设计说明书

1. **设计程序的目的**

使用sklearn的一个回归算法对一份数据集进行机器学习，实现对于结果的预测。完整的流程包括选定模型，训练模型，特征工程，评估模型，进行预测。

1. **设计思路与方法**

* 观察数据集

选定数据集为Sign Language Mnist，其中已经划分好训练集与测试集，并无验证集，是一个手势的数据集，数据集格式是csv形式，其中包括像素值和标签。这是一个多分类问题；数据中有标签则使用监督学习算法，。

* 选用模型

对于该数据集，显然使用线性回归是不可行的，可以采用逻辑回归后进行分类，也可以选用多个弱分类器进行投票，如Adaboost。本实验中选用一个简单的逻辑回归算法作为基线模型，并开始进行训练。一般对于较为复杂的数据集需要使用多个模型多次尝试去找较优的几个算法。

* 评估模型

如果基线模型效果符合预期，可以继续使用；若效果太差需要进行分析，是否需要更换模型，或者仅仅是因为未进行特征处理。在这里采用了准确率，混淆矩阵与分类报告作为评估指数。选出合适的模型算法。

* 特征工程

选定初始的模型后，需要提升基线模型的预测准确率，进行特征工程进行特征的构建。在这个数据集中我仅使用了方向梯度直方图构建了新特征就能获得不错的结果。

* 参数调优

在特征工程结束后，需要对模型算法进行超参数的调整，将默认值调整为适合数据集的参数，可以确定一个参数范围后设置循环，在评估模型过程中保留下最优参数作为最终的参数。

* 确定最终的模型，训练、评估、预测

在特征工程与训练并行过程中确定了最终使用的模型与寻找到的最有用的特征，并训练出最终的模型，训练完成后进行模型的评估，若有多个模型则需要选出最好的一个或者最好的几个进行融合，得到最佳模型后使用测试集进行预测数据。在本实验中未使用模型融合进一步提升预测准确率，只使用了单模型也能得到较高的准确率。

1. **流程图**

