

Student ID: R05944004

Name: Cheng-Wei Wu

# hw 2

## Spam Classification

### Report

---

#### • Program Overview

Different from the hw1, I modularized several parts of my program this time. All the functions are show as follows:

1. sigmoid
2. cross\_entropy
3. read\_data
4. logistic\_regression
5. gen\_model
6. read\_model
7. read\_test
8. gen\_ans

#### • sigmoid (Fig. 4)

It's a normal sigmoid function (support numpy array operation).

#### • cross\_entropy (Fig. 5)

It will compute the cross-entropy for each weight. Since the predicted value may very close to either 0 or 1, which causing operation error to the numpy log function, I add a small value ( $1e-3$ ) to the predicted value which is close to either 0 or 1 to avoid such runtime error. (Fig 2)

#### • read\_data

It's a normal function to read the training data into the format of numpy array. (the data format is same with the hw1 → row: observation, col: variables)

#### • logistic\_regression (Fig. 7)

It's similar with the linear regression version. The only differences are list below:

1. It replaces the results of inner product of weights and data with the same results but under the sigmoid function manipulation.
2. It changes the coefficients of computed gradient value of both weights and bias
3. It changes the way to calculate loss function (root-mean-square error to cross-entropy)

Beyond these difference, I thought that the logistic one is almost the same with linear one.

- **gen\_model**

I used this function to write back the weights and bias to the model file via json format.

- **read\_model**

It's a basic and trivial function which load the generated model.

- **read\_test**

It's almost the same with the function "read\_data" which read the testing data.

- **gen\_ans (Fig. 8)**

Like previous functions, this function is similar with the part of the codes in hw1. It calculates the predicted value of testing data and fills the corresponding answer to the array according to their probability.

Finally, I ran the different functions according to the length of the system parameters and generated the "prediction.csv".

- **Regularization**

Same with the hw1, due to the features of the data, there still exists these 2 phenomena:

1. I proposed my function set to be multivariate linear equation, which has lower probability overfitting the training set than quadratic functions
2. The increment of the error rate of the public set may indicate that my model won't overfit the training data or the public set (which means it will be more adaptable to other data)

- **Learning Rate**

Different from the hw1, in this homework, only the adagrad and adamgrad can converge in specific epochs. The results of using adagrad is similar with using adamgrad.

- **Discussion**

In this assignment, I have implemented 3 different methods:

1. Master → Logistic Regression (#iter: 8000, Adagrad, learning rate: 1 → Accuracy: 0.94333)
2. Method2 → Probabilistic Generative Model (Accuracy: 0.94333) → Fig. 6
3. Method3 → Linear Regression (#iter: 8000, Adagrad, learning rate: 1 → Accuracy: 0.93667)

I had tried some different combinations of the parameters such like #iteration, learning rate, feature extraction, etc. The best one's accuracy in the public leaderboard on the Kaggle is 0.95000, but I forgot the setting of the parameters... The other two methods I implemented are probabilistic generative and linear regression. Following the course lecture, I thought that the accuracy may not have great difference since the size of dataset is not large enough. I also did several experiments on different parameters, such as #iteration, lambda value, different extraction method of features, etc. All these experiments are show in several line charts on the next page. After doing these experiments, I was attracted by the different feature extraction methods' effect; in other words, the difference between feature scale and combinations could result in different accuracy. Since this difference is bigger than I expected, I decided to try other possible combinations and plotted in the line chart (I only try 3 different methods at first). The reason why I came up with this idea is that I noticed that the last several features represent the rare characters which may not appear in the normal email content.

## Figures

### Different Epochs on Optimize Methods

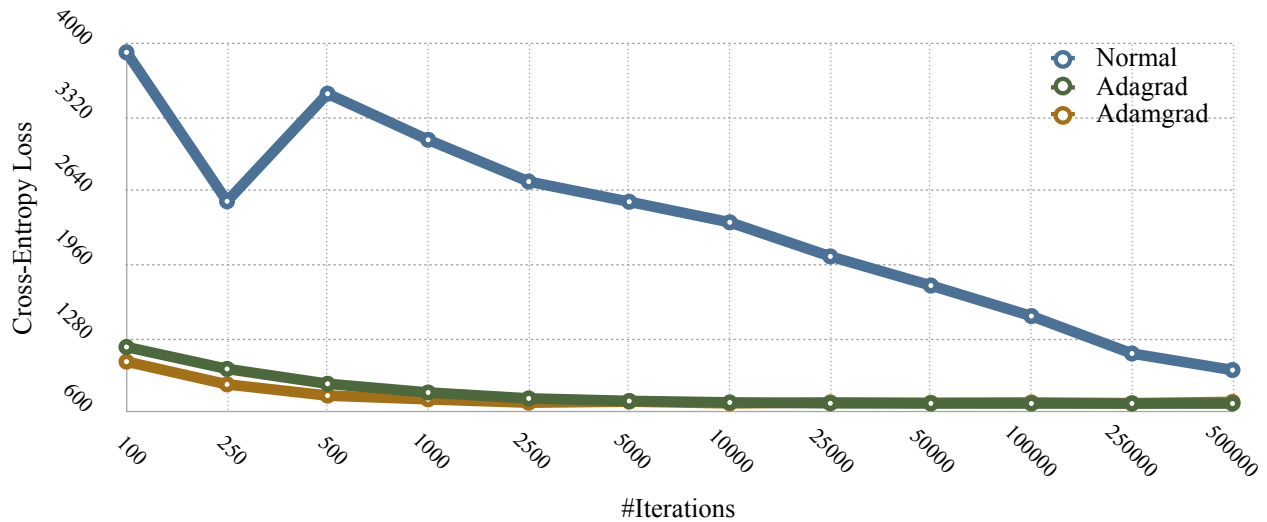


Figure 1.

Experiment results of different epochs on different optimize methods

### Different Feature Extractions (Adamgrad, learning rate = 0.001)

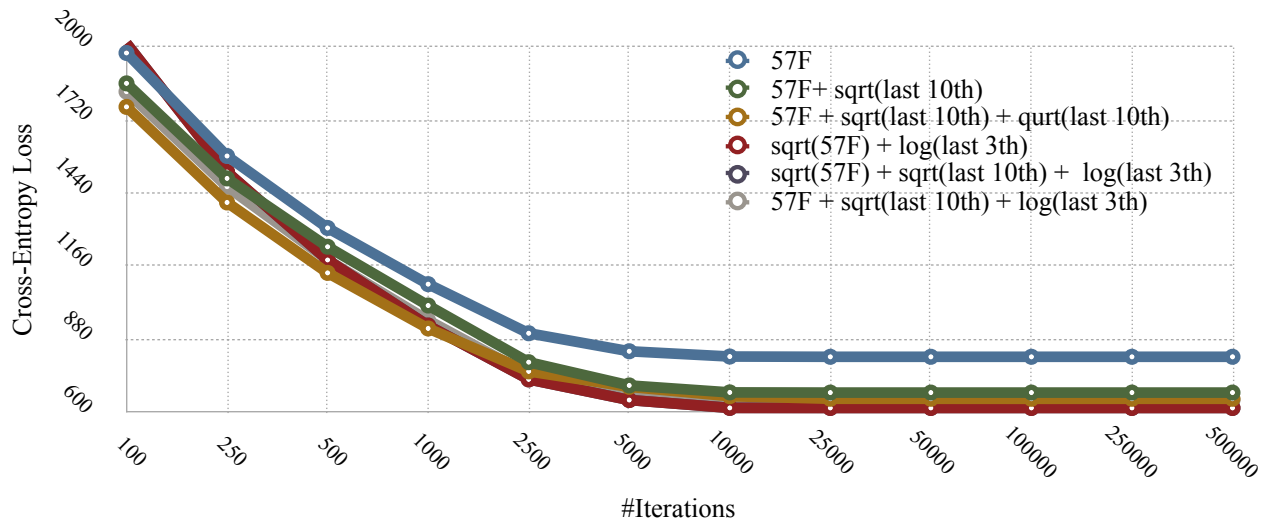


Figure 2.

Experiment results of different feature extraction method (F = feature, Adamgrad, learning rate = 0.001)

### Effect of Regularization (#iter = 10000)

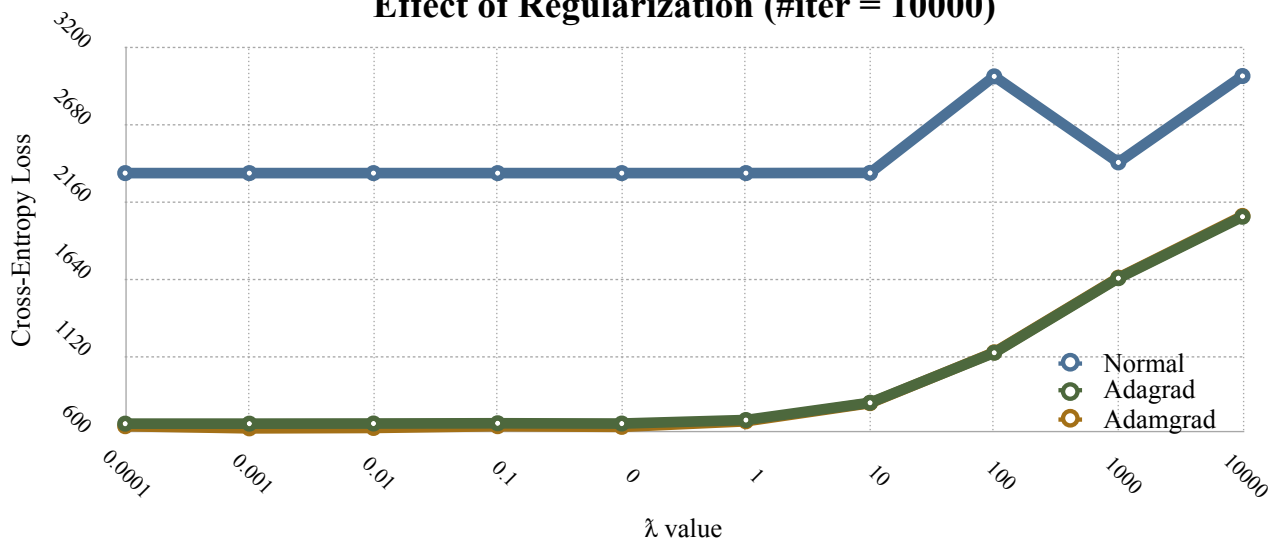


Figure 3.

Experiment results of the effect of regularization (#iter = 10000)

- Codes

```

4  def sigmoid(z):
5      return 1 / (1 + np.exp(-z))

```

Figure. 4

Sigmoid function implemented by numpy

```

7  def cross_entropy(fx_n, y_n):
8      ans = []
9      for i in range(fx_n.shape[0]):
10         if np.isclose(fx_n[i], 0.):
11             ans.append(y_n[i] * np.log(fx_n[i]+1e-3) + (1-y_n[i]) * np.log(1-fx_n[i]))
12         elif np.isclose(1-fx_n[i], 0.):
13             ans.append(y_n[i] * np.log(fx_n[i]) + (1-y_n[i]) * np.log(1-fx_n[i]+1e-3))
14         else:
15             ans.append(y_n[i] * np.log(fx_n[i]) + (1-y_n[i]) * np.log(1-fx_n[i]))
16     return -np.array(ans)

```

Figure. 5

Compute the cross-entropy for each weight (they will be summed as the value of loss function)

```

30 def logistic_regression(X_TRAIN, Y_TRAIN):
31     W, b = np.zeros(X_TRAIN.shape[1]), 0
32     SUM_SQDW, SUM_SQDB = np.zeros(X_TRAIN.shape[1])+1, 0
33     norm, adag, adam = 0.0000001, 0.1, 0.01 # adam-default = 0.001
34     beta1, beta2 = 0.9, 0.999
35     Wmt, Wvt = 0, 0
36     Bmt, Bvt = 0, 0
37     epoch, Lambda, t, eps = 10000, 0, 0, 1e-8
38     for i in range(epoch):
39         fwb = sigmoid(np.dot(X_TRAIN, W) + b)
40         ERR = Y_TRAIN - fwb
41         DW = -1 * np.dot(X_TRAIN.T, ERR)
42         DB = -1 * np.sum(ERR)
43
44         # Compute Loss & Print
45         # if i % 500 == 0:
46             Loss = np.sum(cross_entropy(fwb, Y_TRAIN))
47             print "Iter %7s | Loss: %.7f" % (i, Loss)
48
49         # Regularization
50         DW += Lambda * 2 * W
51
52         # Normal
53         # W -= norm * DW # / X_TRAIN.shape[0]
54         # b -= norm * DB # / X_TRAIN.shape[0]
55
56         # Adagrad
57         # SUM_SQDW += np.square(DW)
58         # SUM_SQDB += np.square(DB)
59         # W -= adag / np.sqrt(SUM_SQDW) * DW # / X_TRAIN.shape[0]
60         # b -= adag / np.sqrt(SUM_SQDB) * DB # / X_TRAIN.shape[0]
61
62         # Adamgrad
63         t += 1
64         Wmt = beta1 * Wmt + (1-beta1) * DW
65         Wvt = beta2 * Wvt + (1-beta2) * np.square(DW)
66         Wmthat = Wmt / (1-np.power(beta1, t))
67         Wvthat = Wvt / (1-np.power(beta2, t))
68         Bmt = beta1 * Bmt + (1-beta1) * DB
69         Bvt = beta2 * Bvt + (1-beta2) * np.square(DB)
70         Bmthat = Bmt / (1-np.power(beta1, t))
71         Bvthat = Bvt / (1-np.power(beta2, t))
72         W -= (adam*Wmthat) / (np.sqrt(Wvthat) + eps)
73         b -= (adam*Bmthat) / (np.sqrt(Bvthat) + eps)
74     return W, b

```

```

24 def gen_mean(X, Y):
25     X_0, X_1 = [], []
26     for i in range(Y.shape[0]):
27         X_0.append(X[i]) if Y[i] == 0 else X_1.append(X[i])
28     X_0, X_1 = np.array(X_0).T, np.array(X_1).T
29     u_0, u_1 = [], []
30     col = X.shape[1]
31     for i in range(col):
32         u_0.append(np.mean(X_0[i][:]))
33         u_1.append(np.mean(X_1[i][:]))
34     return np.array(u_0), np.array(u_1), X_0.shape[1], X_1.shape[1]
35
36
37 def gen_cov(X):
38     return np.cov(X, rowvar=False)
39
40
41 def gen_model(modelpath, u_0, u_1, cov, N_0, N_1):
42     inv_cov = np.linalg.inv(cov)
43     W = np.dot((u_0 - u_1).T, inv_cov).T
44     b = -0.5 * np.dot(np.dot(u_0.T, inv_cov), u_0) + \
45         0.5 * np.dot(np.dot(u_1.T, inv_cov), u_1) + \
46         np.log(N_0 / N_1)
47     with open(modelpath, 'wb') as file:
48         json.dump({'b': b, 'W': list(W)}, file)

```

Figure. 6

Probabilistic Generative Model

Figure. 7

Logistic Regression (normal, adagrad, adamgrad)

```

94 def gen_ans(anspath, X_TEST, W, b):
95     fwb_TEST = sigmoid(np.dot(X_TEST, W) + b)
96     Y_TEST = []
97     for i in fwb_TEST:
98         Y_TEST.append(0) if np.less_equal(i, 0.5) else Y_TEST.append(1)
99     with open(anspath, 'wb') as file:
100         writer = csv.writer(file, delimiter=',')
101         writer.writerow(['id', 'label'])
102         for i in range(len(Y_TEST)): writer.writerow((i+1, Y_TEST[i]))

```

Figure. 8

Generate the predicted results according to the probability (logistic)