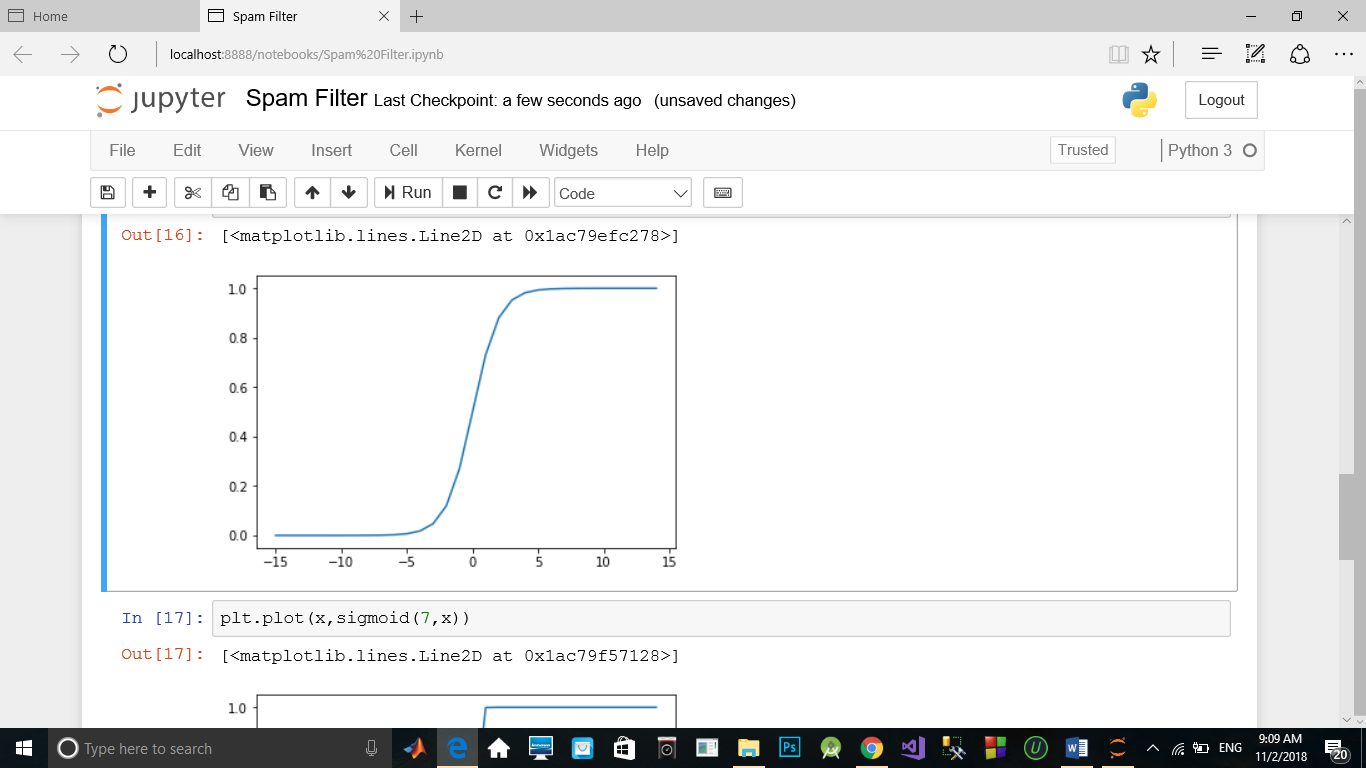
**Assignment 2 Report**

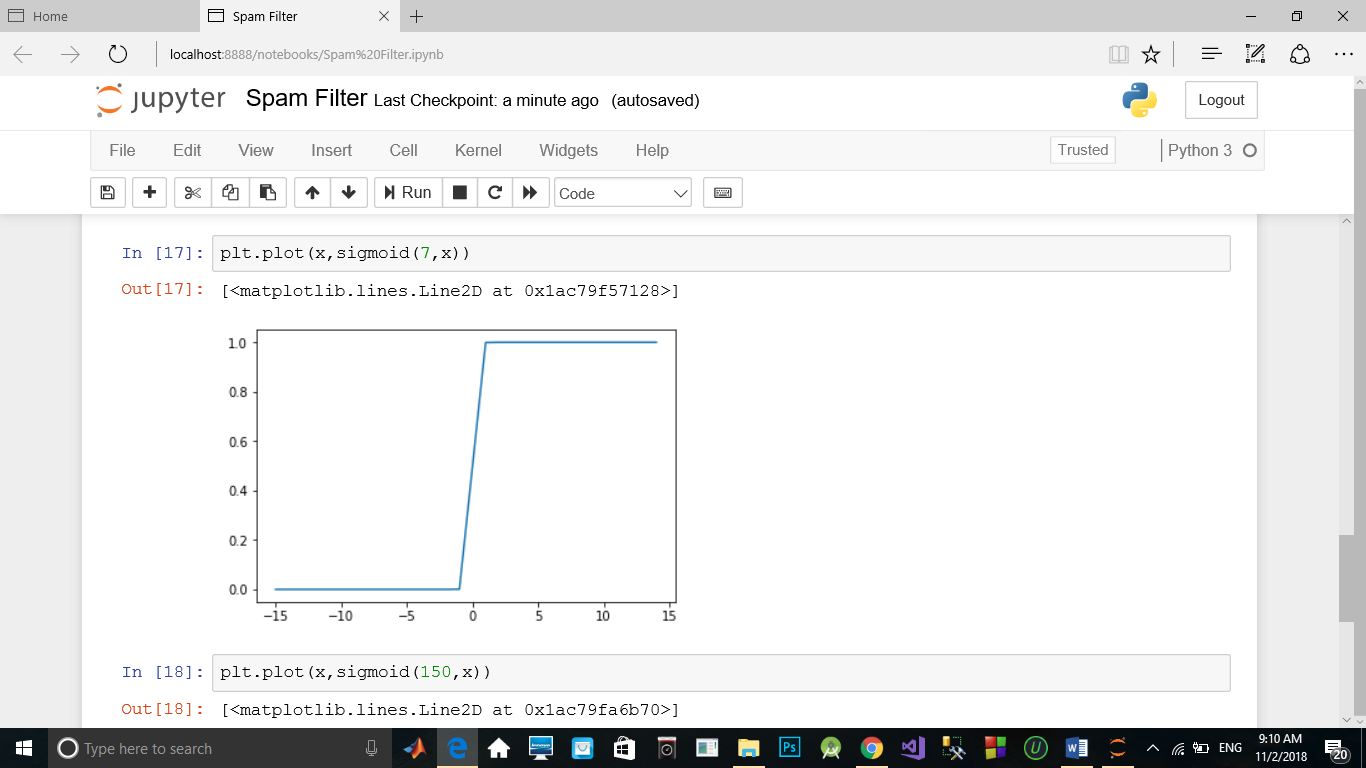
***Mohamed Saleh Mahmoud***

***201501343***

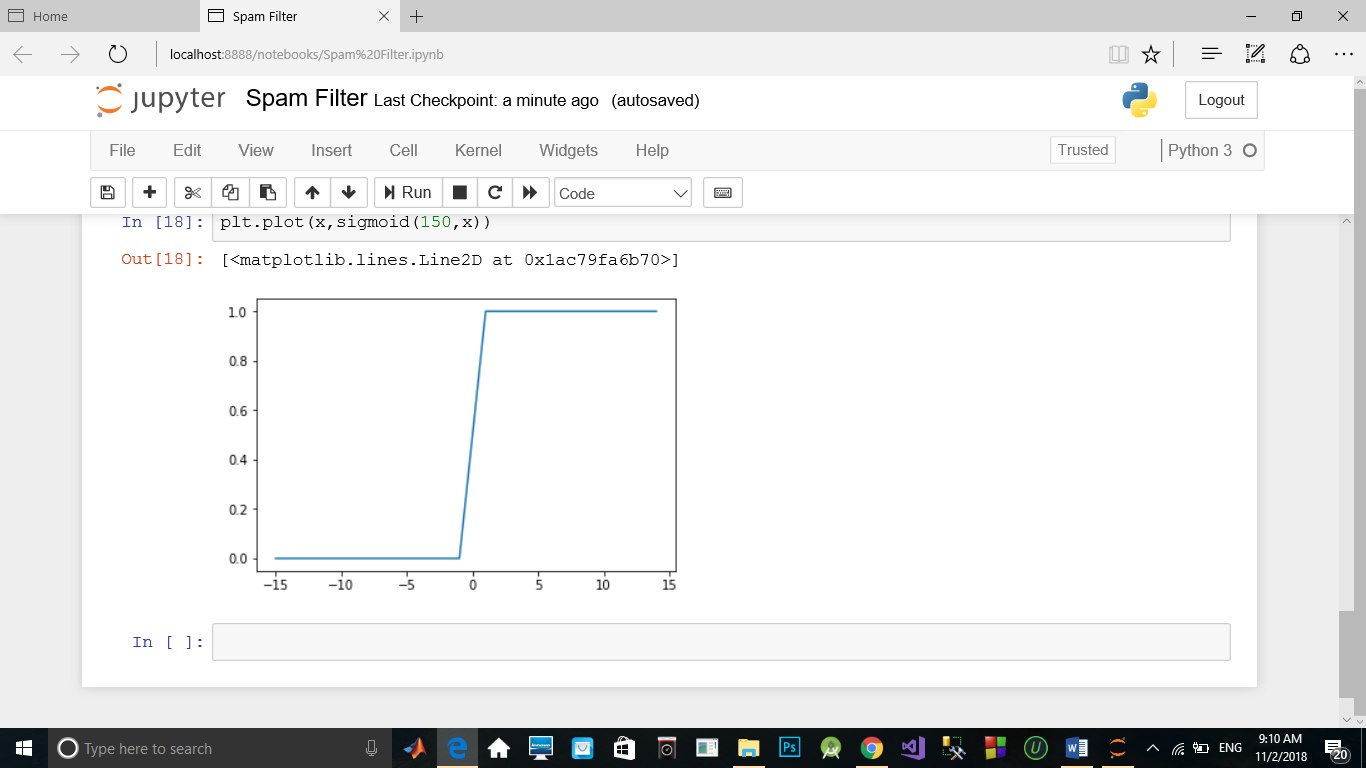
***Problem 1: -***



(In case of w = 1)



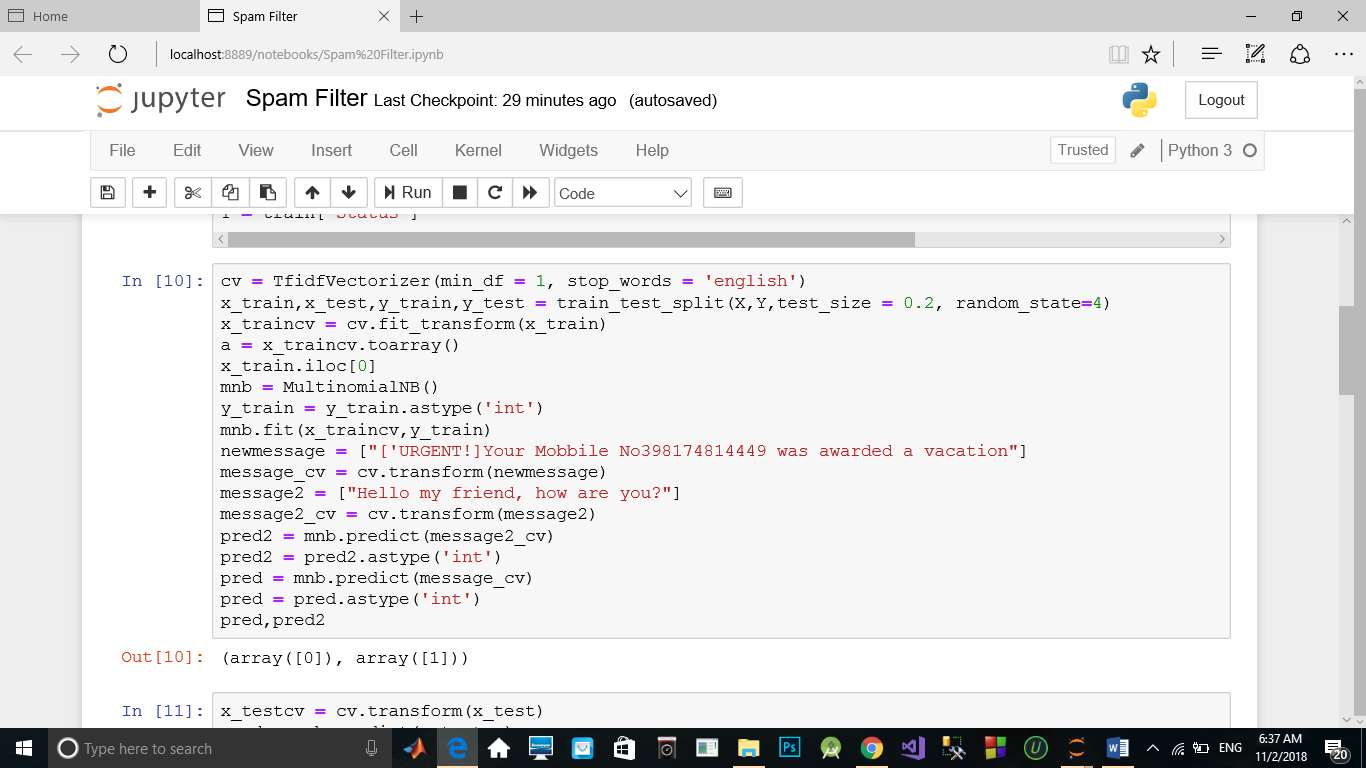
(In case of w = 7)

 (In case of w = 150)

***Discussion:-***

In these plots it’s clear that some points have values above 0.5( the threshold) when we had the weight was 150 which then will be misclassified and overfitted

***Spam Filter Part:-***



The first message was classified as a spam and the second one was ham.

In this part, we used the TfidfVectorizer object to extract the features from the texts and get these values to be inputted to an estimator. In this code the estimator is a naïve bayes estimator called MultinomialNB. The test data were 20% of the original data set and the result of the estimator was almost 96 percent accurate.