说明：

由于星期几以及区域均为类别变量，这里需要对其进行哑变量处理。考虑数据均衡问题，0、1分类比为2.7:1，因此不进行均衡化处理。

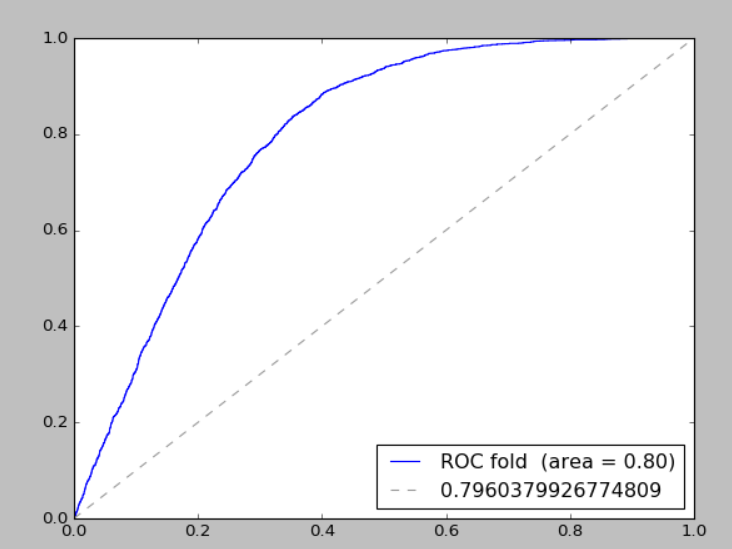
将训练样本拆分训练和测试样本8:2，分别考虑使用逻辑回归模型和xgboost模型来对训练样本进行拟合，并测试模型在测试集上的效果。

lr对应auc为0.796，xgboost对应为0.7945。对应roc曲线如下图。

lr略优于xgboost，这里xgboost的参数选择为：

最大深度3，样本采样系数0.8，列采样0.65，学习速率0.025，最大迭代次数1000。

使用lr来对为止样本进行分类预测，得到结果，这里只保留预测为1的概率值。



最终得到的auc为0.796，并不是十分理想，其中xgboost的参数并没有得到充分的调参处理，下一步会研究一下，gridsearchcv等方法进一步调参优化。