**ML实验报告**

姓名： 韩轶凡 学号： 1809401053 日期： 2020年3月8日

**实验名称： 梯度下降法求解回归问题**

注：本实验中参数b初始值为各位都为1的向量。

W向量初始化为各位都是一个（-10，10）的随机数的随机向量。

其余参数都在各实验组中说明。

**实验结果：**

**实验编号① 步长 = 1 \* 10^(-7)**

第10000次迭代结果为

[-7.48408068 -0.54191628 -6.72478301 -2.26838898 5.12485018 -3.52853194

-3.5495761 5.45465557 -2.60429391 7.10872007]

第20000次迭代结果为

[-8.10351811 -1.19066665 -7.56965766 -2.67175516 4.1101787 -4.16939065

-4.25750791 4.76967856 -3.39526435 6.07032692

第40000次迭代结果为

[-8.36780365 -1.47992941 -8.25568696 -2.51330259 3.0711428 -4.44771715

-4.71826701 4.37213423 -4.05292788 4.97256163]

第100000次迭代结果为

[-7.86831141 -0.99688586 -8.86535808 -0.85602484 1.46890341 -3.92559847

-4.77091439 4.49518802 -4.72418798 3.15460959]

第200000次迭代结果为

[-7.02044613 -0.13444789 -9.464825 1.57935355 -0.49301646 -2.95860518

-4.63020942 4.80691212 -5.55568808 0.76473287]

第500000次迭代结果为

[ -5.49424972 1.52695526 -10.34062203 6.08383166 -3.93432326

-0.84496178 -4.18764467 5.20953405 -7.48213074 -4.06134422]

第1000000次迭代结果为

[ -4.53397556 2.63245793 -10.55227333 9.01018011 -5.91147483

0.97257965 -3.62954011 5.17546594 -9.26855821 -7.74414262]

第2000000次迭代结果为

[ -4.10795074 3.01710332 -10.33468636 10.22024789 -6.53724334

2.03735918 -3.19208438 4.8821317 -10.36840511 -9.6326269 ]

第2460000次迭代结果为

[ -4.06039341 3.02801032 -10.27902176 10.32151272 -6.56869699

2.15770581 -3.13506561 4.81993779 -10.49231592 -9.82438379]

**分析：迭代次数过多，与已知结果仍有差距，考虑缩小步长。**

**实验编号② 步长 = 1 \* 10^(-6)**

第10000次迭代结果为

[-11.04842561 -5.50587185 -5.97766654 4.6961443 0.60841857

-4.9481442 2.36575223 -5.153078 1.85854113 0.80096101]

第20000次迭代结果为

[-9.20676444 -3.759399 -7.05494056 5.720686 -0.93210027 -3.57397055

0.69681279 -2.62062108 -0.55558744 -1.40008286]

第40000次迭代结果为

[-6.81569143 -1.24828541 -8.50958789 7.23509411 -3.07906605 -1.57287806

-1.39738373 0.76778619 -4.09980762 -4.47785369]

第80000次迭代结果为

[-4.77618614 1.3950936 -9.84427686 8.88210713 -5.21631251 0.58392443

-3.04123897 3.83904402 -7.96293016 -7.58808744]

第120000次迭代结果为

[ -4.16867897 2.49720077 -10.27454997 9.58356778 -6.04007936

1.51587935 -3.43301163 4.84136018 -9.63238316 -8.83590241]

第160000次迭代结果为

[ -3.99133621 2.96151057 -10.40119319 9.88117727 -6.36386637

1.92517398 -3.49571969 5.15824275 -10.35518403 -9.35633402]

第500000次迭代结果为

[ -4.51251689 2.86112767 -10.11368947 10.38163073 -6.33290503

2.54930128 -3.44182945 4.70286329 -10.54718439 -9.59207213]

第750000次迭代结果为

[ -4.51189469 2.86061283 -10.11293201 10.38224061 -6.33282358

2.55020599 -3.4415772 4.70196383 -10.54792638 -9.59323482]

第1000000次迭代结果为

[ -4.51188418 2.86060317 -10.11292097 10.38224863 -6.33282148

2.5502181 -3.44157364 4.70195025 -10.54793534 -9.59325105]

**分析：迭代次数仍然过多，后期更新变化程度小，考虑继续缩短步长。**

**实验编号③ 步长 = 1 \* 10^(-5)**

第10000次迭代结果为

[ -3.65318203 1.88960616 -10.21909525 9.59889986 -5.68922417

1.69990067 -3.83315831 4.54777227 -9.78796753 -8.48941924]

第20000次迭代结果为

[ -4.45229061 3.17284382 -10.22397998 10.0207257 -6.56260156

1.85274708 -2.95696105 4.82982076 -10.20275072 -9.50003076]

第40000次迭代结果为

[ -4.56281618 3.3546348 -10.23071999 10.05403439 -6.65928513

1.88682947 -2.83953785 4.88755621 -10.29638319 -9.63064552]

第80000次迭代结果为

[ -4.56462918 3.35768082 -10.23106898 10.05423799 -6.6607174

1.88752157 -2.83769184 4.88881143 -10.29832313 -9.63239591]

第120000次迭代结果为

[ -4.56462983 3.35768184 -10.2310693 10.05423786 -6.66071786

1.88752162 -2.83769141 4.88881201 -10.29832358 -9.63239603]

第130000次迭代结果为

[ -4.56462983 3.35768184 -10.2310693 10.05423786 -6.66071786

1.88752162 -2.83769141 4.88881201 -10.29832358 -9.63239603]

第140000次迭代结果为

[ -4.56462983 3.35768184 -10.2310693 10.05423786 -6.66071786

1.88752162 -2.83769141 4.88881201 -10.29832358 -9.63239603]

第150000次迭代结果为

[ -4.56462983 3.35768184 -10.2310693 10.05423786 -6.66071786

1.88752162 -2.83769141 4.88881201 -10.29832358 -9.63239603]

第160000次迭代结果为

[ -4.56462983 3.35768184 -10.2310693 10.05423786 -6.66071786

1.88752162 -2.83769141 4.88881201 -10.29832358 -9.63239603]

**分析：后期迭代结果没有变化，但仍与标准答案有差距，考虑继续缩短步长来观察结果。**

**实验编号④ 步长 = 1 \* 10^(-4)**

第10000次迭代结果为

[ -4.19337819 3.68709796 -10.52152483 10.18292266 -6.22062163

1.77190176 -3.282353 4.82652605 -10.66668865 -9.66068293]

第20000次迭代结果为

[ -4.1933782 3.68709798 -10.52152483 10.18292267 -6.22062164

1.77190175 -3.282353 4.82652605 -10.66668866 -9.66068293]

第30000次迭代结果为

[ -4.1933782 3.68709798 -10.52152483 10.18292267 -6.22062164

1.77190175 -3.282353 4.82652605 -10.66668866 -9.66068293]

第40000次迭代结果为

[ -4.1933782 3.68709798 -10.52152483 10.18292267 -6.22062164

1.77190175 -3.282353 4.82652605 -10.66668866 -9.66068293]

第50000次迭代结果为

[ -4.1933782 3.68709798 -10.52152483 10.18292267 -6.22062164

1.77190175 -3.282353 4.82652605 -10.66668866 -9.66068293]

第60000次迭代结果为

[ -4.1933782 3.68709798 -10.52152483 10.18292267 -6.22062164

1.77190175 -3.282353 4.82652605 -10.66668866 -9.66068293]

第70000次迭代结果为

[ -4.1933782 3.68709798 -10.52152483 10.18292267 -6.22062164

1.77190175 -3.282353 4.82652605 -10.66668866 -9.66068293]

第80000次迭代结果为

[ -4.1933782 3.68709798 -10.52152483 10.18292267 -6.22062164

1.77190175 -3.282353 4.82652605 -10.66668866 -9.66068293]

第90000次迭代结果为

[ -4.1933782 3.68709798 -10.52152483 10.18292267 -6.22062164

1.77190175 -3.282353 4.82652605 -10.66668866 -9.66068293]

第100000次迭代结果为

[ -4.1933782 3.68709798 -10.52152483 10.18292267 -6.22062164

1.77190175 -3.282353 4.82652605 -10.66668866 -9.66068293]

第110000次迭代结果为

[ -4.1933782 3.68709798 -10.52152483 10.18292267 -6.22062164

1.77190175 -3.282353 4.82652605 -10.66668866 -9.66068293]

第120000次迭代结果为

[ -4.1933782 3.68709798 -10.52152483 10.18292267 -6.22062164

1.77190175 -3.282353 4.82652605 -10.66668866 -9.66068293]

**分析：到后期迭代结果仍然没有变化。**

**实验总结：**

由于第一次尝试，并不了解步长可能对实验带来的影响，因而一开始选择用较小的步长来进行实验（实验①），在实验中发现后期的结果变化幅度较小，且仍与预期结果有偏差，于是考虑逐步减小步长，继续进行实验（实验②③④），但是发现，到后期的迭代结果没有任何变化。同时，我还做了一组步长为10^(-3)的对照组，结果没有记录在实验报告中，但是可以发现随着迭代次数增加，后期迭代结果即w的数值呈指数增长，与预期结果相差甚远。

综上所述，根据本实验结果，认为步长在10^（-4）到10^（-6）效果较好，但是，仍未计算得到老师所给的标准答案，分析原因可能为input与output的结果都为浮点数，多次迭代导致小数部分的误差被累加，进而与标准答案产生偏差。

需要指出的，本次实验的循环条件为判断loss function的值为0即终止循环，但是，也曾尝试过将循环条件改为判断loss function小于某一值即终止，仍无法得到理想的结果，故该部分实验并未记录。

可以总结得到，当步长较短时，后期无明显变化，而步长较长时，会出现急速的增长情况。本次实验结果基本与标准答案接近，但仍存在一些细节问题，导致无法做到与标准答案相同，需要在日后继续修正、改善，同时希望老师能够给出指导，帮助指出其中的导致错误的原因，以便学习经验教训。

**实验名称： 最小二乘法求解回归问题**

**实验结果：**

W = [ -3.99999867 2.99999638 -9.99999869 10.00000127 -5.99999705

1.99999934 -3.00000113 4.99999976 -10.00000057 -9.00000042]

**实验总结：**

相对于最小二乘法来说，梯度下降法须要归一化处理以及选取学习速率 ，且需多次迭代更新来求得最终结果，而最小二乘法则不需要。  
 相对于梯度下降法来说，最小二乘法须要求解（XTX）的逆，其计算量为O（n3），当训练数据集过于庞大的话，其求解过程非常耗时，而梯度下降法耗时相对较小。  
 所以，当模型相对简单，训练数据集相对较小，用最小二乘法较好；对于更复杂的学习算法或者更庞大的训练数据集，用梯度下降法较好，一般当特征变量小于10的四次方时，使用最小二乘法较稳妥，而大于10的四次方时，则应该使用梯度下降法来降低计算量。

在本实验中，显然最小二乘法比梯度下降法具有优势。