

Programming:

3.3 见 least\_sq\_L1.py

3.4 画出三个 train set 共 15 个结果图，如图 1 所示

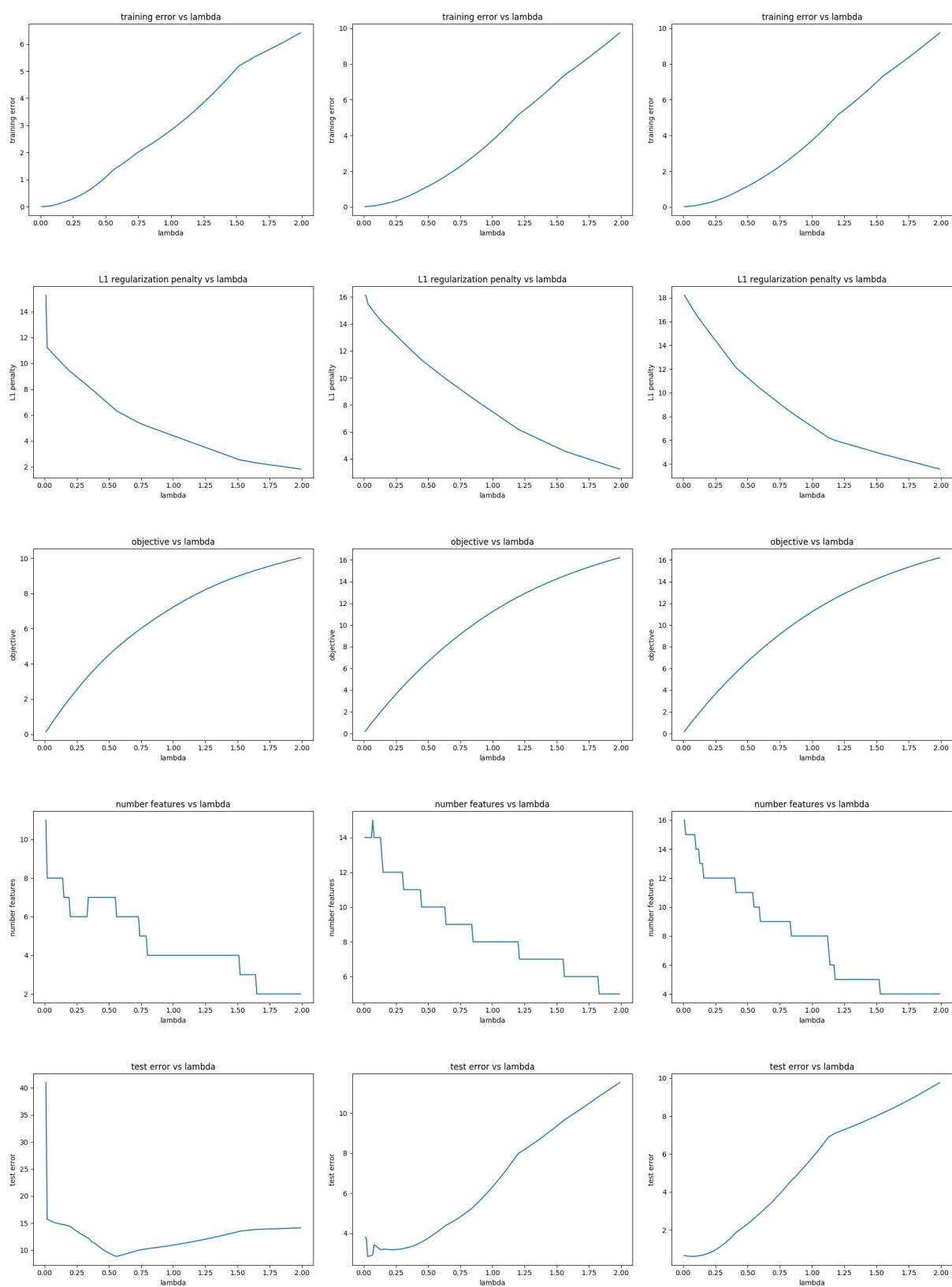


图 1: 五种特征与  $\lambda$  关系

其中第一列为 train\_small set 结果图，第二列为 train\_mid set 结果图，第三列为 train\_large set 结果图。每一行为同一种特征与 lambda 关系。

首先分析这五个特征与 lambda 的关系：从图中易知，训练误差 training error 与 lambda 正相关；L1 正则项 L1 regularization penalty 与 lambda 负相关；目标函数 objective function 与 lambda 正相关，且为上凸函数，曲线平滑；特征数 number of features 与 lambda 负相关，且随着 lambda 增大呈现阶梯状下降，最大特征数为 16 不超过数据维数  $D=16$ ；测试误差 test error 随着 lambda 增大先减小后增大。

图像中有几个不平滑点，可能是数据问题，或者计算过程中出现奇点：第一列 lambda=0 时，L1 惩罚项和测试误差均很大；第二列 lambda 较小时，特征数有一个跳变，测试误差有一段不平滑。

最小测试误差与对应的 lambda 如下表 1 所示。

表 1：三个测试集的最小测试误差与对应的 lambda

|            | train_small | train_mid   | train_large |
|------------|-------------|-------------|-------------|
| 最小测试误差     | 8.85        | 2.84 (3.18) | 0.61        |
| 对应的 lambda | 0.56        | 0.03 (0.25) | 0.07        |

由图像分析结合奇点可知，train\_mid 最小测试误差对应的 lambda 应为 0.25 附近。所以可以得出结论，随着训练集规模增大，最小测试误差减小，lambda 减小。在实际运用中，可以根据训练数据的规模来选择 lambda 的适当值，不能太大也不能太小，随着规模增大要减小 lambda 的值，从而最大程度地降低测试误差。