import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

data = pd.read\_csv(r'C:\Users\18801\Documents\WeChat Files\wxid\_8vs9i3nhli4u22\FileStorage\File\2019-05\2017fall-ml-hw1-master\train.csv')

pm2\_5 = data[data['class']=='PM2.5'].ix[:, 3:]

tempxlist = []

tempylist = []

for i in range(15):

tempx = pm2\_5.iloc[:, i:i+9]

tempx.columns = np.array(range(9))

tempy = pm2\_5.iloc[:, i+9]

tempy.columns = ['1']

tempxlist.append(tempx)

tempylist.append(tempy)

xdata = pd.concat(tempxlist)

x = np.array(xdata, float)

ydata = pd.concat(tempylist)

y = np.array(ydata, float)

x = np.concatenate((np.ones((x.shape[0], 1)), x), axis=1)

w = np.zeros((len(x[0])))

lr = 10

iteration = 10000

s\_grad = np.zeros(len(x[0]))

for i in range(iteration):

tem = np.dot(x, w)

loss = y - tem

grad = np.dot(x.transpose(), loss)\*(-2)

s\_grad += grad\*\*2

ada = np.sqrt(s\_grad)

w = w - lr\*grad/ada

testdata = pd.read\_csv(r'C:\Users\18801\Documents\WeChat Files\wxid\_8vs9i3nhli4u22\FileStorage\File\2019-05\2017fall-ml-hw1-master\test.csv')

pm2\_5\_test = testdata[testdata['class']=='PM2.5'].ix[:, 2:]

x\_test = np.array(pm2\_5\_test, float)

x\_test\_b = np.concatenate((np.ones((x\_test.shape[0], 1)), x\_test), axis=1)

y\_star = np.dot(x\_test\_b, w)

y\_pre = pd.read\_csv(r'C:\Users\18801\Documents\WeChat Files\wxid\_8vs9i3nhli4u22\FileStorage\File\2019-05\2017fall-ml-hw1-master\sampleSubmission.csv', encoding='gbk')

y\_pre.value = y\_star

real = pd.read\_csv(r'C:\Users\18801\Documents\WeChat Files\wxid\_8vs9i3nhli4u22\FileStorage\File\2019-05\2017fall-ml-hw1-master\ans.csv')

erro = abs(y\_pre.value - real.value).sum()/len(real.value)

print(erro)

输出：4.97442948413