CS260 Homework 3

Yang Yang 804522285

Oct 5, 2015

Collaborators: Qian Li

1 Question 1

After converting all training data into bag-of-words features vectors, from the matlab result we know that the most frequently 3 words are:

{(enron: 600), (will: 451), (please: 291)}

2 Question 2

For unregularized logistic regression:

$$w^{t+1} = w^t - \eta \sum_i \{\sigma(w^{t^T} x_i + b^t) - y_i\} x_i \tag{2.1}$$

$$b^{t+1} = b^t - \eta \sum_i \{\sigma(w^{t^T} x_i + b^t) - y_i\} \tag{2.2}$$

For regularized logistic regression:

$$w_j^{t+1} = w_j^t - \eta \{\sum_i \{\sigma(w^{t^T} x_i + b^t) - y_i\} x_{i,j} + 2\lambda w_j^t\} \tag{2.3}$$

$$b^{t+1} = b^t - \eta \sum_i \{\sigma(w^{t^T} x_i + b^t) - y_i\} \tag{2.4}$$

3 Question 3

3.1 a

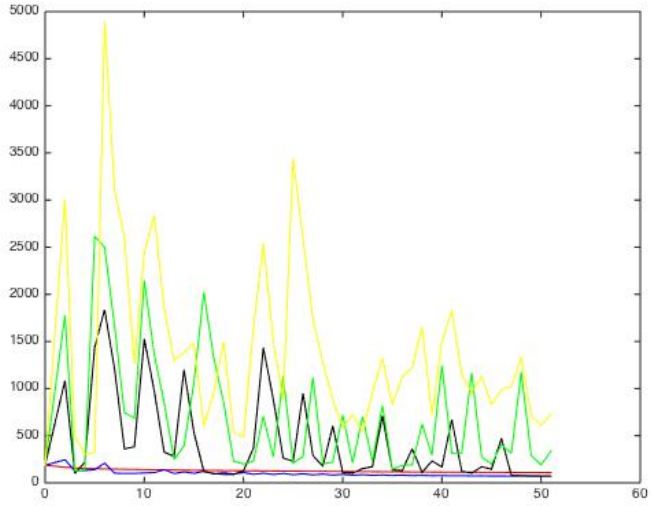

Figure 1 cross-entropy function value with respect to the number of steps for Ionosphere(yellow represent step size = 0.5; green represent step size = 0.1; black represent step size = 0.05; blue represent step size = 0.01; red represent step size = 0.001)

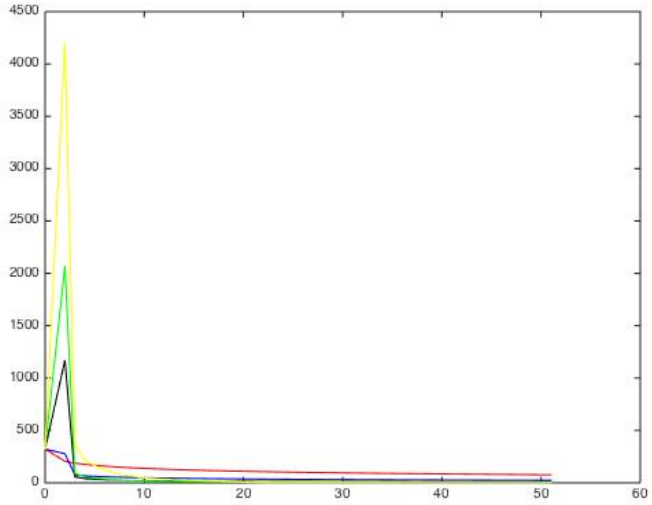

Figure 2 cross-entropy function value with respect to the number of steps for EmailSpam(yellow represent step size = 0.5; green represent step size = 0.1; black represent step size = 0.05; blue represent step size = 0.01; red represent step size = 0.001)

3.2 b

$L2norm(withoutregularization)$	0.001	0.01	0.05	0.1	0.5
Ionosphere	1.4855	4.6573	18.4769	38.5544	185.9563
EmailSpam	2.5806	7.8507	27.4529	53.4542	264.7956

4 Question 4

4.1 Part a

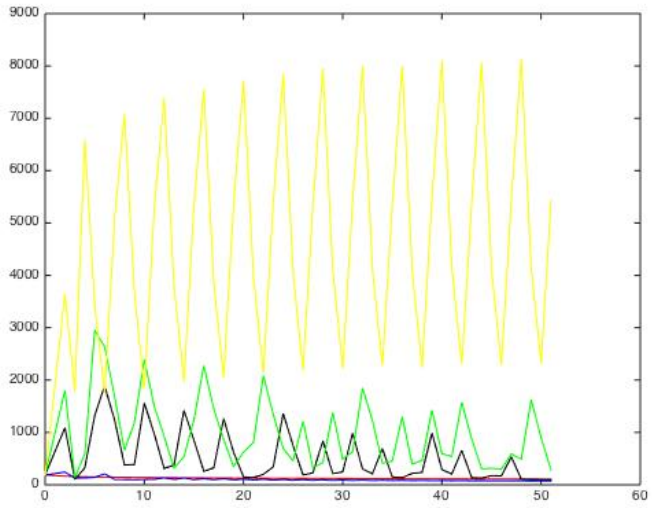

Figure 3 cross-entropy function with regularization value with respect to the number of steps for Ionosphere(yellow represent step size = 0.5; green represent step size = 0.1; black represent step size = 0.05; blue represent step size = 0.01; red represent step size = 0.001)

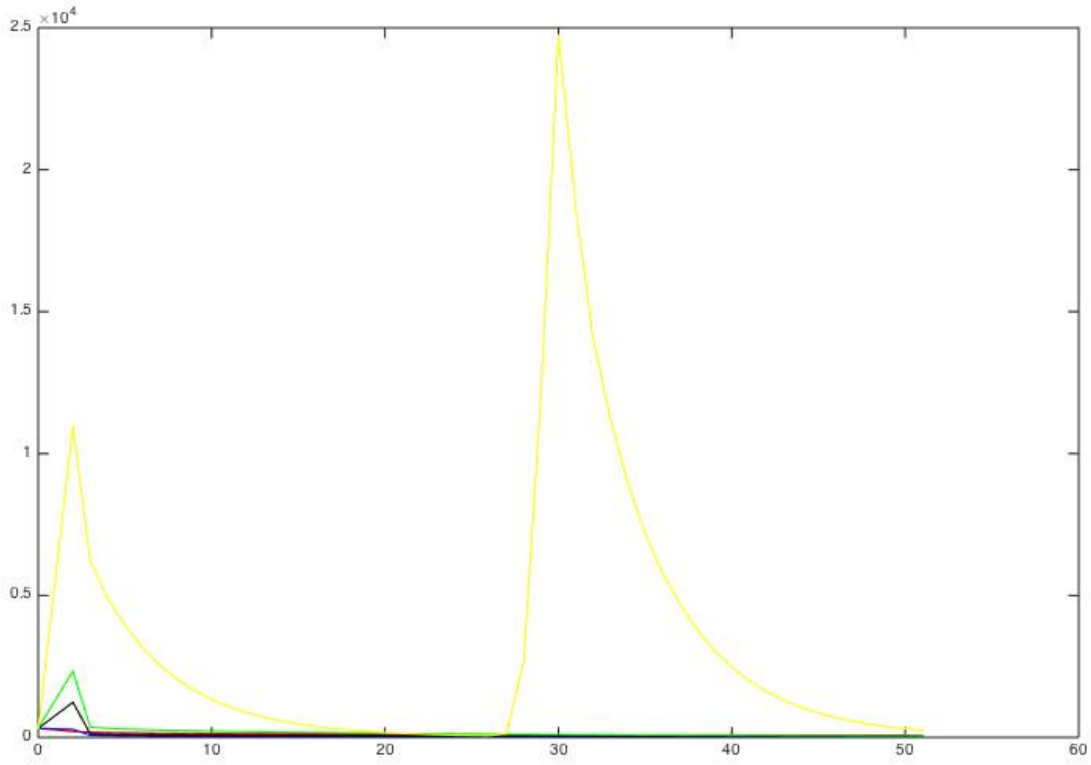

Figure 4 cross-entropy function with regularization value with respect to the number of steps for EmailSpam(yellow represent step size = 0.5; green represent step size = 0.1; black represent step size = 0.05; blue represent step size = 0.01; red represent step size = 0.001)

4.2 Part b

$L2norm(withregularization)$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35
Ionosphere	4.6573	4.5766	4.4990	4.4242	4.3521	4.2829	4.2164	4.1526
EmailSpam	7.8507	7.6080	7.3784	7.1611	6.9557	6.7614	6.5777	6.4041
$L2norm(withregularization)$	0.4	0.45	0.5					
Ionosphere	4.0908	4.0300	3.9712					
EmailSpam	6.2400	6.0848	5.9382					

4.3 Part c

Entropy function with respect to regularization cofficient for Ionosphere are shown below:

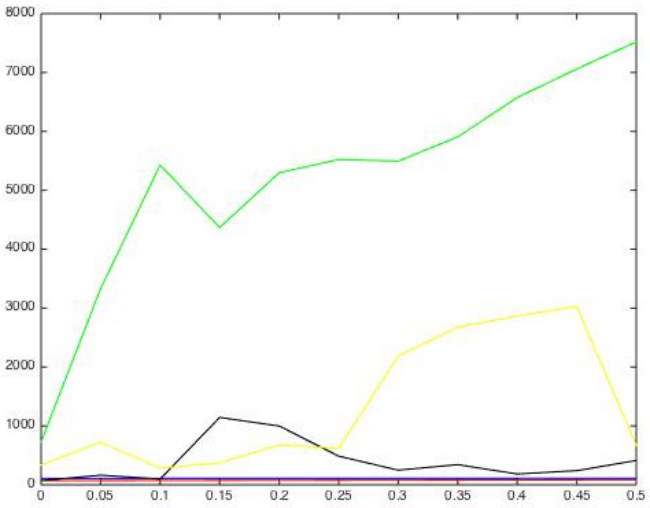

Figure 5 entropy function with respect to regularization cofficient for Ionosphere(blue line: represent step size = 0.001, red line represent step size= 0.01, black line represent step size = 0.05, yellow line represent step size= 0.1, green line represent step size= 0.5)

Entropy function with respect to regularization cofficient for EmailSpam are shown below:

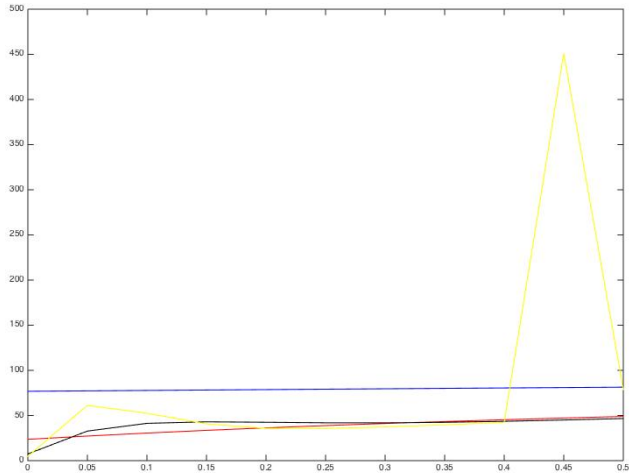

Figure 6 entropy function with respect to regularization cofficient for EmailSpam(blue line: represent step size = 0.001, red line represent step size= 0.01, black line represent step size = 0.05, yellow line represent step size= 0.1)

In the Figure 5, I did not include the situation when step size equals to 0.5. This is because the error function in this situation will be too large. You can see this situation in the below Figure 6.

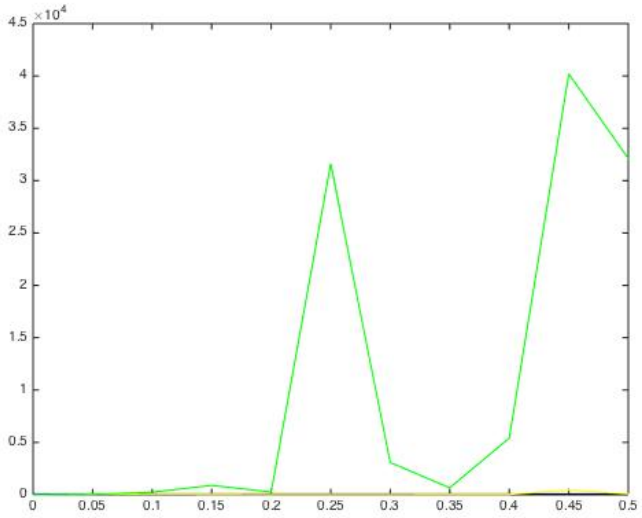

Figure 7 entropy function with respect to regularization cofficient for EmailSpam(blue line: represent step size = 0.001, red line represent step size= 0.01, black line represent step size = 0.05, yellow line represent step size= 0.1, green line represent step size= 0.5)

5 Question 5

For unregularized logistic regression:

$$\bigtriangledown \varepsilon_t(w) = \sum_i \{\sigma(w^T x_i + b) - y_i\} x_i \tag{5.1}$$

$$\mathbb{H}(w) = \sum_i \{x_i x_i^T \sigma(\omega^T x_i + b)[1 - \sigma(\omega^T x_i + b)]\} \tag{5.2}$$

$$\bigtriangledown \varepsilon_t(b) = \sum_i \{\sigma(w^T x_i + b) - y_i\} \tag{5.3}$$

$$\mathbb{H}(b) = \sum_i \{\sigma(\omega^T x_i + b)[1 - \sigma(\omega^T x_i + b)]\} \tag{5.4}$$

Thus, we can get the update function as follows

$$\begin{aligned} w^{t+1} &= w^t - \mathbb{H}^{-1} \bigtriangledown \varepsilon_t \\ &= w^t - \{\sum_i \{x_i x_i^T \sigma(w^{tT} x_i + b^t)[1 - \sigma(w^{tT} x_i + b^t)]\}\}^{-1} \{\sum_i \{\sigma(w^{tT} x_i + b^t) - y_i\} x_i\} \end{aligned} \tag{5.5}$$

$$\begin{aligned} b^{t+1} &= b^t - \mathbb{H}^{-1} \bigtriangledown \varepsilon_t \\ &= b^t - \{\sum_i \{\sigma(w^{tT} x_i + b)[1 - \sigma(w^{tT} x_i + b)]\}\}^{-1} \{\sum_i \{\sigma(w^{tT} x_i + b^t) - y_i\}\} \end{aligned} \tag{5.6}$$

For regularized logistic regression:

$$\frac{\partial}{\partial w_j}(\varepsilon_t(w_j)) = \sum_i \{\sigma(w^T x_i + b) - y_i\} x_{i,j} + 2\lambda w_j \tag{5.7}$$

$$\frac{\partial^2}{\partial w w^T}(\varepsilon_t(w_j)) = \sum_i \{x_i x_i^T \sigma(\omega^T x_i + b)[1 - \sigma(\omega^T x_i + b)]\} + 2\lambda E \tag{5.8}$$

$$\bigtriangledown \varepsilon_t(b) = \sum_i \{\sigma(w^T x_i + b) - y_i\} \tag{5.9}$$

$$\mathbb{H}(b) = \sum_i \{\sigma(\omega^T x_i + b)[1 - \sigma(\omega^T x_i + b)]\} \tag{5.10}$$

Thus, we can get the update function as follows

$$\begin{aligned} w_j^{t+1} &= w_j^t - \mathbb{H}^{-1} \bigtriangledown \varepsilon_t \\ &= w_j^t - \{\sum_i \{x_{i,j} x_{i,j}^T \sigma(w^{tT} x_i + b^t)[1 - \sigma(w^{tT} x_i + b^t)]\} + 2\lambda E\}^{-1} \{\sum_i \{\sigma(w^{tT} x_i + b^t) - y_i\} x_{i,j} + 2\lambda w_j^t\} \end{aligned} \tag{5.11}$$

$$\begin{aligned} b^{t+1} &= b^t - \mathbb{H}^{-1} \bigtriangledown \varepsilon_t \\ &= b^t - \{\sum_i \{\sigma(\omega^T x_i + b)[1 - \sigma(\omega^T x_i + b)]\}\}^{-1} \{\sum_i \{\sigma(w^T x_i + b) - y_i\}\} \end{aligned} \tag{5.12}$$

6 Question 6

6.1 Part a

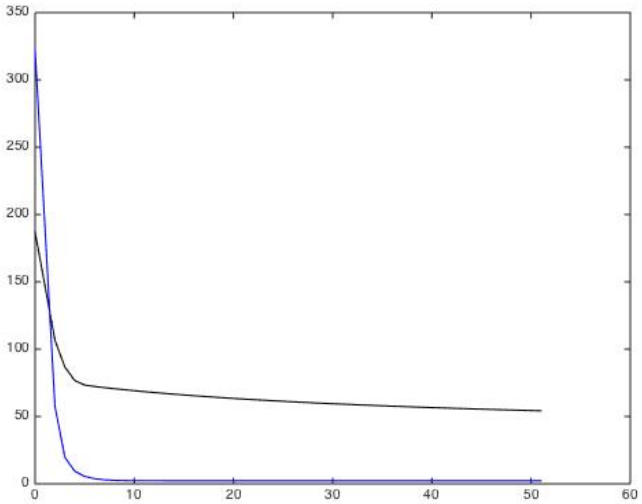

Figure 8 cross-entropy function with regularization value with respect to the number of steps for Ionosphere(black represent ionosphere and blue represent EmailSpam)

6.2 b

For ionosphere, the L2 norm for w is 11.6866
For EmailSpam, the L2 norm for w is 603.3280

6.3 c

For ionosphere, the error function value is 40.0021
For EmailSpam, the error function value is 3697.6

7 Question 7

7.1 a

For $\lambda = 0$, it's the same as question 6.
For $\lambda = 0.05$, it's the same as question 6.

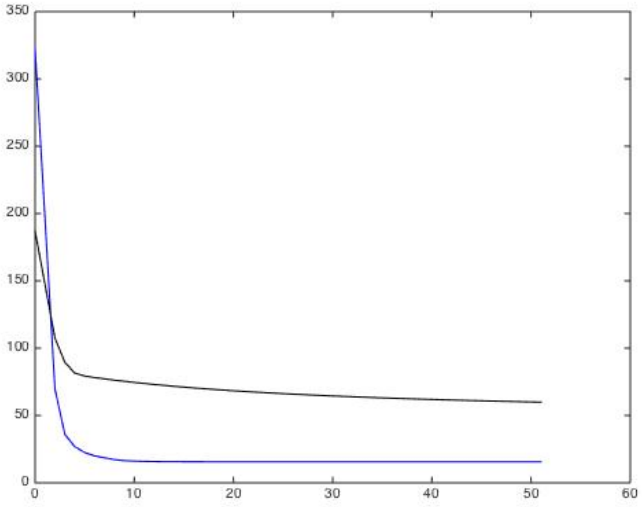

Figure 9 When $\lambda = 0.05$, cross-entropy function with regularization value with respect to the number of steps (black represent ionosphere and blue represent EmailSpam)

For $\lambda = 0.1$, it's the same as question 6.

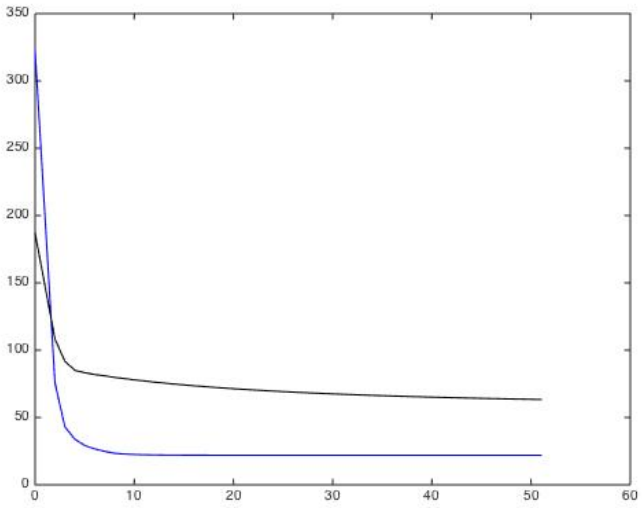

Figure 10 When $\lambda = 0.1$, cross-entropy function with regularization value with respect to the number of steps (black represent ionosphere and blue represent EmailSpam)

For $\lambda = 0.15$, it's the same as question 6.

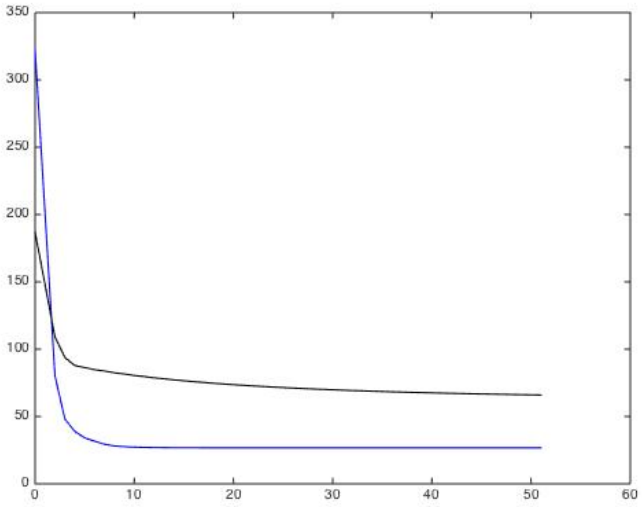

Figure 11 When $\lambda = 0.15$, cross-entropy function with regularization value with respect to the number of steps (black represent ionosphere and blue represent EmailSpam)

For $\lambda = 0.2$, it's the same as question 6.

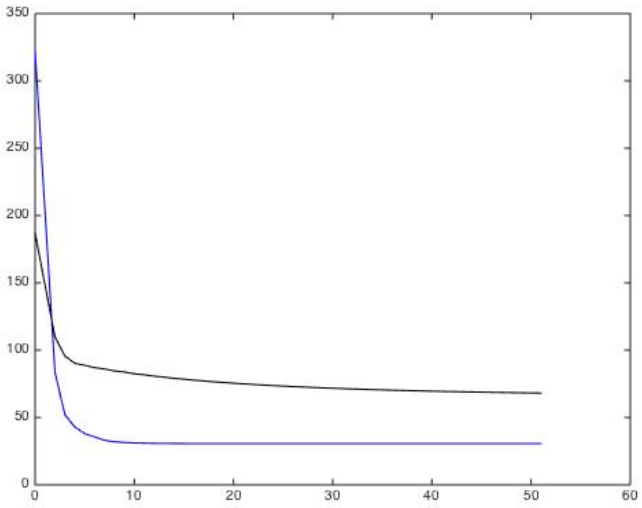

Figure 12 When $\lambda = 0.2$, cross-entropy function with regularization value with respect to the number of steps (black represent ionosphere and blue represent EmailSpam)

For $\lambda = 0.25$, it's the same as question 6.

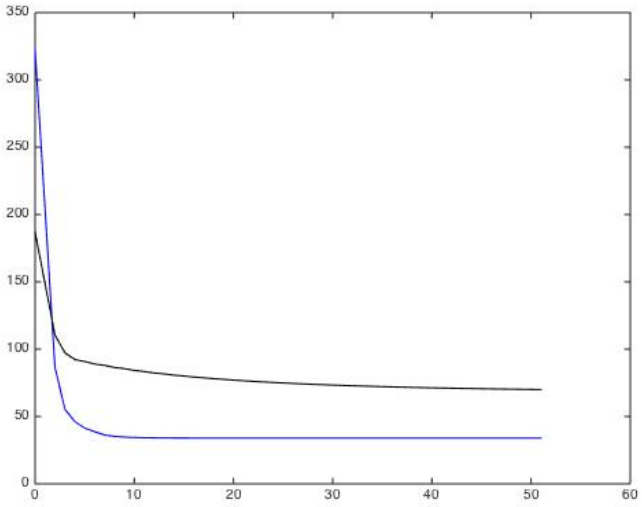

Figure 13 When $\lambda = 0.25$, cross-entropy function with regularization value with respect to the number of steps (black represent ionosphere and blue represent EmailSpam)

For $\lambda = 0.3$, it's the same as question 6.

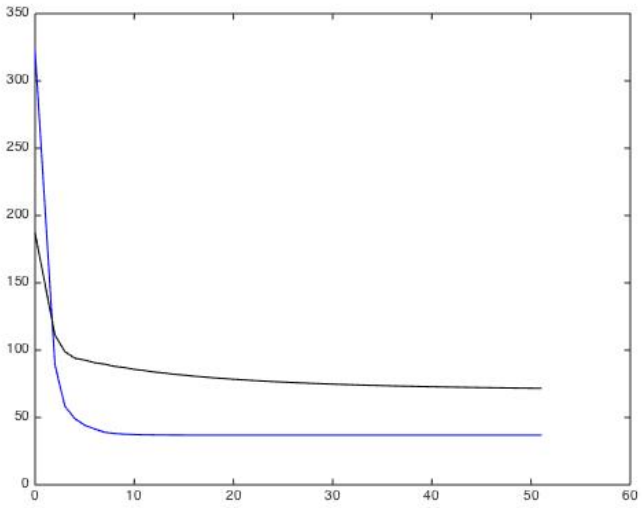

Figure 14 When $\lambda = 0.3$, cross-entropy function with regularization value with respect to the number of steps (black represent ionosphere and blue represent EmailSpam)

For $\lambda = 0.35$, it's the same as question 6.

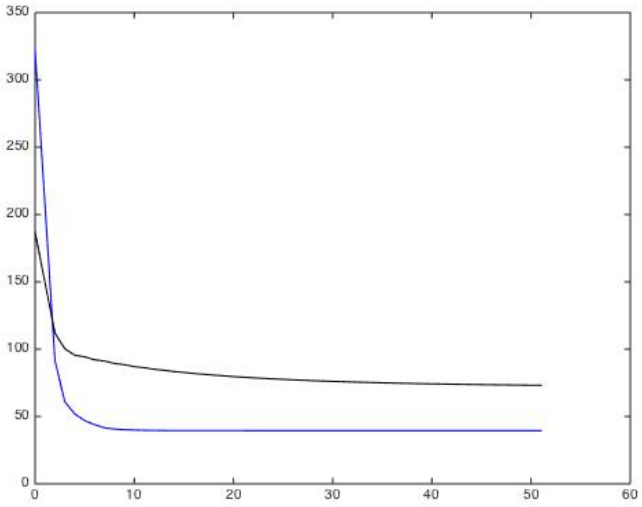

Figure 15 When $\lambda = 0.35$, cross-entropy function with regularization value with respect to the number of steps (black represent ionosphere and blue represent EmailSpam)

For $\lambda = 0.4$, it's the same as question 6.

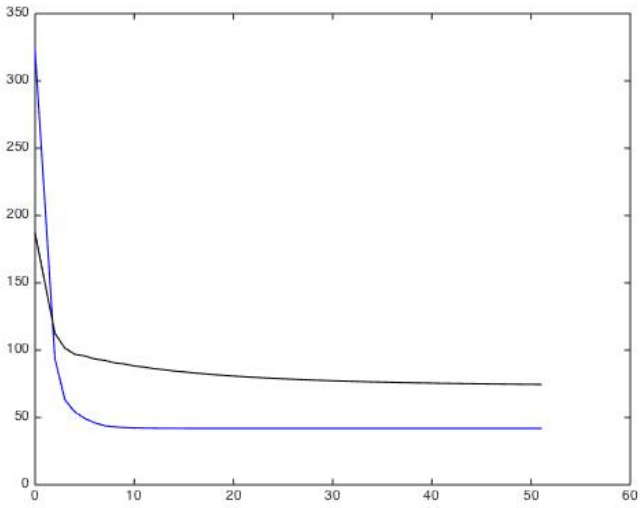

Figure 16 When $\lambda = 0.4$, cross-entropy function with regularization value with respect to the number of steps (black represent ionosphere and blue represent EmailSpam)

For $\lambda = 0.45$, it's the same as question 6.

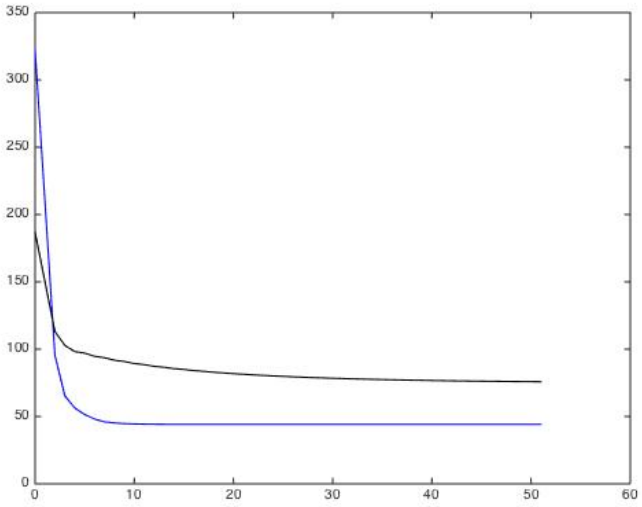

Figure 17 When $\lambda = 0.45$, cross-entropy function with regularization value with respect to the number of steps (black represent ionosphere and blue represent EmailSpam)

For $\lambda = 0.5$, it's the same as question 6.

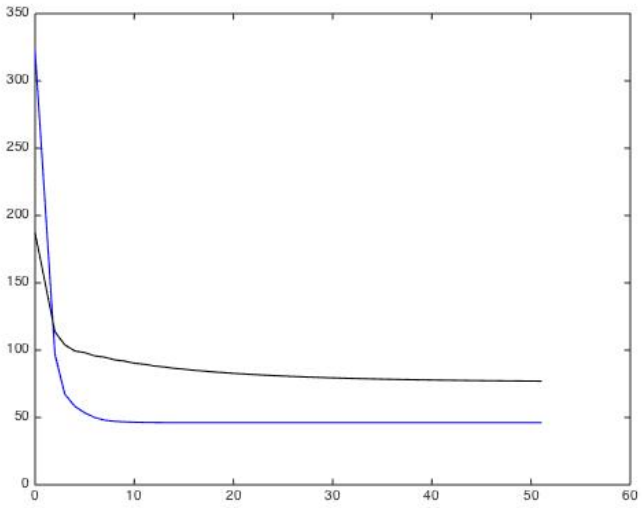

Figure 18 When $\lambda = 0.5$, cross-entropy function with regularization value with respect to the number of steps (black represent ionosphere and blue represent EmailSpam)

For $\lambda = 0.5$, it's the same as question 6.

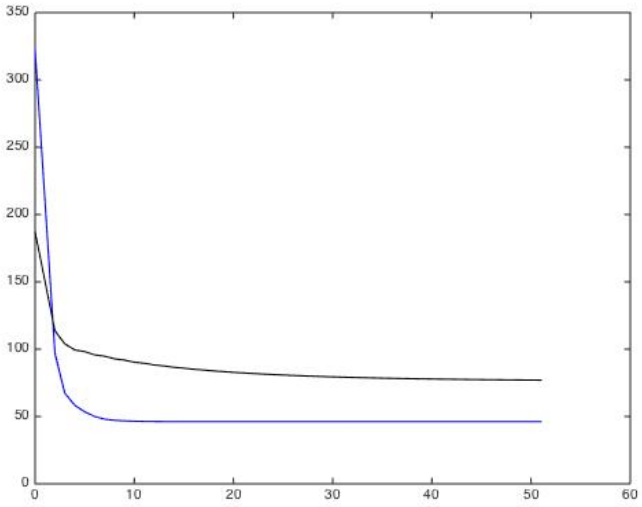

Figure 19 When $\lambda = 0.5$, cross-entropy function with regularization value with respect to the number of steps(black represent ionosphere and blue represent EmailSpam)

To compare the result, I put all Ionosphere plots in one picture and all EmailSpam plots in one picture. The pictures are shown below:

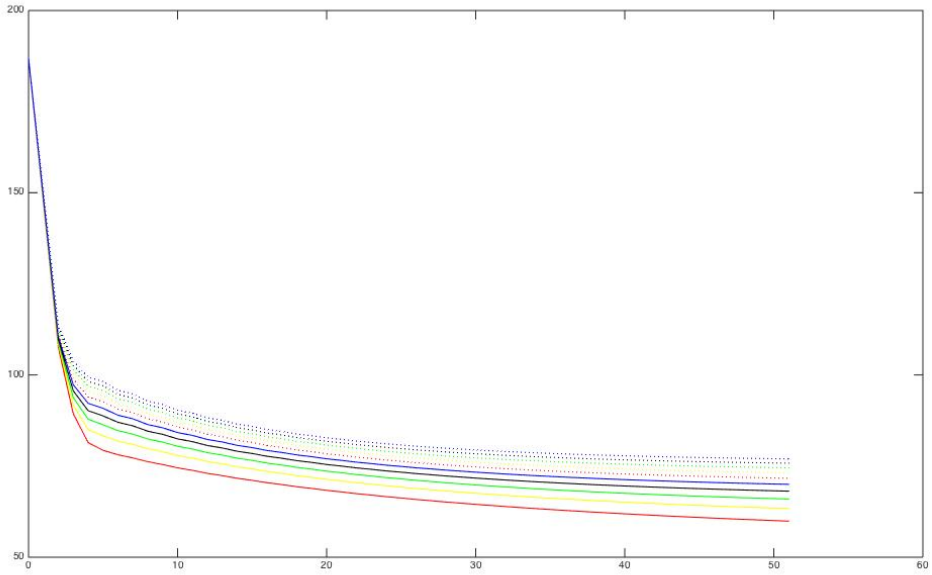

Figure 20 cross-entropy function value with respect to the number of steps for Ionosphere(red solidline: $\lambda = 0.05$, yellow solidline: $\lambda = 0.1$, green solidline:$\lambda = 0.15$, black solidline: $\lambda = 0.2$, blue solidline: $\lambda = 0.25$,red dashed line: $\lambda = 0.3$, yellow dashed line: $\lambda = 0.35$, green dashed line:$\lambda = 0.4$, black dashed line: $\lambda = 0.45$, blue dashed line: $\lambda = 0.5$,)

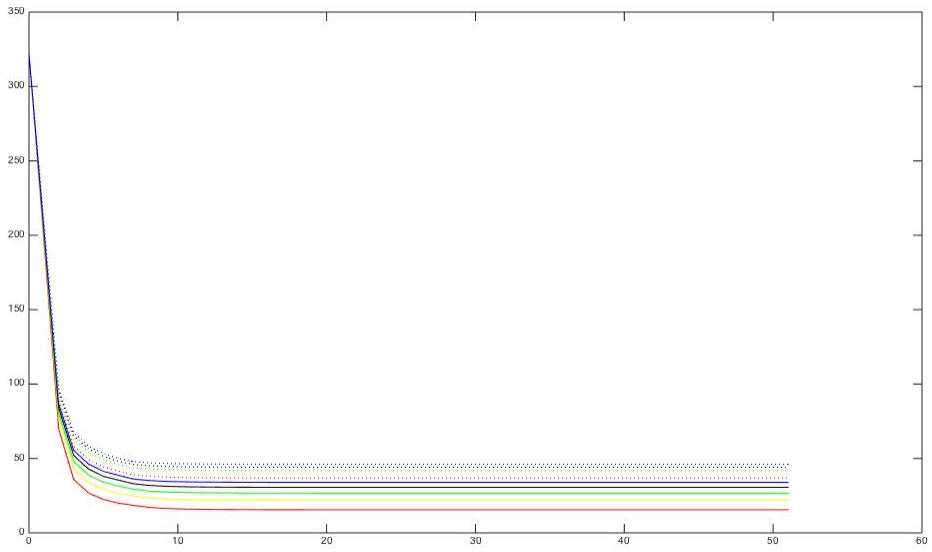

Figure 21 cross-entropy function value with respect to the number of steps for EmailSpam(red solidline: $\lambda = 0.05$, yellow solidline: $\lambda = 0.1$, green solidline:$\lambda = 0.15$, black solidline: $\lambda = 0.2$, blue solidline: $\lambda = 0.25$,red dashed line: $\lambda = 0.3$, yellow dashed line: $\lambda = 0.35$, green dashed line:$\lambda = 0.4$, black dashed line: $\lambda = 0.45$, blue dashed line: $\lambda = 0.5$,)

7.2 Part b

$L2norm(withregularization)$	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35
Ionosphere	11.68	8.57	7.29	6.54	6.03	5.65	5.36	5.11
EmailSpam	603.3	12.46	10.36	9.22	8.4543	7.88	7.4395	7.07
$L2norm(withregularization)$	0.4	0.45	0.5					
Ionosphere	4.91	4.74	4.58					
EmailSpam	6.76	6.50	6.27					

7.3 Part c

error Function	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35
Ionosphere	40.0	32.80	31.10	30.97	31.37	31.9247	32.5178	33.10
EmailSpam	3697.6	122.35	116.12	113.92	113.0654	112.8121	112.88	113.15
error Function	0.4	0.45	0.5					
Ionosphere	33.67	34.21	34.72					
EmailSpam	113.54	114.0	114.51					

8 Question 8

From Figure 1 to Figure 4, we can see that if step size is too small, the convergence will be slow. But if step size is too large, the result will be unstable. It may now show convergence.

From the table in 4.2, we can see that magnitude of w will be smaller with the increase of λ.

If the step size is too large, the error function will increase steeply with the increase of step size. And it will increase very slowly with the increase of λ

9 Question 9

When using gradient descent, you should choose the step size very carefully. But using newton's method could not consider this problem and it can convergence at the second iteration. However, the calculation of newton's method is more complex than gradient descent. It will take much longer time to compute newton's method than gradient descent.