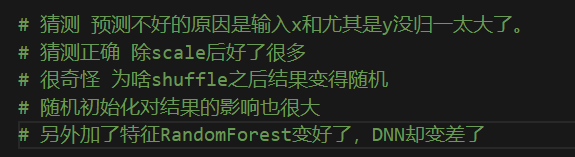
HW12 实验报告

最终结果

自己探索：

深度学习的探索



特征处理的探索：

文本

描述已自动生成

最好正好0.349几

抄的树洞 #5101417：

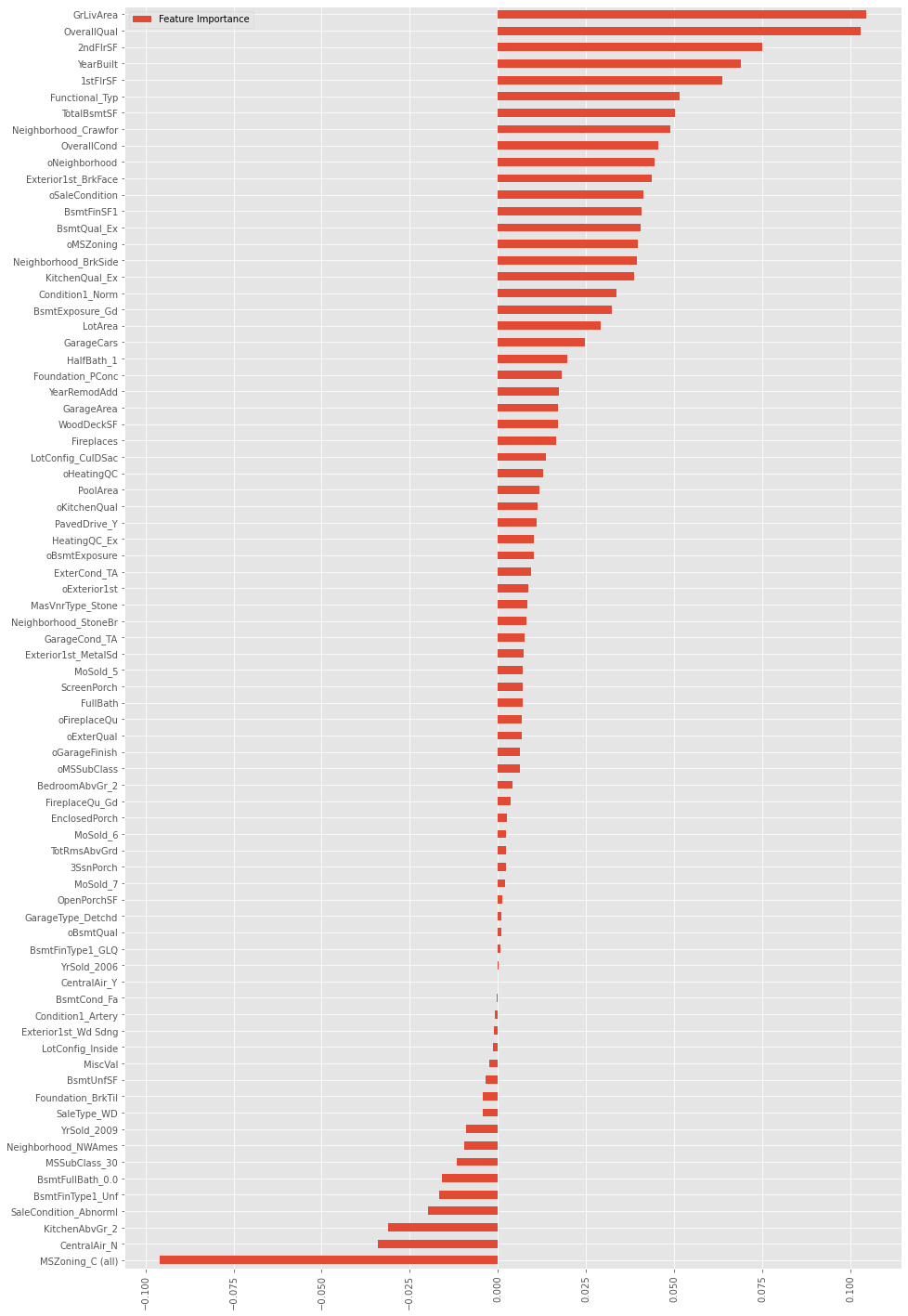
文本

描述已自动生成

在这基础上试了很多，只有把BOROUGH换成one-hot能上升一点点点点，0.3122，但几乎不怎么影响，就算了。

（另外改了一点点逻辑，树洞上数据处理时用了test真值，数据映射时我只用了训练集，性能几乎没掉，证明树洞的这个映射方法真的是有效的：最关键的Neighborhood统计均值，将200多个分类变为一个连续属性）

Lasso得到的特征相关度：



Anyway，挫败感非常强啊。

非常的困惑：

1、为什么深度学习的效果远不如RandomForest、XGBRegressor、ExtraTreesRegressor等机器学习算法。

2、为什么集成起来的机器学习算法能提高效果，怎么AverageWeight。

3、RobustScaler是什么，怎么对数据缩放了。为什么还取log最后exp放缩居然不会炸（因为深度学习取log与exp炸了，我神经网络的归一最后直接除100000）

只能说这两次作业是一个pandas库练习。