

**The Experiment Report of**

***Machine Learning***

**College Software College**

**Subject Software Engineering**

**Members**

**Student ID 201530661741**

**E-mail yimingzhao0@qq.com**

**Tutor**   **Mingkui Tan**

**Date submitted** **2017. .**

**1. Topic: Logistic regression, logistic classification and gradient decent method**

**2. Time: 2017.12.2 PM 2:00-5:00**

**3. Reporter: Yiming Zhao**

**4. Purposes:**

* **Understand the differences between the gradient decent and stochastic gradient decent method.**
* **Understand the connection between the logistic regression and linear regression.**
* **Practicing the SVM on the bigger data set.**

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

**Using a9a data set from LIBSVM Data, including** **32561 / 16281(testing) samples and 123 features in each sample.**

**6. Experimental steps:**

* **Read the data set using sklearn’s function**
* **Randomly initializing the parameters.**
* **Selecting a proper loss function and using NAG, Adam, Adadelta, RMSProp to optimize the loss function.**

**7. Code:**

from sklearn import datasets as ds  
import numpy as np  
from numpy import random  
import matplotlib.pyplot as plt  
  
  
def sigmoid(input\_):  
 return 1 / (1 + np.exp(-input\_))  
  
  
def train(x\_train, y\_train, x\_test, y\_test, method, iters, test\_errors):  
 max\_iterations = 100  
 theta = random.rand(num\_features + 1)  
 num\_test\_samples, num\_test\_features = x\_test.shape  
  
 if method == 'sgd':  
 lr = 0.01  
  
 for i in range(max\_iterations):  
 output = sigmoid(np.dot(x\_train[i], theta))  
 error = output - y\_train[i]  
 theta = theta - lr \* np.dot(x\_train[i], error)  
  
 predict\_error = 0  
 for j in range(num\_test\_samples):  
 predict\_output = sigmoid(np.dot(x\_test[j], theta))  
 predict\_error -= y\_test[j] \* np.log(predict\_output) + (1 - y\_test[j]) \* np.log(1 - predict\_output)  
 print(str(i) + '\t' + str(predict\_error / num\_test\_samples))  
  
 iters.append(i)  
 test\_errors.append(predict\_error / num\_test\_samples)  
  
 if method == 'nag':  
 lr = 0.01  
 miu = 0.9  
 momentum = np.zeros(num\_features + 1)  
  
 for i in range(max\_iterations):  
 output = sigmoid(np.dot(x\_train[i], theta - lr \* miu \* momentum))  
 error = output - y\_train[i]  
 grad = np.dot(x\_train[i], error)  
 momentum = momentum \* lr + grad  
 theta = theta - lr \* momentum  
  
 predict\_error = 0  
 for j in range(num\_test\_samples):  
 predict\_output = sigmoid(np.dot(x\_test[j], theta))  
 predict\_error -= y\_test[j] \* np.log(predict\_output) + (1 - y\_test[j]) \* np.log(1 - predict\_output)  
 print(str(i) + '\t' + str(predict\_error / num\_test\_samples))  
  
 iters.append(i)  
 test\_errors.append(predict\_error / num\_test\_samples)  
  
 if method == 'rmsprop':  
 lr = 0.1  
 expectation = 1  
 rho = 0.95  
 delta = 10e-7  
  
 for i in range(max\_iterations):  
 output = sigmoid(np.dot(x\_train[i], theta))  
 error = output - y\_train[i]  
 grad = np.dot(x\_train[i], error)  
 norm = grad \* grad  
 expectation = rho \* expectation + (1 - rho) \* norm  
 theta = theta - lr \* grad / (np.sqrt(expectation) + delta)  
  
 predict\_error = 0  
 for j in range(num\_test\_samples):  
 predict\_output = sigmoid(np.dot(x\_test[j], theta))  
 predict\_error -= y\_test[j] \* np.log(predict\_output) + (1 - y\_test[j]) \* np.log(1 - predict\_output)  
 print(str(i) + '\t' + str(predict\_error / num\_test\_samples))  
 iters.append(i)  
 test\_errors.append(predict\_error / num\_test\_samples)  
  
 if method == 'adam':  
 delta = 10e-8  
 rho1 = 0.9  
 rho2 = 0.999  
 lr = 0.1  
 s = 0  
 r = 0  
  
 for i in range(max\_iterations):  
 output = sigmoid(np.dot(x\_train[i], theta))  
 error = output - y\_train[i]  
 grad = np.dot(x\_train[i], error)  
  
 s = rho1 \* s + (1 - rho1) \* grad  
 r = rho2 \* r + (1 - rho2) \* grad \* grad  
 s\_hat = s / (1 - rho1)  
 r\_hat = r / (1 - rho2)  
 delta\_theta = (-lr \* s\_hat) / (np.sqrt(r\_hat) + delta)  
 theta = theta + delta\_theta  
  
 predict\_error = 0  
 for j in range(num\_test\_samples):  
 predict\_output = sigmoid(np.dot(x\_test[j], theta))  
 predict\_error -= y\_test[j] \* np.log(predict\_output) + (1 - y\_test[j]) \* np.log(1 - predict\_output)  
 print(str(i) + '\t' + str(predict\_error / num\_test\_samples))  
 iters.append(i)  
 test\_errors.append(predict\_error / num\_test\_samples)  
  
 if method == 'adadelta':  
 r = 0  
 e = 0  
 miu = 0.9  
 delta = 10e-7  
 lr = 10  
  
 for i in range(max\_iterations):  
 output = sigmoid(np.dot(x\_train[i], theta))  
 error = output - y\_train[i]  
 grad = np.dot(x\_train[i], error)  
  
 r = miu \* r + (1 - miu) \* grad \* grad  
 delta\_theta = (-lr \* grad \* np.sqrt(e + delta)) / (np.sqrt(r + delta))  
 theta = theta + delta\_theta  
 e = miu \* e + (1 - miu) \* e \* e  
  
  
 predict\_error = 0  
 for j in range(num\_test\_samples):  
 predict\_output = sigmoid(np.dot(x\_test[j], theta))  
 predict\_error -= y\_test[j] \* np.log(predict\_output) + (1 - y\_test[j]) \* np.log(1 - predict\_output)  
 print(str(i) + '\t' + str(predict\_error / num\_test\_samples))  
 iters.append(i)  
 test\_errors.append(predict\_error / num\_test\_samples)  
  
if \_\_name\_\_ == '\_\_main\_\_':  
 x\_train, y\_train = ds.load\_svmlight\_file('./data/a9a')  
 x\_test, y\_test = ds.load\_svmlight\_file('./data/a9a.t')  
  
 num\_samples, num\_features = x\_train.shape  
 num\_test\_samples, num\_test\_features = x\_test.shape  
  
 x\_train = x\_train.toarray()  
 temp = np.ones(shape=[32561, 1], dtype=np.float32)  
 x\_train = np.concatenate([x\_train, temp], axis=1)  
 x\_test = x\_test.toarray()  
 temp = np.zeros(shape=[16281, 1], dtype=np.float32)  
 temp1 = np.ones(shape=[16281, 1], dtype=np.float32)  
 x\_test = np.concatenate([x\_test, temp, temp1], axis=1)  
  
 for i in range(0, len(y\_train)):  
 if y\_train[i] == -1:  
 y\_train[i] = 0  
 for i in range(0, len(y\_test)):  
 if y\_test[i] == -1:  
 y\_test[i] = 0  
  
  
 methods = ['sgd', 'nag', 'rmsprop', 'adadelta', 'adam']  
 for method in methods:  
 iters = []  
 test\_errors = []  
 train(x\_train, y\_train, x\_test, y\_test, method, iters, test\_errors)  
 plt.plot(iters, test\_errors, label=method)  
  
plt.xlabel('Iteration')  
plt.ylabel('Loss')  
plt.legend()  
plt.show()

from sklearn import datasets as ds  
import numpy as np  
from numpy import random  
import matplotlib.pyplot as plt  
import os  
  
C = 1  
feature\_size = 123  
bias = np.zeros(shape=[feature\_size + 1, 1])  
bias[len(bias)-1][0] = 1.  
  
  
def compute\_loss(x, y, w):  
 # x.shape = [batch\_size, feature\_size + 1], y.shape = [batch\_size, 1], w.shape = [feature\_size + 1, 1]  
 pred = np.matmul(x, w)  
 hinge\_loss = np.maximum(1 - y \* pred, 0)  
 loss = np.mean(hinge\_loss \*\* 2) + C \* np.sum((w - bias) \*\* 2)  
 return loss  
  
  
def compute\_gradient(x, y, w):  
 # x.shape = [batch\_size, feature\_size + 1], y.shape = [batch\_size, 1], w.shape = [feature\_size + 1, 1]  
 pred = np.matmul(x, w)  
 hinge\_loss = np.maximum(1 - y \* pred, 0)  
 hinge\_loss\_gradient = -np.matmul(x.transpose(), hinge\_loss \* y) / len(y)  
 norm\_gradient = 2 \* C \* (w - bias)  
 gradient = hinge\_loss\_gradient + norm\_gradient  
 return gradient # gradient.shape = [feature\_size + 1, 1], the same as w  
  
  
global\_list = {'NAG\_momentum': 0, 'RMS\_expectation': 1, 'ADAM\_s': 0, 'ADAM\_r': 0,  
 'ADADELDA\_r': 0, 'ADADELDA\_e': 0}  
  
  
def optimizer(method, parameter\_list, x, y, w):  
 global global\_list  
 if method == 'SGD':  
 # sgd\_para\_list = {'learning\_rate':0.01}  
 lr = parameter\_list['learning\_rate']  
 w -= lr \* compute\_gradient(x, y, w)  
 if method == 'NAG':  
 # nag\_para\_list = {'miu':0.9, 'learning\_rate':0.01}  
 lr = parameter\_list['learning\_rate']  
 miu = parameter\_list['miu']  
 momentum = global\_list['NAG\_momentum']  
 gradient = compute\_gradient(x, y, w - momentum \* lr \* miu)  
 momentum = momentum \* lr + gradient  
 w -= lr \* momentum  
 global\_list['NAG\_momentum'] = momentum  
 if method == 'RMSProp':  
 # rmsprop\_para\_list = {'delta':10e-7, 'rho':0.95, 'learning\_rate':0.1}  
 lr = parameter\_list['learning\_rate']  
 expectation = global\_list['RMS\_expectation']  
 rho = parameter\_list['rho']  
 delta = parameter\_list['delta']  
 gradient = compute\_gradient(x, y, w)  
 expectation = rho \* expectation + (1 - rho) \* (gradient \*\* 2)  
 global\_list['RMS\_expectation'] = expectation  
 w -= lr \* gradient / (np.sqrt(expectation) + delta)  
 if method == 'ADAM':  
 # adam\_para\_list = {'delta':10e-8, 'rho1':0.9, 'rho2':0.999, 'learning\_rate':0.1}  
 delta = parameter\_list['delta']  
 rho1 = parameter\_list['rho1']  
 rho2 = parameter\_list['rho2']  
 lr = parameter\_list['learning\_rate']  
 s = global\_list['ADAM\_s']  
 r = global\_list['ADAM\_r']  
 gradient = compute\_gradient(x, y, w)  
 s = rho1 \* s + (1 - rho1) \* gradient  
 r = rho2 \* r + (1 - rho2) \* (gradient \*\* 2)  
 s\_hat = s / (1 - rho1)  
 r\_hat = r / (1 - rho2)  
 w -= (lr \* s\_hat) / (np.sqrt(r\_hat) + delta)  
 global\_list['ADAM\_s'] = s  
 global\_list['ADAM\_r'] = r  
 if method == 'Adadelta':  
 # adadelta\_para\_list = {'delta':10e-7, 'miu':0.9, 'learning\_rate':0.1}  
 r = global\_list['ADADELDA\_r']  
 e = global\_list['ADADELDA\_e']  
 miu = parameter\_list['miu']  
 delta = parameter\_list['delta']  
 lr = parameter\_list['learning\_rate']  
 grad = compute\_gradient(x, y, w)  
 r = miu \* r + (1 - miu) \* (grad \*\* 2)  
 w -= (lr \* grad \* np.sqrt(e + delta)) / (np.sqrt(r + delta))  
 e = miu \* e + (1 - miu) \* (w \*\* 2)  
 global\_list['ADADELDA\_r'] = r  
 global\_list['ADADELDA\_e'] = e  
  
  
if \_\_name\_\_ == '\_\_main\_\_':  
 x\_train, y\_train = ds.load\_svmlight\_file('./data/a9a')  
 x\_test, y\_test = ds.load\_svmlight\_file('./data/a9a.t')  
  
 num\_samples, num\_features = x\_train.shape  
 num\_test\_samples, num\_test\_features = x\_test.shape  
  
 x\_train = x\_train.toarray()  
 temp = np.ones(shape=[num\_samples, 1], dtype=np.float32)  
 x\_train = np.concatenate([x\_train, temp], axis=1)  
 x\_test = x\_test.toarray()  
 temp = np.zeros(shape=[num\_test\_samples, 1], dtype=np.float32)  
 temp1 = np.ones(shape=[num\_test\_samples, 1], dtype=np.float32)  
 x\_test = np.concatenate([x\_test, temp, temp1], axis=1)  
 y\_train = y\_train.reshape([len(y\_train), 1])  
 y\_test = y\_test.reshape([len(y\_test), 1])  
  
  
 def shuffle\_train():  
 global x\_train, y\_train  
 rng\_state = np.random.get\_state()  
 np.random.shuffle(x\_train)  
 np.random.set\_state(rng\_state)  
 np.random.shuffle(y\_train)  
  
 batch\_size = 1024  
 data\_size = num\_samples  
  
 def feed\_data(batch\_count):  
 if (1 + batch\_count) \* batch\_size <= data\_size:  
 feed\_dict = {'x': x\_train[batch\_count \* batch\_size:(batch\_count + 1) \* batch\_size],  
 'y': y\_train[batch\_count \* batch\_size:(batch\_count + 1) \* batch\_size]}  
 else:  
 feed\_dict = {'x': x\_train[batch\_count \* batch\_size:data\_size],  
 'y': y\_train[batch\_count \* batch\_size:data\_size]}  
 return feed\_dict  
  
 methods = ['SGD', 'NAG', 'RMSProp', 'Adadelta', 'ADAM']  
 sgd\_para\_list = {'learning\_rate': 0.01}  
 nag\_para\_list = {'miu': 0.9, 'learning\_rate': 0.01}  
 rmsprop\_para\_list = {'delta': 10e-7, 'rho': 0.95, 'learning\_rate': 0.1}  
 adam\_para\_list = {'delta': 10e-8, 'rho1': 0.9, 'rho2': 0.999, 'learning\_rate': 0.1}  
 adadelta\_para\_list = {'delta': 10e-7, 'miu': 0.9, 'learning\_rate': 0.1}  
 para\_list = {'SGD': sgd\_para\_list, 'NAG': nag\_para\_list, 'RMSProp': rmsprop\_para\_list,  
 'Adadelta': adadelta\_para\_list, 'ADAM': adam\_para\_list}  
 for method in methods:  
 iters = []  
 test\_errors = []  
 w = np.random.rand(feature\_size+1, 1)  
 parameter\_list = para\_list[method]  
 count = 0  
 for i in range(0, 3):  
 shuffle\_train()  
 for batch\_count in range(0, int(data\_size / batch\_size) + 1):  
 feed\_dict = feed\_data(batch\_count)  
 iters.append(count)  
 count += 1  
 optimizer(method=method, parameter\_list=parameter\_list, x=feed\_dict['x'], y=feed\_dict['y'], w=w)  
 test\_errors.append(compute\_loss(x\_test, y\_test, w))  
 plt.plot(iters, test\_errors, label=method)  
  
plt.xlabel('Iteration')  
plt.ylabel('Loss')  
plt.legend()  
plt.show()

(Fill in the contents of 8-11 respectively for logistic regression and linear classification)

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

Randomly initializing.

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

**10. Experimental results and curve:**(Fill in this content for various methods of gradient descent respectively)

## Hyper-parameter selection:

SGD:

Learning rate = 0.01

NAG:

Learning rate = 0.01

Miu = 0.9

Momentum\_ini =

RMSProp:

Learning rate = 0.1

Expectation = 1

delta=

Adam:  
delta = 10e-8

rho1 = 0.9

rho2 = 0.999

lr = 0.1

s = 0

r = 0

Adadelta

r = 0

        e = 0

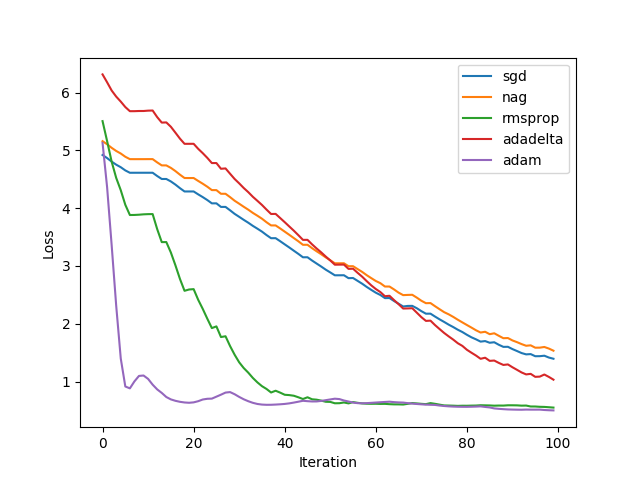
        miu = 0.9

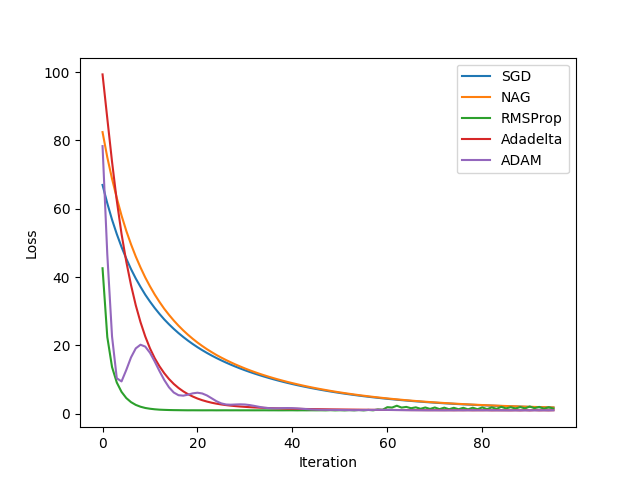
        delta = 10e-7

        lr = 10

## Predicted Results (Best Results):

## Loss curve:





**11. Results analysis:**

**12. Similarities and differences between logistic regression and linear classification：**

**13. Summary:**