

```

from scipy import stats
import urllib
import numpy as np
import scipy.optimize
import pylab as py
import string

def parseData(fname):
    for l in urllib.urlopen(fname):
        yield l

def sigmoid(x):
    return 1/(1+np.exp(-x))

print "Reading data..."
data = list(parseData("new_train3.txt"))
print "done"
d = []
y = []
for i in range(len(data)):
    sub = string.split(data[i])
    for j in range(len(sub)):
        sub[j] = int(sub[j])
    d.append(sub)
    y.append(0.0)

print "Reading data..."
data = list(parseData("new_train5.txt"))
print "done"
for i in range(len(data)):
    sub = string.split(data[i])
    for j in range(len(sub)):
        sub[j] = int(sub[j])
    d.append(sub)
    y.append(1.0)

ita = [0.001 for i in range(64)]
w = [0.0 for i in range(64)]
log_like = []
for i in range(5000):
    print i, ": "
    gradient = [0.0 for q in range(64)]
    log = 0.0
    for k in range(len(d)):
        coeff = np.array([y[k] - sigmoid(np.dot(w,d[k])) for q in range(64)])
        log += y[k]*np.log(sigmoid(np.dot(w,d[k]))) + (1-y[k])*np.log(sigmoid(-np
            .dot(w,d[k])))
        gradient += np.multiply(coeff ,d[k])
    print "gradient, ", gradient
    print "log: ", log
    log_like.append(log)
    w = w + np.multiply(ita, gradient)
    print "w: ", w
print "log_like: ", log_like

```

```
print "Reading data..."
data = list(parseData("new_test3.txt"))
print "done"
d_d = []
y_y = []
for i in range(len(data)):
    sub = string.split(data[i])
    for j in range(len(sub)):
        sub[j] = int(sub[j])
    d_d.append(sub)
    y_y.append(0.0)

print "Reading data..."
data = list(parseData("new_test5.txt"))
print "done"
for i in range(len(data)):
    sub = string.split(data[i])
    for j in range(len(sub)):
        sub[j] = int(sub[j])
    d_d.append(sub)
    y_y.append(1.0)

sum = 0.0
for i in range(400):
    if sigmoid(np.dot(w,d_d[i])) > 0.5:
        diff = 1.0 - y_y[i]
    else:
        diff = 0.0 - y_y[i]
    sum += np.abs(diff)
sum /= 400
print "error on test data 3: ", sum
sum1 = sum
sum = 0.0
for i in range(400, 800):
    if sigmoid(np.dot(w,d_d[i])) > 0.5:
        diff = 1.0 - y_y[i]
    else:
        diff = 0.0 - y_y[i]
    sum += np.abs(diff)
sum /= 400
print "error on test data 5: ", sum

for i in range(len(data)):
    sub = string.split(data[i])
    for j in range(len(sub)):
        sub[j] = int(sub[j])
    d.append(sub)
    y.append(1.0)

sum = 0.0
for i in range(700):
    if sigmoid(np.dot(w,d[i])) > 0.5:
        diff = 1.0 - y[i]
    else:
        diff = 0.0 - y[i]
    sum += np.abs(diff)
```

```
sum /= 700
print "error on training data 3: ", sum
sum = 0.0
for i in range(700,1400):
    if sigmoid(np.dot(w,d[i])) > 0.5:
        diff = 1.0 - y[i]
    else:
        diff = 0.0 - y[i]
    sum += np.abs(diff)
sum /= 700
print "error on training data 5: ", sum
```