Ian Zimmer

0028719394

1 Late Day used

CS373 Homework 3

2 Perceptron Theory

2.1

F(x) = { 1 if Sum($w_j x_j$) > bias }

{ 0 if Sum($w_j x_j$) <= bias }

Yes because by allowing the hyperplane to be moved from the origin, the algorithm can represent situations where the threshold isn't the origin.

2.2

a i: Will not give > .95 classification accuracy because the data is not dispersed linearly and perceptron will not be able to find a linear threshold that separates the positive from the negative. If you can add dimensions to the algorithm, then the data can be separated and it can achieve >.95.

a ii: Will not give > .95 classification accuracy because the data is not dispersed linearly and perceptron will not be able to find a linear threshold that separates the positive from the negative. If you can add dimensions to the algorithm, then the data can be separated and it can achieve >.95.

b i: Will not give > .95 classification accuracy because the data cannot be divided around the origin.

b ii: Will not give > .95 classification accuracy because the data is not dispersed linearly and perceptron will not be able to find a linear threshold that separates the positive from the negative. If you can add dimensions to the algorithm, then the data can be separated around a place other than the origin and it can achieve >.95.

c i: Will give > .95 classification accuracy because the data is separated binarily around the origin

c ii: Will give > .95 classification accuracy because the bias could be 0 and the data would be separated binarily around the origin

d i: Will not give > .95 classification accuracy because any threshold that goes through the origin could not separate the data into two classes well

d ii: Will give > .95 classification accuracy because the bias would provide a threshold away from the origin that would binarily separate the data

2.3

Matching: b=b

Does not match: b = b+ry

3 Naïve Bayes Theory

3.1

$P(c^+|d) = P(d|c^+)/ P(d)$

3.2

With each document being length L, there will need to be 2^L+2 parameters.

3.3

Since there is the unigram assumption, we only need L+1 parameters.

3.4

$P(c^+)$ = (total c^+ documents)/((total c+ documents) + (total c- documents))

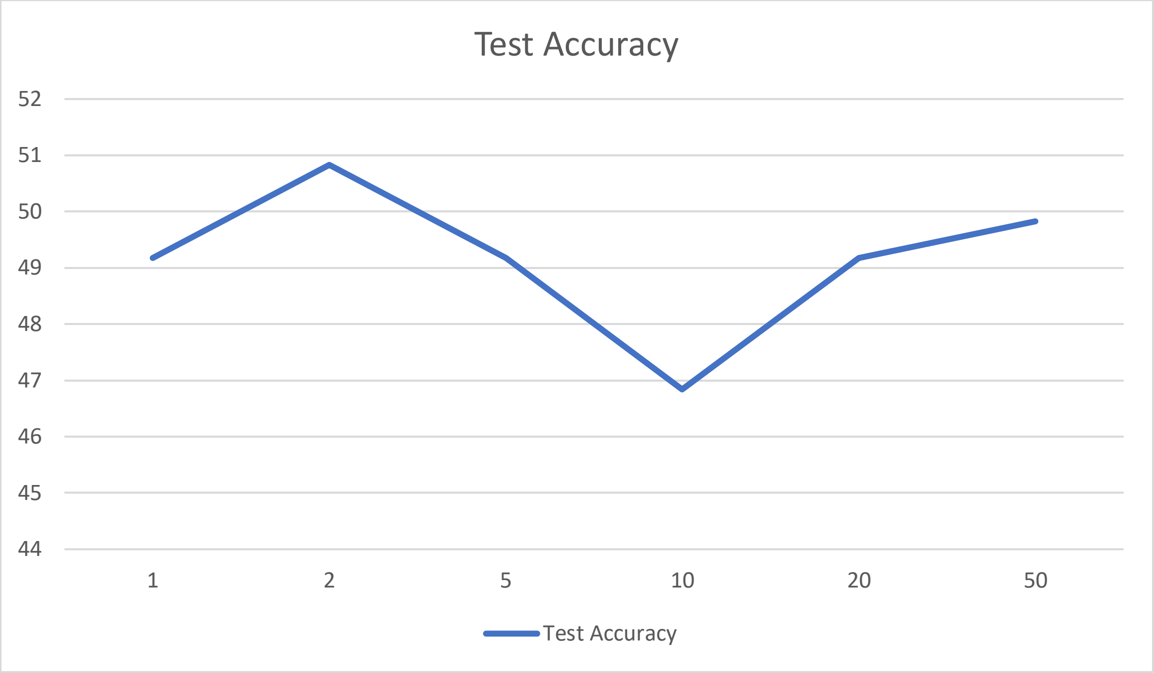
P(c-) = (total c- documents)/((total c+ documents) + (total c- documents))

4 Analysis

4.1

The performance is worse than the respective models with a bias term. This is because the hyperplane is only allowed to be drawn through the origin which is less accurate than a hyperplane that can be drawn anywhere.

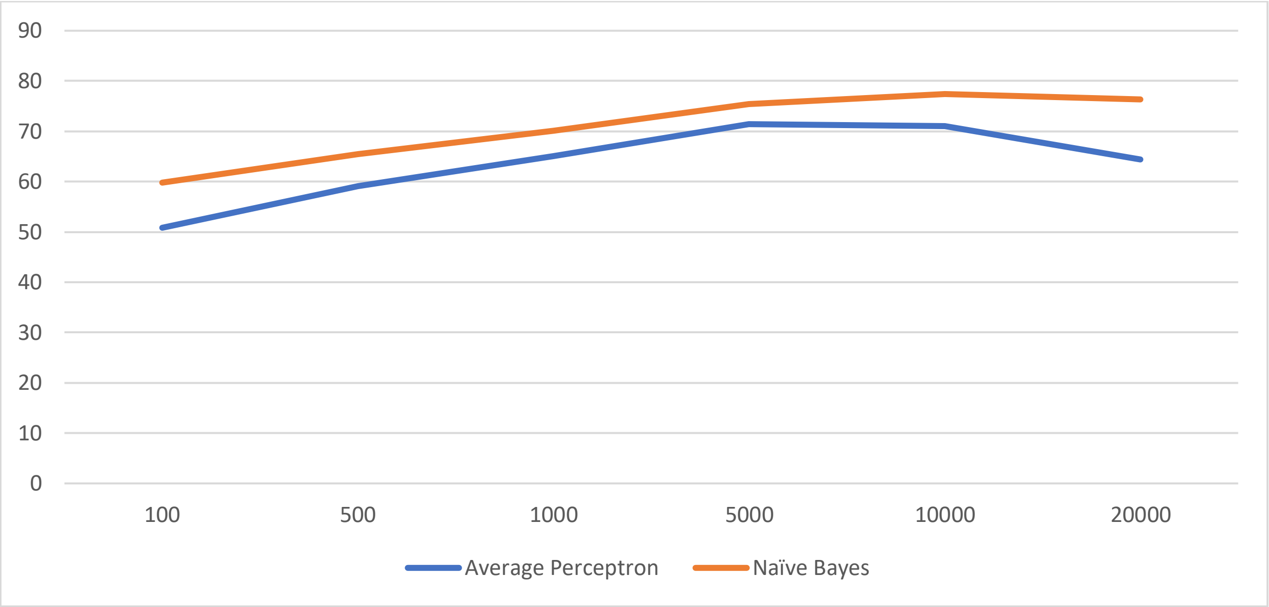
4.2



Iterations	Test Accuracy
1	49.17
2	50.83
5	49.17
10	46.84
20	49.17
50	49.83

No it does not. Since perceptron looks at a singular point, the algorithm is unaware of the future and can only account for it once it has been reached. It will not be able to reach 100 even when the whole dataset has been traversed because perceptron weights the overall set and provides a general prediction.

4.3



	Average Perceptron	Naïve Bayes
100	50.833	59.8

500	59.14	65.45
1000	65.12	70.1
5000	71.43	75.42
10000	71.1	77.41
20000	64.45	76.41

The performance of both algorithms increases as the vocabulary size increases until it gets to size 10000, then It levels off. This could be because the a word could occur only once in one review and slightly skew other predictions.