系统工程第5次作业

张博睿 自75 2017011537

程序详见

# 1.

### （1）使用PCA+线性回归结果

**算法流程**

|  |
| --- |
| **输入：**数据和标签。 |
| （1）对数据进行归一化并保存数据的均值和方差  （2）计算归一化数据的协方差矩阵  （3）对协方差矩阵进行特征值分解，并根据选择占比较大的特征值和特征向量（主成分）。  （4）计算降维后的数据  （5）通过OLS计算和的回归系数并进行F检验。  （6）根据保存的均值和方差恢复关于的回归系数。 |
| **输出：**最终的回归系数和相关中间过程结果。 |

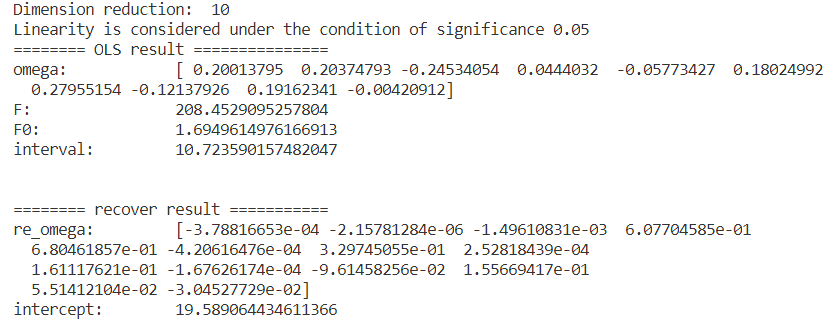
**最终结果如下**

|  |  |
| --- | --- |
| 项目 | 结果 |
| 显著性水平 |  |
| 显著性检验 |  |
| 计算 |  |
| 显著性检验 | 存在线性 |
| 病态特征值阈值 |  |
| 降维维度 |  |
| 置信区间 |  |

回归系数为

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 项目 |  |  |  |  |
| 系数 |  |  |  |  |
| 项目 |  |  |  |  |
| 系数 |  |  |  |  |
| 项目 |  |  |  |  |
| 系数 |  |  |  |  |
| 项目 |  |  |  |  |
| 系数 |  |  |  |  |

**运行的原始结果如下**

****

### （2）直接使用病态回归建模

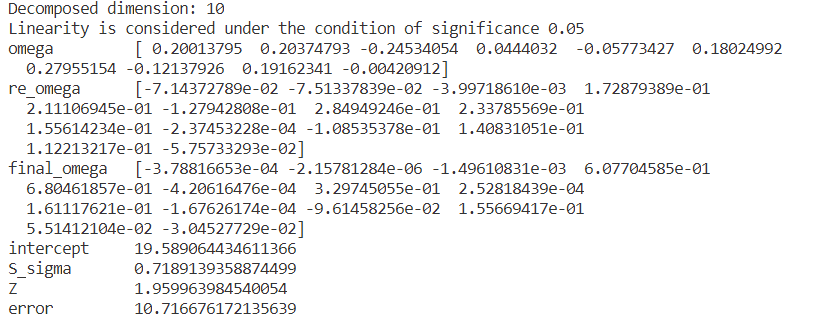
**最终结果如下**

|  |  |
| --- | --- |
| 项目 | 结果 |
| 显著性水平 |  |
| 显著性检验 |  |
| 计算 |  |
| 显著性检验 | 存在线性 |
| 病态特征值阈值 |  |
| 降维维度 |  |
| 置信区间 |  |

回归系数为

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 项目 |  |  |  |  |
| 系数 |  |  |  |  |
| 项目 |  |  |  |  |
| 系数 |  |  |  |  |
| 项目 |  |  |  |  |
| 系数 |  |  |  |  |
| 项目 |  |  |  |  |
| 系数 |  |  |  |  |

**运行的原始结果如下**



### （3）分析

可以看到，使用**PCA+线性回归**和**直接使用病态回归**的结果基本一致，下面对比分析两者的不同点和相同点。

**不同点：**

|  |  |  |
| --- | --- | --- |
|  | +线性回归 | 病态回归 |
| 优化目标 |  |  |
| 特征分解 | 计算协方差矩阵（除以n-1） | 没有除以n-1 |

**相同点：**

在处理实际问题是，两者都是通过计算协方差矩阵的特征分解，并通过选择较大特征值的方式来达到优化目标。

**分析：**

通过实验发现，两种方法得到的结果基本是一致的。回到两种方法的目标函数可以发现，尽管两种方法的优化目标存在一定的差异，但是在实际运算的时候可以用相似的方法（特征值分解）来优化目标函数。

由于主成分分析的时候强调了协方差矩阵，计算时除以了系数n-1，而病态回归问题没有。但是这在选择特征值和特征向量（主成分）时并没有太多的影响（可能存在数值计算的差别）。最终导致两种方法的效果基本是一致的。