1. **实验：**

确定假设为：

使用梯度下降方法，分别对单参量，双参量，三参量，……，七参量进行实验。

使用训练数据集为：全部训练集

大训练数据集为2015年1月1日到2015年5月31日。中间两个奇异点，第十周3月6日到3月14日的春节和5月1日到5月3日的劳动节。将其取出之后，总共20周数据。将前19周作为训练集，最后第20周作为测试集。

1. **实验结果为：**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 迭代次数  参量数 | 15000  0.003 | 150000  0.003 | 150000  0.0003 | 150000  0.00003 | 1500000  0.00003 |
| 1 | 69.22% | 69.22% | 69.22% | 69.22% | 69.22% |
| 2 | 69.28% | 69.28% | 69.28% | 69.28% | 69.28% |
| 3 | 53.79% | 53.79% | **74.98%** | 74.98% | 74.98% |
| 4 | 48.88% | 48.88% | 68.77% | 68.77% | 68.77% |
| 5 | 47.18% | 47.18% | 67.46% | 67.46% | 67.46% |
| 6 | 44.68% | 44.68% | 62.99% | 62.99% | 62.99% |
| 7 | 41.87% | 41.87% | 50.73% | 50.73% | 50.73% |

1. **想法：**

在线性回归的尝试中，使用三参量，迭代次数为150000，学习率为0.0003的情况下，能达到最佳的预测精度。即模型为：

|  |  |  |  |
| --- | --- | --- | --- |
| w  Flavor | W1 | W2 | W3 |
| 1 | 0.5531 | 0.1107 | 0.5471 |
| 2 | 0.0094 | 0.6357 | 0.3685 |
| 3 | 0.2333 | 0.2087 | 0.3880 |
| 4 | 0.2109 | 0.1689 | 0.4116 |
| 5 | 0.6612 | 0.3890 | 0.0023 |
| 6 | 0.4272 | 0.3807 | 0.0398 |
| 7 | 0.3232 | 0.4912 | -0.2321 |
| 8 | 0.0617 | 0.7448 | 0.3355 |
| 9 | 0.4951 | 0.1660 | 0.0725 |
| 10 | 0.1467 | 0.1200 | 0.1467 |
| 11 | 0.1986 | 0.2792 | 0.4261 |
| 12 | 0.2827 | -0.0484 | 0.1143 |
| 13 | 0 | 0 | 0 |
| 14 | 0.1001 | 0.1422 | 0.1586 |
| 15 | 0.0068 | -0.0201 | 0.0536 |

线性回归训练出来的模型有两大问题：

1. 数据中出现负数等，这应该不正确，而且并没有遵循，时间越近，权重越大的准则。因此下一步考虑逻辑回归，使用来进行尝试。
2. 在回归过程中，对于奇异点的处理，是直接忽略，而且使用的数据集也并没有考虑最后12月份和2016年1月份的数据，不知道如何处理。