Assignment 1[1]

1.修改梯度下降代码*Gradient descent.ipynb*中的自定义函数*gradient_descent*使其能自动选择最优步长参数*learning_rate*。

提示： 最优步长是使得目标函数达到最小的步长值，可将$[0,1]$区间等分，即$\{0,0.01,0.02,\cdots,0.99,1\}$，在每次迭代时分别计算*learning_rate*取区间内不同值时目标函数的值，再选择其中使得目标函数值最小的步长值。

[1]Due date: 6:30pm 3/17/2025

2.从清华大学开放中文词库下载财经和地名语料库，将其导入Python，删除词频数，将每个词作为一个元素构建一个词典列表。以该词典为基础，利用代码*re and segmentation.ipynb*中的自定义函数*cut*对如下文本进行分词。

“据中指研究院统计，自*9*月*30*日人民银行发布消息决定下调首套个人住房公积金贷款利率以来，截至*10*月*9*日，已有杭州、济南、吉林等至少*30*个地区的公积金管理中心发布相关通知，落实下调首套个人住房公积金贷款利率*0.15*个百分点。近期，包括下调首套个人住房公积金贷款利率、阶段性放宽部分城市首套住房贷款利率下限、支持居民换购住房个人所得税退税优惠政策等接连出台。”

3.修改代码*re and segmentation.ipynb*中的自定义函数*cut*，使其能够使用逆向最大匹配法和双向最大匹配法实现分词，并以问题**2**中的词典和文本测试三种不同分词方法的效果，哪种分词方法的准确率较高？

4.以中文酒店评论.*txt*为文本库（其中，review为用户评论，label为标签，label=1为正面评论，label=-1为负面评论）。

(i)将数据导入Python，分别统计正面评论和负面评论的条数。

(ii)分别对每条评论进行分词并统计词数（或句子长度）。

(iii)分别绘制正面评论和负面评论句子长度的直方图。

(iv)删除(ii)中每条评论分词后得到的非形容词，只保留每条评论分词后得到的形容词。

(v)分别绘制正面评论和负面评论高频形容词的词云图。

提示：需将(iv)中每条评论的形容词列表按正负评论分别合并为两个大列表后才能绘制词云图。