使用说明书

1. **运行环境**

* Python：pandas，numpy，matplotlib，seaborn，sklearn，skimage，pickle，tqdm，timeit，joblib
* 安装有Jupyter
* S3 Browser

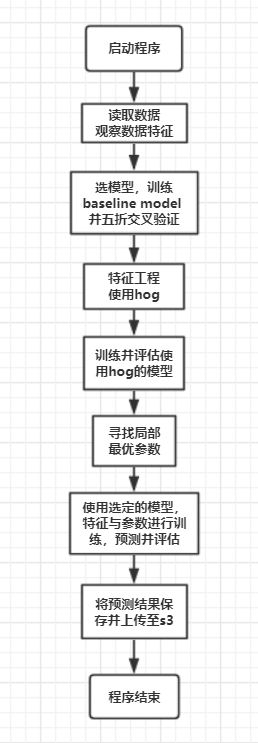
1. **运行前**

* 将运行环境安装准备好
* 先将数据集archive.zip（kaggle数据集中的Sign Language MNIST）下载解压好，本程序只使用其中的sign\_mnist\_train.csv与sign\_mnist\_test.csv两个文件。
* 将代码中的数据集文件路径修改正确
* 将要保存的预测结果的文件路径修改正确
* 将S3 Browser的配置信息填写完整

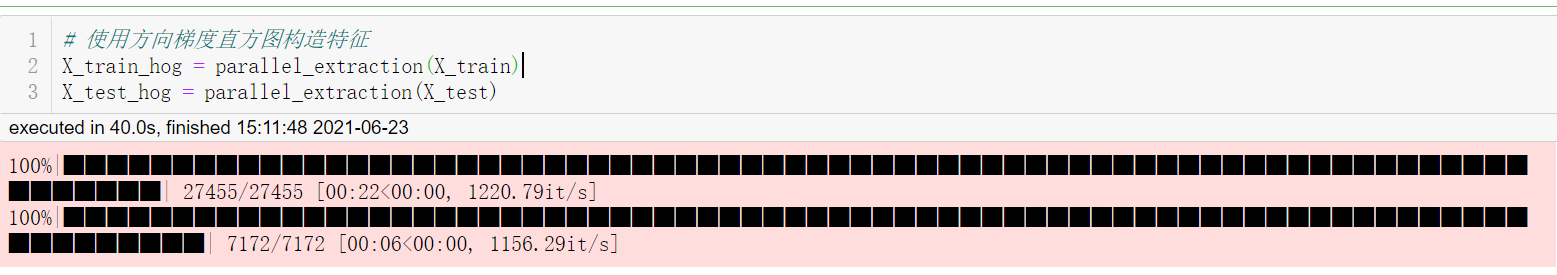
1. **运行时**

在将运行前的准备完成后，可以直接运行。运行过程中可以看到每一块代码的运行结果，也可以跳过基准模型，寻找新特征模型，寻找更好的参数模型，直接到最终选择的模型进行训练。模型采用逻辑回归进行训练，算法采用拟牛顿法的lbfgs进行收敛。基线模型采用了五折交叉验证进行评估。在特征工程中只使用了特征提取，使用方向梯度直方图作为新特征进行训练。确定下选用的特征后，开始寻找较优参数，使用主成分分析，在一个范围内循环调整pca值与估计量，寻找局部最优组合。最终训练完成后，使用测试集进行预测，并在最终的模型中采用准确率，混淆矩阵，以及分类报告进行评估。最后将预测结果上传写入S3中存储。

1. **流程图与部分运行结果图**







**寻找局部最优参数**

