

**The Experiment Report of**

**Deep *Learning***

**College Software College**

**Subject Software Engineering**

**Members**  Tengyun Wang

**Student ID 201710106574**

**E-mail wty9391@gmail.com**

**Tutor**   **Mingkui Tang**

**Date submitted** **2017.12 .07**

1. **Topic: Linear Regression, Linear Classification and Gradient Descent**
2. **Time: 2017.12.02**

**3. Reporter: Tengyun Wang**

**4. Purposes:**

* 1. Further understand of linear regression and gradient descent.
  2. Conduct some experiments under small scale dataset.
  3. Realize the process of optimization and adjusting parameters.

**5. Data sets and data analysis:**

Linear Regression uses Housing in LIBSVM Data, including 506 samples and each sample has 13 features. You are expected to download scaled edition. After downloading, you are supposed to divide it into training set, validation set.

Linear classification uses Australian in LIBSVM Data, including 690 samples and each sample has 14 features. You are expected to download scaled edition. After downloading, you are supposed to divide it into training set, validation set.

**6. Experimental steps:**

* **Linear Regression and Gradient Descent**

1. Load the experiment data. You can use load\_svmlight\_file function in sklearn library.
2. Devide dataset. You should divide dataset into training set and validation set using train\_test\_split function. Test set is not required in this experiment.
3. Initialize linear model parameters. You can choose to set all parameter into zero, initialize it randomly or with normal distribution.
4. Choose loss function and derivation: Find more detail in PPT.
5. Calculate gradient toward loss function from all samples.
6. Denote the opposite direction of gradient as .
7. Update model: . is learning rate, a hyper-parameter that we can adjust.
8. Get the loss under the training set and by validating under validation set.
9. Repeate step 5 to 8 for several times, and drawing graph of as well as with the number of iterations.

* **Linear Classification and Gradient Descent**

1. Load the experiment data.
2. Divide dataset into training set and validation set.
3. Initialize SVM model parameters. You can choose to set all parameter into zero, initialize it randomly or with normal distribution.
4. Choose loss function and derivation: Find more detail in PPT.
5. Calculate gradient toward loss function from all samples.
6. Denote the opposite direction of gradient as .
7. Update model: . is learning rate, a hyper-parameter that we can adjust.
8. Select the appropriate threshold, mark the sample whose predict scores greater than the threshold as positive, on the contrary as negative. Get the loss under the trainin set and by validating under validation set.
9. Repeate step 5 to 8 for several times, and drawing graph of as well as with the number of iterations.

**7. Code:**

* **Linear Regression and Gradient Descent**

1. # write your code here#!/usr/bin/env python3
2. # -\*- coding: utf-8 -\*-
3. """
4. Created on Sat Dec  2 09:58:38 2017
5. @author: wty
6. """
7. **import** numpy as np
8. **import** scipy
9. **import** matplotlib.pyplot as plt
11. **from** sklearn.datasets **import** load\_svmlight\_file
12. **from** sklearn.model\_selection **import** train\_test\_split
14. **def** h(w,X):
15. **return** X.dot(w)
17. **def** error(w,X,Y):
18. **return** h(w,X) - Y
20. **def** L(w,X,Y,lamda = 0.0):
21. num\_records,num\_features  = np.shape(X)
22. e = error(w,X,Y)
23. regulation\_loss = 1.0/2 \* lamda \* w.transpose().dot(w)
24. loss = 1.0/2 \* 1.0/float(num\_records) \* e.transpose().dot(e)\
25. + regulation\_loss
26. **return** loss[0][0]
28. **def** g(w,X,Y,lamda = 0.0):
29. num\_records,num\_features  = np.shape(X)
31. # L2 norm
32. **return** 1.0/num\_records \* X.transpose().dot(error(w,X,Y)) \
33. + lamda \* w
35. **def** MSE(y,y\_hat):
36. **return** ((y-y\_hat)\*\*2).sum()/len(y)
38. data = load\_svmlight\_file("./resources/housing\_scale.txt")
39. # add interception
40. X = scipy.sparse.hstack(\
41. (scipy.sparse.csr\_matrix(np.ones((len(data[1]),1))),data[0]))
42. Y = data[1].reshape((len(data[1]),1))
44. num\_records,num\_features  = np.shape(X)
46. X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(\
47. X, Y, test\_size=0.33,random\_state=5)
48. Y\_train = Y\_train.reshape((len(Y\_train),1))
49. Y\_test = Y\_test.reshape((len(Y\_test),1))
51. # initialize w
52. w = np.random.normal(size=(num\_features,1))
54. lamda = 0.5
55. eta = 0.05
56. loss\_train = []
57. loss\_test = []
58. MSE\_train = []
59. MSE\_test = []
60. max\_iterate = 30
61. **for** epoch **in** range(max\_iterate):
62. #print("epoch:",epoch)
63. loss\_train.append(L(w,X\_train,Y\_train,lamda))
64. loss\_test.append(L(w,X\_test,Y\_test,lamda))
65. MSE\_train.append(MSE(Y\_train,h(w,X\_train)))
66. MSE\_test.append(MSE(Y\_test,h(w,X\_test)))
67. w = w - eta \* g(w,X\_train,Y\_train,lamda)
69. plt.subplot(211)
70. train\_loss\_line = plt.plot(range(max\_iterate),loss\_train,label='train loss')
71. test\_loss\_line = plt.plot(range(max\_iterate),loss\_test,label='test loss')
72. ax=plt.gca()
73. ax.set(xlabel='Epoch', ylabel='Loss with l2 norm')
74. plt.legend()
76. plt.subplot(212)
77. train\_mse\_line = plt.plot(range(max\_iterate),MSE\_train,label='train MSE')
78. test\_mse\_line = plt.plot(range(max\_iterate),MSE\_test,label='test MSE')
79. ax=plt.gca()
80. ax.set(xlabel='Epoch', ylabel='MSE')
81. plt.legend()
83. plt.show()

* **Linear Classification and Gradient Descent**

1. # -\*- coding: utf-8 -\*-
3. **import** numpy as np
4. **import** scipy
5. **import** matplotlib.pyplot as plt
7. **from** sklearn.datasets **import** load\_svmlight\_file
8. **from** sklearn.model\_selection **import** train\_test\_split
10. **def** h(w,X):
11. **return** X.dot(w)
13. **def** hinge\_loss(w,X,Y,C=1.0):
14. num\_records,num\_features  = np.shape(X)
15. zero = np.zeros((num\_records,1))
16. margin = 1 - C \* Y \* h(w,X)
17. **return** np.max([zero,margin],axis=0)
19. **def** L(w,X,Y,lamda=0.0,C=1.0):
20. num\_records,num\_features  = np.shape(X)
21. e = hinge\_loss(w,X,Y,C)
22. regulation\_loss = 1.0/2 \* lamda \* w.transpose().dot(w)
23. loss = 1.0/float(num\_records) \* e.sum()  + regulation\_loss
24. **return** loss[0][0]
26. **def** g(w,X,Y,lamda=0.0,C=1.0):
27. num\_records,num\_features  = np.shape(X)
28. e = hinge\_loss(w,X,Y,C)
29. indicator = np.zeros((num\_records,1))
30. indicator[np.nonzero(e)] = 1
32. **return** - 1.0/float(num\_records) \* C \
33. \* X.transpose().dot(Y \* indicator).sum(axis=1).reshape((num\_features,1)) \
34. + lamda \* w
36. **def** predict(w,X,threshold=0.5):
37. raw = h(w,X)
38. raw[raw<=threshold] = -1
39. raw[raw>threshold] = 1
40. **return** raw
42. **def** accuracy(w,X,Y,threshold=0.5):
43. num\_records,num\_features  = np.shape(X)
44. P = predict(w,X,threshold)
46. is\_right = P \* Y
47. is\_right[is\_right < 0] = 0
49. **return** 1.0/num\_records \* np.count\_nonzero(is\_right)
51. data = load\_svmlight\_file("./resources/australian\_scale.txt")
52. # add interception
53. X = scipy.sparse.hstack(\
54. (scipy.sparse.csr\_matrix(np.ones((len(data[1]),1))),data[0]))
55. Y = data[1].reshape((len(data[1]),1))
57. num\_records,num\_features  = np.shape(X)
59. X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(\
60. X, Y, test\_size=0.33,random\_state=42)
61. Y\_train = Y\_train.reshape((len(Y\_train),1))
62. Y\_test = Y\_test.reshape((len(Y\_test),1))
64. # initialize w
65. w = np.random.normal(size=(num\_features,1))
67. lamda = 0.1
68. eta = 0.2
69. C = 1.0
70. threshold=0.7
71. max\_iterate = 100
72. loss\_train = []
73. loss\_test = []
74. accuracy\_train = []
75. accuracy\_test = []
77. **for** epoch **in** range(max\_iterate):
78. #print("epoch:",epoch)
79. loss\_train.append(L(w,X\_train,Y\_train,lamda,C))
80. loss\_test.append(L(w,X\_test,Y\_test,lamda,C))
81. accuracy\_train.append(accuracy(w,X\_train,Y\_train,threshold))
82. accuracy\_test.append(accuracy(w,X\_test,Y\_test,threshold))
83. w = w - eta \* g(w,X\_train,Y\_train,lamda,C)
85. fig, ax = plt.subplots()
86. ax\_e = ax.twinx()
87. train\_loss\_line = ax.plot(range(max\_iterate),loss\_train,label='train loss')
88. test\_loss\_line = ax.plot(range(max\_iterate),loss\_test,label='test loss')
89. train\_accuracy\_line = ax\_e.plot(range(max\_iterate),accuracy\_train,'r',label='train accuracy')
90. test\_accuracy\_line = ax\_e.plot(range(max\_iterate),accuracy\_test,'g',label='test accuracy')
92. ax.set(xlabel='Epoch', ylabel='Loss with l2 norm')
93. ax\_e.set\_ylabel('Accuracy with threshold='+str(threshold))
95. ax.legend(loc=4)
96. ax\_e.legend(loc=1)
97. plt.show()

(Fill in the contents of 8-12 respectively for linear regression and linear classification)

**8. Selection of validation (hold-out, cross-validation, k-folds cross-validation, etc.):**

* **Linear Regression and Gradient Descent**

hold-out，Train Set: Test Set = 3 : 1, Exhaustive Grid Search

param\_grid = [

{'lamda': [0, 0.01, 0.1, 0.5, 1], 'eta': [0.05, 0.1, 0.15, 0.2, 0.3]}

]

The score is the opposite number of **RMSE**



* **Linear Classification and Gradient Descent**

hold-out，Train Set: Test Set = 3 : 1, Exhaustive Grid Search

param\_grid = [

{'lamda': [0, 0.1, 0.5],

'eta': [0.01, 0.02, 0.05, 0.1],

'C': [1, 2],

'threshold': [0.4,0.5,0.6]}

]

The score is the **accuracy** of classification model



**9. The initialization method of model parameters:**

* **Linear Regression and Gradient Descent**

**w = np.random.normal(size=(num\_features,1))**

* **Linear Classification and Gradient Descent**

**w = np.random.normal(size=(num\_features,1))**

**10. The selected loss function and its derivatives:**

* **Linear Regression and Gradient Descent**
* **Linear Classification and Gradient Descent**

**11. Experimental results and curve:**

* **Linear Regression and Gradient Descent**

Printing the best 5 models loss curves

Figure of {'eta': 0.05, 'lamda': 0}

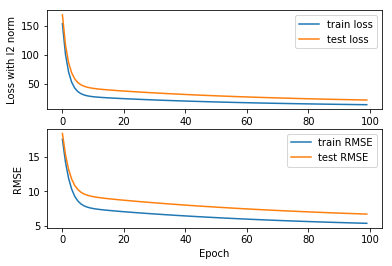


Figure of {'eta': 0.05, 'lamda': 0.01}

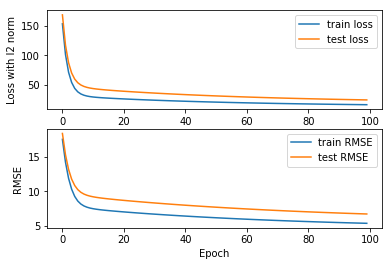


Figure of {'eta': 0.05, 'lamda': 0.1}

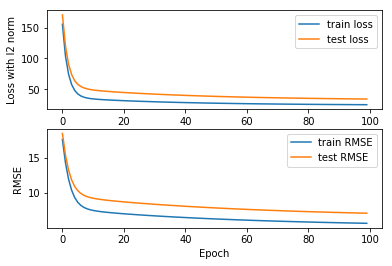


Figure of {'eta': 0.05, 'lamda': 0.5}

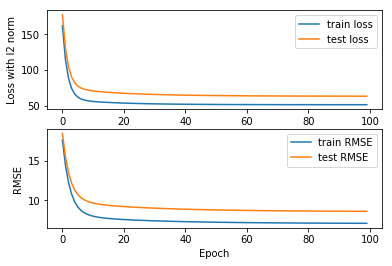
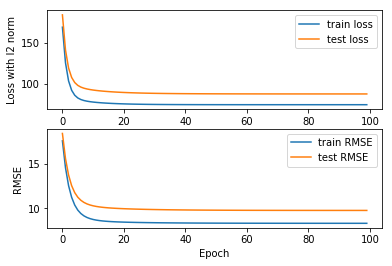


Figure of {'eta': 0.05, 'lamda': 1}



* **Linear Classification and Gradient Descent**

Printing the best 5 models loss curves

Figure of {'C': 1, 'eta': 0.05, 'lamda': 0.1, 'threshold': 0.6}

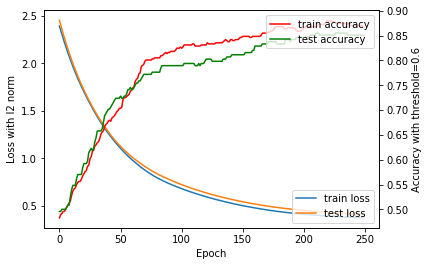


Figure of {'C': 1, 'eta': 0.1, 'lamda': 0.5, 'threshold': 0.4}

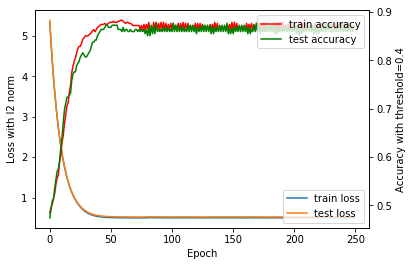


Figure of {'C': 1, 'eta': 0.1, 'lamda': 0.1, 'threshold': 0.6}

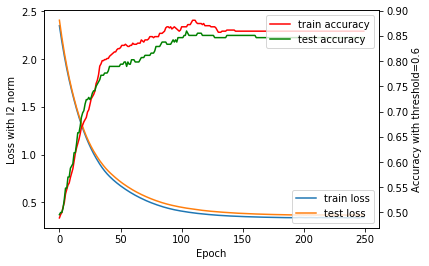


Figure of {'C': 1, 'eta': 0.05, 'lamda': 0.5, 'threshold': 0.4}

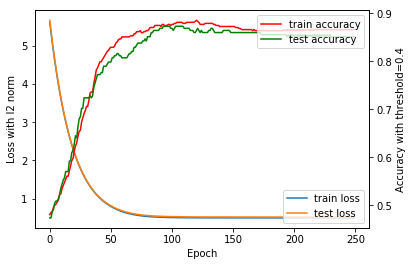
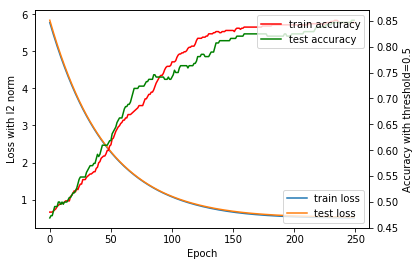


Figure of {'C': 1, 'eta': 0.02, 'lamda': 0.5, 'threshold': 0.5}



* **12. Results analysis:**

Gradient decent is a valid method to optimize both regression problem and classification problem.

At the beginning of optimization, the loss decrease quickly and the accuracy of model increase sharply. With epoch goes by, the model verges to be optimized.

The greater the learning rate is, the quicker the loss decreases. But large learning rate may lead to vibrating.

The metrics to evaluate models are important. In genera, a smaller loss leads to better metrics. But there are exceptions, see the relationship between accuracy and loss in classification experiments.

**13. Similarities and differences between linear regression and linear classification:**

In linear regression, the evaluation metric is often MSE(mean square error).

In linear classification, the evaluation metric is often accuracy. And the threshold is a important factor to influence the accuracy.

**14. Summary:**

Gradient decent is a valid method to optimize both regression problem and classification problem.

There are many differences between linear regression and linear classification.