CS6375 Assignment 3 report

Name: Haoda LE
Netid: HXL180046

1. Naïve Bayes

NB accuracy summary

Before remove stop words	**After removed stop words**
0.9665271966527197	0.9686192468619247

Analysis
After removed the stop words, the accuracy will increase, because without the noise data, the W vector will be more focused.

Running result shows below:
Before remove stop words

```
(base) Haodas-MBP:Assignment 3 haodale$ python NB.py
spam accuracy: 126 / 130 = 0.9692307692307692
ham accuracy: 336 / 348 = 0.9655172413793104
total accuracy: 0.9665271966527197
```

After removed stop words

```
(base) Haodas-MBP:Assignment 3 haodale$ python NB_filtered_stopwords.py
spam accuracy: 128 / 130 = 0.9846153846153847
ham accuracy: 335 / 348 = 0.9626436781609196
total accuracy: 0.9686192468619247
```

2. Logistic Regression

LR accuracy summary
Iteration times = 10

Lambda	**Before remove stop words**	**After removed stop words**
0.01	0.9142259414225942	0.9246861924686193
0.05	0.9142259414225942	0.9246861924686193
0.1	0.9100418410041841	0.9246861924686193

Iteration times = 50

Lambda	**Before remove stop words**	**After removed stop words**
0.01	0.9309623430962343	0.9435146443514645
0.05	0.9309623430962343	0.9435146443514645
0.1	0.9330543933054394	0.9435146443514645

Analysis
For the whole test, the learning rate is fixed as 0.01.
For same iteration times and lambda, after removed the stop words, the accuracy will increase, because without the noise data, the W vector will be more focused.
For same lambda, the accuracy will increase with the iteration times, because the W vector will be closer to converge.

For the same iteration times and in the appropriate range, change lambda will increase the accuracy, because the lambda will penalize and limit the W vector.

Running result shows below:

Iteration times: 10, Lambda = 0.01, before remove stop words

```
(base) Haodas-MBP:Assignment 3 haodale$ python LR.py 0.01 10
running iteration:  1 2 3 4 5 6 7 8 9 10
spam accuracy: 108 / 130 = 0.8307692307692308
ham accuracy: 329 / 348 = 0.9454022988505747
total accuracy: 0.9142259414225942
```

Iteration times: 10, Lambda = 0.01, after **removed** stop words

```
(base) Haodas-MBP:Assignment 3 haodale$ python LR_filtered_stopwords.py 0.01 10
running iteration:  1 2 3 4 5 6 7 8 9 10
spam accuracy: 103 / 130 = 0.7923076923076923
ham accuracy: 339 / 348 = 0.9741379310344828
total accuracy: 0.9246861924686193
```

Iteration times: 10, Lambda = 0.05, before remove stop words

```
(base) Haodas-MBP:Assignment 3 haodale$ python LR.py 0.05 10
running iteration:  1 2 3 4 5 6 7 8 9 10
spam accuracy: 108 / 130 = 0.8307692307692308
ham accuracy: 329 / 348 = 0.9454022988505747
total accuracy: 0.9142259414225942
```

Iteration times: 10, Lambda = 0.05, after **removed** stop words

```
(base) Haodas-MBP:Assignment 3 haodale$ python LR_filtered_stopwords.py 0.05 10
running iteration:  1 2 3 4 5 6 7 8 9 10
spam accuracy: 103 / 130 = 0.7923076923076923
ham accuracy: 339 / 348 = 0.9741379310344828
total accuracy: 0.9246861924686193
```

Iteration times: 10, Lambda = 0.1, before remove stop words

```
(base) Haodas-MBP:Assignment 3 haodale$ python LR.py 0.1 10
running iteration:  1 2 3 4 5 6 7 8 9 10
spam accuracy: 106 / 130 = 0.8153846153846154
ham accuracy: 329 / 348 = 0.9454022988505747
total accuracy: 0.9100418410041841
```

Iteration times: 10, Lambda = 0.1, after **removed** stop words

```
(base) Haodas-MBP:Assignment 3 haodale$ python LR_filtered_stopwords.py 0.1 10
running iteration:  1 2 3 4 5 6 7 8 9 10
spam accuracy: 103 / 130 = 0.7923076923076923
ham accuracy: 339 / 348 = 0.9741379310344828
total accuracy: 0.9246861924686193
```

Iteration times: 50, Lambda = 0.01, before remove stop words

```
(base) Haodas-MBP:Assignment 3 haodale$ python LR.py 0.01 50
running iteration:  1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
spam accuracy: 109 / 130 = 0.8384615384615385
ham accuracy: 336 / 348 = 0.9655172413793104
total accuracy: 0.9309623430962343
```

Iteration times: 50, Lambda = 0.01, after **removed** stop words

```
(base) Haodas-MBP:Assignment 3 haodale$ python LR_filtered_stopwords.py 0.01 50
running iteration:  1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
spam accuracy: 111 / 130 = 0.8538461538461538
ham accuracy: 340 / 348 = 0.9770114942528736
total accuracy: 0.9435146443514645
```

Iteration times: 50, Lambda = 0.05, before remove stop words

```
(base) Haodas-MBP:Assignment 3 haodale$ python LR.py 0.05 50
running iteration:  1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
spam accuracy: 109 / 130 = 0.8384615384615385
ham accuracy: 336 / 348 = 0.9655172413793104
total accuracy: 0.9309623430962343
```

Iteration times: 50, Lambda = 0.05, after **removed** stop words

```
(base) Haodas-MBP:Assignment 3 haodale$ python LR_filtered_stopwords.py 0.05 50
running iteration:  1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
spam accuracy: 111 / 130 = 0.8538461538461538
ham accuracy: 340 / 348 = 0.9770114942528736
total accuracy: 0.9435146443514645
```

Iteration times: 50, Lambda = 0.1, before remove stop words

```
(base) Haodas-MBP:Assignment 3 haodale$ python LR.py 0.1 50
running iteration:  1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
spam accuracy: 110 / 130 = 0.8461538461538461
ham accuracy: 336 / 348 = 0.9655172413793104
total accuracy: 0.9330543933054394
```

Iteration times: 50, Lambda = 0.1, after **removed** stop words

```
(base) Haodas-MBP:Assignment 3 haodale$ python LR_filtered_stopwords.py 0.1 50
running iteration:  1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
spam accuracy: 111 / 130 = 0.8538461538461538
ham accuracy: 340 / 348 = 0.9770114942528736
total accuracy: 0.9435146443514645
```