Homework 2: Multivariate Linear Regression

Student ID: Student Name:

Lectured by 梁上松, Sun Yat-sen University

In this homework, you will investigate multivariate linear regression using Gradient Descent and Stochastic Gradient Descent. You will also examine the relationship between the cost function, the convergence of gradient descent, overfitting problem, and the learning rate.
在本次作业中，你将探讨使用梯度下降法（随机梯度下降法）的多变量线性回归模型。你将探讨损失函数、梯度下降法的收敛、过拟合问题和学习率等之间的关系。

Download the file "dataForTraining.txt" in the attached files called "Homework 2". This is a training dataset of apartment prices in Haizhu District, Guangzhou, Guangdong, China, where there are 50 training instances, one line per one instance, formatted in three columns separated with each other by a whitespace. The data in the first and the second columns are sizes of the apartments in square meters and the distances to the Double-Duck-Mountain Vocational Technical College in kilo-meters, respectively, while the data in the third are the corresponding prices in billion RMB. Please build a multivariate linear regression model with the training instances by script in any programming languages to predict the prices of the apartments. For evaluation purpose, please also download the file "dataForTesting.txt" (the same format as that in the file of training data) in the same folder.
请在文件夹"作业 2"中下载文件名为"dataForTraining.txt"的文件。该文件包含广东省广州市海珠区的房价信息，里面包含 50 个训练样本数据。文件有三列，第一列对应房的面积（单位：平方米），第二列对应房子距离双鸭山职业技术学院的距离（单位：千米），第三列对应房子的销售价格（单位：万元）。每一行对应一个训练样本。请使用提供的 50 个训练样本来训练多变量回归模型以便进行房价预测，请用（随机）梯度下降法的多变量线性回归模型进行建模。为了评估训练效果，请文件夹中下载测试数据集"dataForTesting.txt"（该测试文件里的数据跟训练样本具有相同的格式，即第一列对应房子面积，第二列对应距离，第三列对应房子总价）。

Exercise 1: How many parameters do you use to tune this linear regression model? Please use Gradient Descent to obtain the optimal parameters. Before you train the model, please set the number of iterations to be 1500000, the learning rate to 0.00015, the initial values of all the parameters to 0.0. During training, at every 100000 iterations, i.e., 100000 , 200000,…, 1500000, report the current training error and the testing error in a figure (you can draw it by hands or by any software). What can you find in the plots? Please analyze the plots.
Exercise 1: 你需要用多少个参数来训练该线性回归模型？请使用梯度下降方法训练。训练时，请把迭代次数设成 1500000，学习率设成 0.00015，参数都设成 0.0。在训练的过程中，每迭代 100000 步，计算训练样本对应的误差，和使用当前的参数得到的测试样本对应的误差。请画图显示迭代到达 100000 步、200000 步、… … 1500000 时对应的训练样本的误差和测试样本对应的误差（图可以手画，或者用工具画图）。从画出的图中，你发现什么？请简单分析。

3 个参数，y (总价) = a0 + a1 * x1(面积) + a2 * x2(距离);
回归模型：y=a0+a1*x1+a2*x2
学习率：0. 00015

回归模型	学习率	迭代次数(*100000)	参数 a0	参数 a1	参数 a2	训练误差	测试误差	发现
		1	0.10812813247090658	10.692963570990205	-1.5249960902955981	2.6362554159157623E7	2989081.797270165	训练误差小，测试误差大
		2	72.18797769726687	6.919449271079721	-72.37587220002476	4750.2916493746 75	3781.8308121 11741	训练误差收敛，测试误差小

		3	72.187977697 26687	6.919449271079721	-72.37587220002476	4750.2916493746 75	3781.8308121 11741	训练误差收敛，测试误差小
		4	72.187977697 26687	6.919449271079721	-72.37587220002476	4750.2916493746 75	3781.8308121 11741	训练误差收敛，测试误差小
		5	72.187977697 26687	6.919449271079721	-72.37587220002476	4750.2916493746 75	3781.8308121 11741	训练误差收敛，测试误差小
		6	72.187977697 26687	6.919449271079721	-72.37587220002476	4750.2916493746 75	3781.8308121 11741	训练误差收敛，测试误差小
		7	72.187977697 26687	6.919449271079721	-72.37587220002476	4750.2916493746 75	3781.8308121 11741	训练误差收敛，测试误差小
		8	72.187977697 26687	6.919449271079721	-72.37587220002476	4750.2916493746 75	3781.8308121 11741	训练误差收敛，测试误差小
		9	72.187977697 26687	6.919449271079721	-72.37587220002476	4750.2916493746 75	3781.8308121 11741	训练误差收敛，测试误差小
		10	72.187977697 26687	6.919449271079721	-72.37587220002476	4750.2916493746 75	3781.8308121 11741	训练误差收敛，测试误差小
		11	72.187977697 26687	6.919449271079721	-72.37587220002476	4750.2916493746 75	3781.8308121 11741	训练误差收敛，测试误差小
		12	72.187977697 26687	6.919449271079721	-72.37587220002476	4750.2916493746 75	3781.8308121 11741	训练误差收敛，测试误差小
		13	72.187977697 26687	6.919449271079721	-72.37587220002476	4750.2916493746 75	3781.8308121 11741	训练误差收敛，测试误差小
		14	72.187977697 26687	6.919449271079721	-72.37587220002476	4750.2916493746 75	3781.8308121 11741	训练误差收敛，测试误差小
		15	72.187977697 26687	6.919449271079721	-72.37587220002476	4750.2916493746 75	3781.8308121 11741	训练误差收敛，测试误差小

发现：应该以测试误差为标准
随机梯度下降方法收敛比较快
训练误差随着迭代而收敛

Exercise 2: Now, you change the learning rate to a number of different values, for instance, to 0.0002 (you may also change the number of iterations as well) and then train the model again. What can you find? Please conclude your findings.
Exercise 2: 现在，你改变学习率，比如把学习率改成 0.0002（此时，你可以保持相同的迭代次数也可以改变迭代次数），然后训练该回归模型。你有什么发现？请简单分析。

学习率：0.0002
迭代 100000， 训练误差为 3.453137881561735E7，测试误差为 4333215.5044527985
迭代 200000-1500000，训练误差为 1502.1993468036187，测试误差为 2266.039871215753
发现：

随着迭代而收敛，训练误差和测试误差都有所减小，学习率的选取影响训练误差和测试误差，选取学习率对回归模型的建立准确率影响很大

Exercise 3: Now, we turn to use other optimization methods to get the optimal parameters. Can you use Stochastic Gradient Descent to get the optimal parameters? Plots the training error and the testing error at each K-step iterations (the size of K is set by yourself). Can you analyze the plots and make comparisons to those findings in Exercise 1?
Exercise 3: 现在，我们使用其他方法来获得最优的参数。你是否可以用随机梯度下降法获得最优的参数？请使用随机梯度下降法画出迭代次数（每 K 次，这里的 K 你自己设定）与训练样本和测试样本对应的误差的图。比较 Exercise 1 中的实验图，请总结你的发现。

迭代（*200）	训练误差（保留两位）	测试误差（保留两位）	发现		
0	3.45	4333215.50	训练误差小	测试误差大	未收敛
1	951183.77	122746.76	训练误差大	测试误差大	未收敛
2	205095.82	20929.36	训练误差减小	测试误差减小	未收敛
3	41630.29	2295.27	训练误差减小	测试误差减小	未收敛
4	7619.46	248.98	训练误差减小	测试误差减小	未收敛
5	1442.10	838.02	训练误差减小	测试误差最小	未收敛
6	791.84	1477.53	训练误差最小	测试误差增大	未收敛
7	1006.97	1864.95	训练误差增大	测试误差增大	未收敛
8	1229.93	2068.81	训练误差增大	测试误差增大	未收敛
9	1363.94	2170.51	训练误差增大	测试误差增大	未收敛
10	1434.26	2220.09	训练误差增大	测试误差增大	未收敛
11	1469.30	2244.01	训练误差增大	测试误差增大	未收敛
12	1486.38	2255.49	训练误差增大	测试误差增大	未收敛
13	1494.61	2260.99	训练误差增大	测试误差增大	未收敛
14	1498.57	2263.62	训练误差增大	测试误差增大	未收敛
15	1500.46	2264.88	训练误差增大	测试误差增大	未收敛
16	1501.36	2265.48	训练误差收敛	测试误差收敛	收敛
17	1501.80	2265.77	训练误差收敛	测试误差收敛	收敛
18	1502.00	2265.91	训练误差收敛	测试误差收敛	收敛
19	1502.10	2265.97	训练误差收敛	测试误差收敛	收敛
20	1502.15	2266.01	训练误差收敛	测试误差收敛	收敛
21	1502.17	2266.02	训练误差收敛	测试误差收敛	收敛
22	1502.18	2266.03	训练误差收敛	测试误差收敛	收敛
23	1502.19	2266.03	训练误差收敛	测试误差收敛	收敛
24	1502.19	2266.03	训练误差收敛	测试误差收敛	收敛

发现：
stochastic gradient descent （随机梯度下降），
速度比较快，迭代 3000 多就开始收敛，
但是收敛性能不太好，可能在最优点附近晃来晃去，hit 不到最优点
两次参数的更新也有可能互相抵消掉，造成目标函数震荡的比较剧烈
可以借用批梯度下降方法（batch gradient descent）的优点调整算法