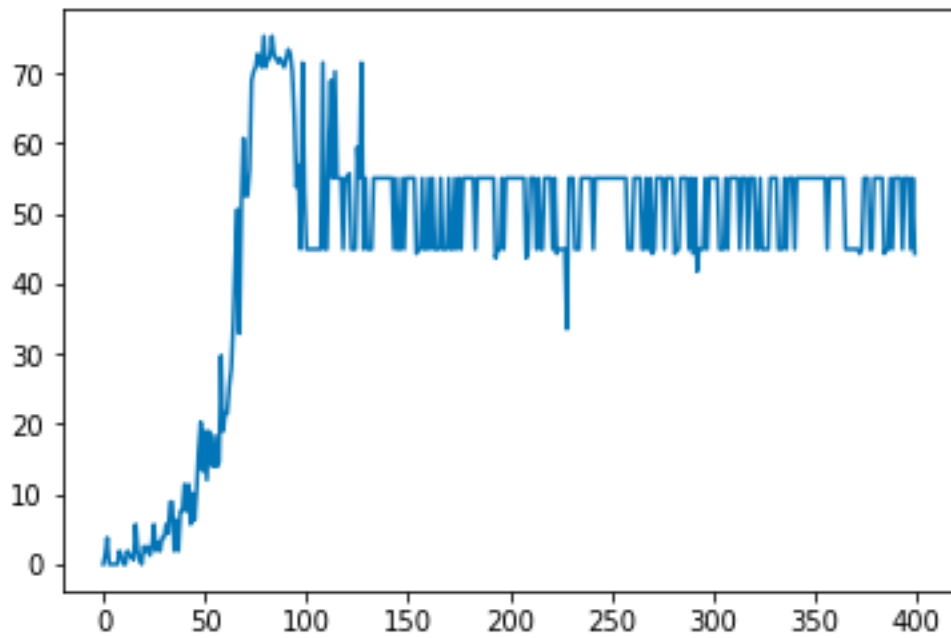
For the GSC concatenate data set, basically I use the same code from the concatenation of human observed data set and do the same process for the GSC data set. The accuracy hasn't changed much, but the graph is totally different from the human observed graph. The accuracy has obvious fluctuations in the beginning of learning data, then the accuracy starts to be more stable and constant.

4. Linear Regression of GSC subtract data set

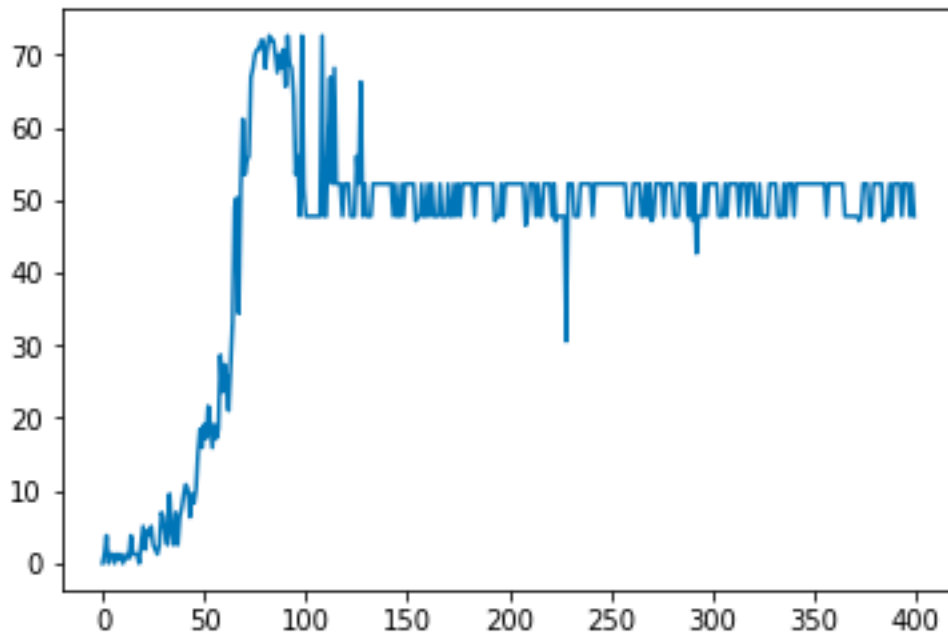
Accuracy for training data:



Accuracy for validation data:



Accuracy for testing data:



For the GSC subtract data set, I use the same code from human observed subtract data set to get the data frame for the GSC for subtraction. The accuracy that I get is similar to the GSC concatenate data set, but subtraction doesn't have much fluctuations, and in the beginning, the accuracy goes up quickly, and come down to 50% quickly as well. The only thing that is similar to GSC concatenate data set is the accuracy become more stable and constant at around 50%.