CS 373 Spring 2019: Homework 3

Youngsik Yoon, 0029846135

April 1, 2019

2 Perceptron

2.1 Theory

1. The equation with the bias term included is

   $$f(x) = \begin{cases} 1 & \text{if } \sum w_j x_j + b > 0 \\ 0 & \text{if } \sum w_j x_j + b \leq 0 \end{cases}$$

   where b is the bias. Perceptron with a bias term is more expressive compared to the one without bias because the separation of the data no longer has to pass through the origin. The bias term allows the separation to be translated and not pass through the origin. This allows a more diverse set of distributions to be classified with higher accuracy.

2. (a) Neither (i) nor (ii) will give a high classification accuracy because the data cannot be separated linearly.

   (b) Neither (i) nor (ii) will give a high classification accuracy because the data cannot be separated linearly.

   (c) Both (i) and (ii) will give a high classification accuracy because the data can be separated linearly.

   (d) Only (ii) will give a high classification accuracy because the bias allows a translation. (i) will not give a high classification accuracy because the data cannot be separated linearly if the origin needs to be passed.

3. When the classifier label doesn't match the gold label during training, the update rule for *bias* of a vanilla perceptron is

   $$b = b + \gamma \cdot y$$

   When the classifier label does match the gold label, the *bias* is not updated.

3 Naive Bayes

3.1 Theory

1. The equation for $P(c^+|d)$ in terms of $P(d|c^+)$ using Bayes theorem is

   $$P(c^+|d) = \frac{P(d|c^+)P(c^+)}{P(d)}$$

2. In order to correctly estimate $P(d|c^+)$ for any given document without making independence assumptions, 2^l parameters are needed.

3. With the unigram assumption, V parameters are needed.

4. The equation to estimate $P(c^+)$ is

   $$P(c^+) = \frac{\text{size}(c^+)}{\text{size}(c^+) + \text{size}(c^-)}$$

   , and the equation to estimate $P(c^-)$ is

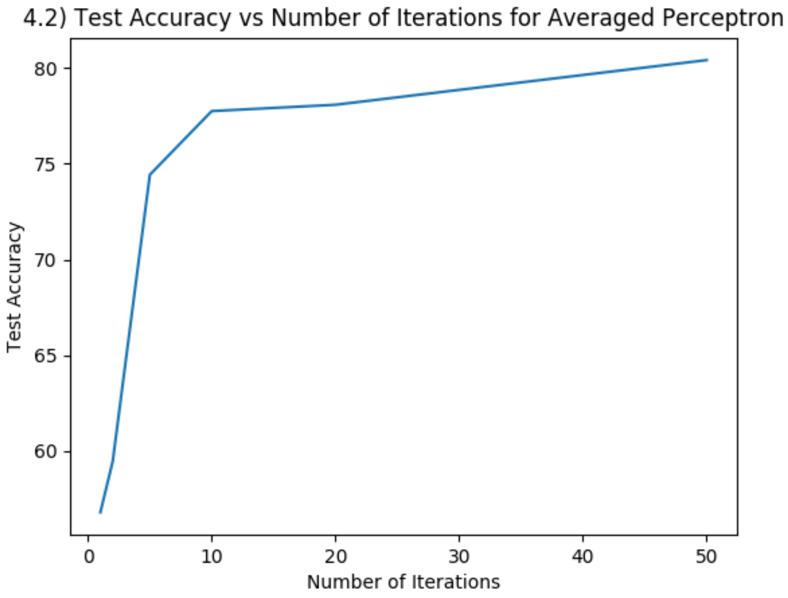
   $$P(c^-) = \frac{\text{size}(c^-)}{\text{size}(c^+) + \text{size}(c^-)}$$

   where size(c^+) and size(c^-) is the total number of the positive and negative class respectively.

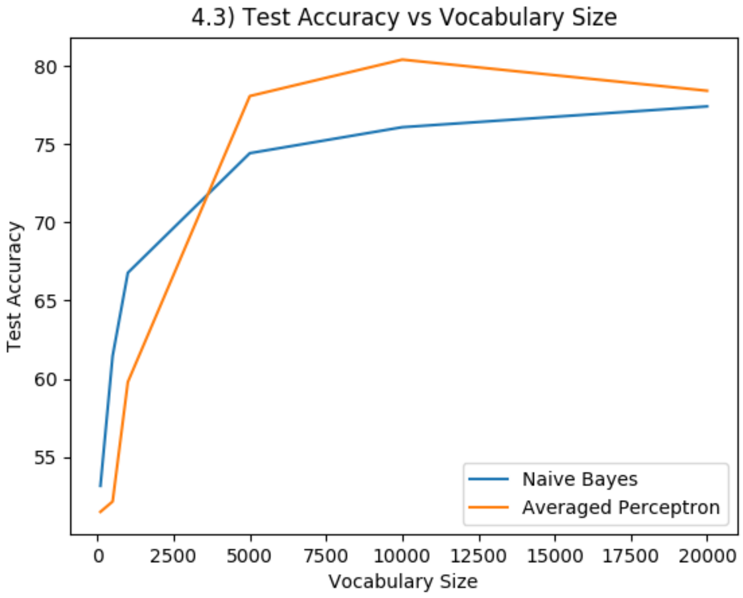
4 Analysis

1. The performance is better with the bias term because bias term allows more classification scenarios to have higher accuracy.

2. The training accuracy does converge after 21 iterations.



3. The performance increases as vocabulary size increases because the model is considering more words now.


4.3) Test Accuracy vs Vocabulary Size
Test Accuracy
Vocabulary Size
Naive Bayes
Averaged Perceptron