SpaceshipTitanic问题

此处对问题进行翻译和修饰

可视化

特征工程和可视化

（1）Spent特征，四个花销：找出最高的和总消费

（2）缺失值处理：缺失值作为onehot的新值、numeric的用mean

（3）cabin特征：cabin的raw数据指向了一串字母和数字，用斜杠隔开，需要人工区分开之后以cabinA、cabinB、cabinC进行区分

（4）surname和兄弟姐妹数量：surname数量和兄弟姐妹是否一致——也即兄弟姐妹数量是否全在ship上

训练模型选择

选择GBDT，用GridSearch进行超参调整。

最终参数和结果