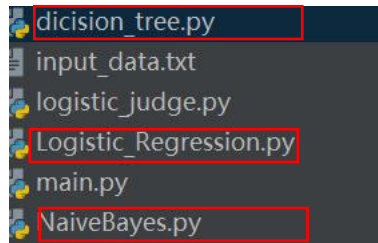


本实验在 `pyspark` 环境下进行，通过 `pycharm` 进行编程  
`data` 文件夹为训练样本集，数据集需要规模大于 100 行，具体表头内容应为  
`id,features,class` 的形式

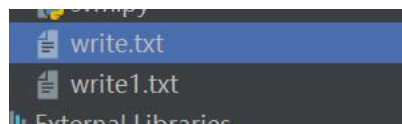
`input_data` 为需要选择分类方法的数据集 `D`

算法集有三个算法，分别为逻辑回归，朴素贝叶斯，决策树算法。



`logistic_judge` 为分类 `D` 所选择的合适算法

`write.txt` 和 `write1.txt` 为中间过程文件。



`assusement.py` 文件表示对数据集 `D` 的处理，结果写入 `write1` 文件。