Name- Faisal Tariq Vora

PUID- 0028770609

CS373 HW 3:

2. Perceptron (45 points)

2.1 Theory

2.1.1Equation: $f(x) = 1$ if $\Sigma w_j x_j + b_j > 0$

0 if $\Sigma w_j x_j + b_j <= 0$

The perceptron with a bias term is more expressive as the one without the bias, the separation line passes through the origin. With the bias term we can put the separation line anywhere on the plane. There is a way to overcome this problem by changing the features. But for this problem I am going to assume that we are not going to change the original features.

2.1.2

i) a) No, it would not give a high accuracy as the data points are indistinguishable in the circles and a single line cannot divide them properly.

b) No, it would not give a high accuracy as the data points are indistinguishable in the circles and a single line cannot divide them properly.

c) Yes, it can give a high classification accuracy as at x axis point 0 we can draw a straight line up that separates the two classes with a high accuracy.

d) No, without the bias the line will have to pass through 0. Even if we draw a 45 degree line towards the right there will be a lot of purple|violet dots will be on the wrong side of the line.

ii) a) No, it would not give a high accuracy as the data points are indistinguishable in the circles and a single line cannot divide them properly.

b) No, it would not give a high accuracy as the data points are indistinguishable in the circles and a single line cannot divide them properly.

c) Yes, as mentioned for without bias, we can separate it by drawing a straight line up at 90 degree from the 0 point

d) Yes, With the bias we will be able to separate them with high accuracy as we can draw a line from point 3.7 straight up which separates the two class with high accuracy

2.1.3 The update rule for bias term b of a vanilla perceptron given the learning rate r and the label y when the classifier label does not match during training is

$b = b + ry$,

When it matches, b remains the same : b = b.

Problem 3: Naïve Bayes

3.1 theory

1. The equation of P(c+|d) in terms of P(d|c+) is follows:

P(c+|d) = (P(d|c+)*P(c+))/P(d)

For Naïve Bayes we can drop the denominator as we will be comparing the two probabilities of the two classes(+ and -)

2. We will need to learn 2V+2 parameters in order to be able to correctly estimate P(d|c+) for any given document without making independent assumptions

3. We will need to learn 2V parameters in order to be able to correctly be able to estimate P(d|c+)

4. P(c+) = (size(c+)|(size (c-)+size (c+))

P(c-) = (size(c-)|(size (c-)+size (c+))

Part 4. Analysis:

1

With the bias term the values are

Perceptron Results:

Accuracy: 80.73, Precision: 79.22, Recall: 82.43, F1: 80.79

Averaged Perceptron Results:

Accuracy: 79.73, Precision: 77.70, Recall: 82.43, F1: 79.99

And the values *without the bias term* are:

Perceptron Results:

Accuracy: 78.07, Precision: 79.70, Recall: 74.32, F1: 76.92

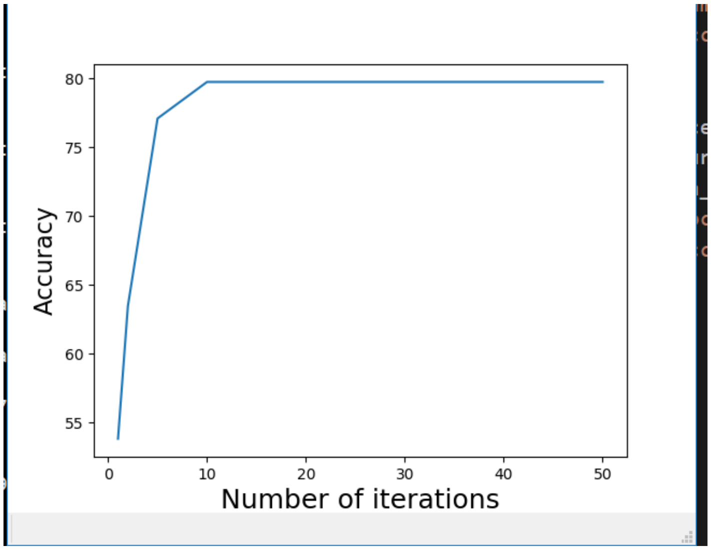
Averaged Perceptron Results:

Accuracy: 81.06, Precision: 82.26, Recall: 78.37, F1: 80.27

We can see that the accuracy for perceptron decreased without the bias term and even though the accuracy for the average perceptron increases, it does not increase by a lot. Having the bias term increases the overall accuracy.

2. number of iterations

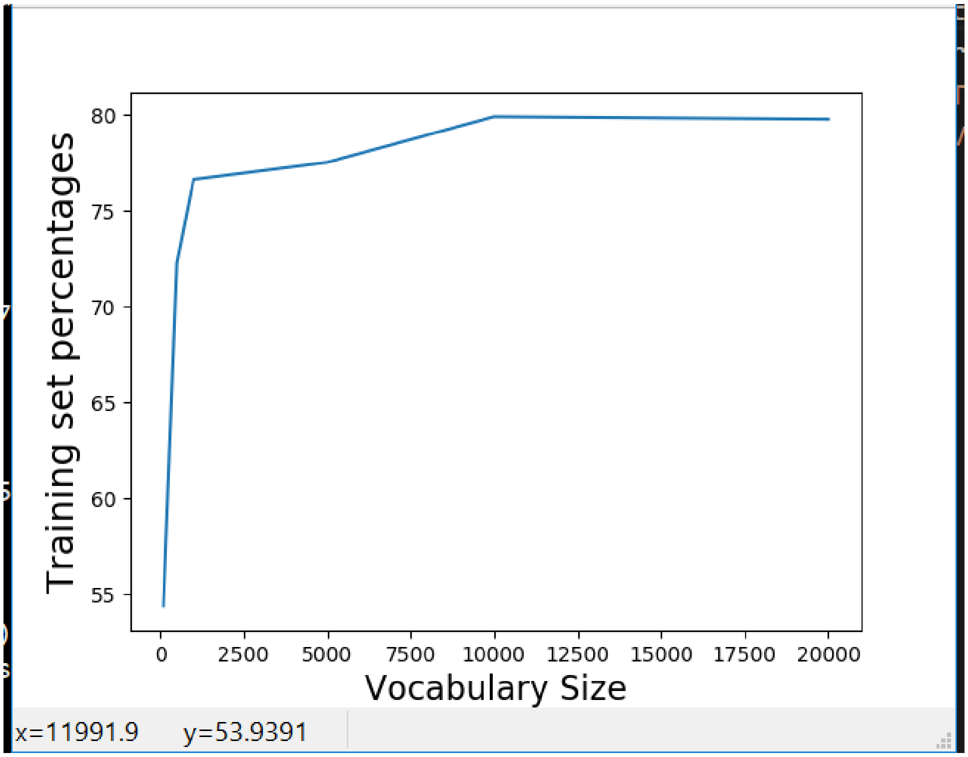
1 - 65.78,2 - 70.10,5 - 74.42,10 - 72.09,20 - 76.74,50 - 77.08



^Graph for Number of Iterations vs accuracy for average perceptron

Perceptron does not converge after a certain number of iterations. After some iterations it remains constant. This is because increasing the number of iterations does not always improve the trained vector. After some point the accuracy does not reach 100, which is good as we are making sure that we are not memorizing the data.

3.



As we increase the vocab list the accuracy increases steadily. After a certain vocab size, increasing it more does not increase the accuracy. It remains stable if the vocab size is increased even further after a certain vocab size as seen from the graph above.