Note：绿色字体为该步骤的工作内容是什么；红色字体代表补充说明

一：模型训练（针对train\_data）

1. 对train\_data进行one-hot/递进编码处理，得到train\_data1
2. 对train\_data1进行Imputation，得到train\_data2
3. 将train\_data2中SaleCondition及SalePrice列去除，使用PCA，得到PCA模型
4. 取出train\_data2中SaleCondition为Normal的obserbations，然后去除SaleCondition及SalePrice列，然后用已得的PCA模型进行降维，然后将SalePrice列加入，得到train\_data3。使用train\_data3训练模型（模型待定），得到模型A
5. 循环过程：

每个循环中取出train\_data2中SaleCondiation的一类的observations，然后去除SaleCondition及SalePrice列，然后用已得的PCA模型进行降维，得到train\_data4，用模型A对train\_data4进行predict，然后将真实SalePrice与predict得到的price相除，得到一个比值，观察该比值是否有个集中趋势，从而确定该类型SaleCondiation的系数的估计alpha（具体如何确定系数的估计视出来的结果定）

二：检验模型效果（针对test\_data）

1. 对数据进行one-hot/递进编码处理，得到test\_data1
2. 对test\_data1进行Imputation，得到test\_data2
3. 预测：将test\_data2中的SaleCondition列去除，用已得的PCA模型降维，然后使用模型A预测，将预测值乘以对应的SaleCondition的系数（Normal为1），得到最终预测价格（test\_data中没有SalePrice，所以应该是需要将预测结果上传到kaggle，kaggle给出指标）