数据预处理

首先我们读取了NFLIS\_DATA和美国社会经济数据的数据，为了使其能够矩阵化，我们首先清洗了非数字值，并根据其含义填入了恰当的数字值（‘（X）’及’\*\*’,’\*\*\*\*\*’等字符串值进行了清0，而表现增长使用了数字1，表现下降使用了数字-1）。对于误差较大或者不需要的数据进行了删除。对于维度不相同的数据，进行了padding操作。

总结

问题要求我们给出药物事件增长的原因的分析，为减少药物事件提出建议，我们在进行了数据分析和建模之后，成功的完成了任务。首先，使用了一个简单的神经网络对2010年到2017年各地药物事件数进行建模分析，找出了OH州拥有有利于该州毒品事件数增长的因子，然后再通过统计分析，得到了一个可视化模型支持我们的结论，并且找出了该州毒品事件数剧烈增长是由于其部分城市的毒品事件数剧烈增长是，而不是一个州的毒品事件数的普遍增长。我们再使用2010年至2017年美国社会经济数据构建了我们第二个神经网络模型，该模型具有良好的非线性表达能力，但由于层数较深，无法通过分析权值来解释模型，所以使用了可视化模型进行辅助分析，找出了可能对毒品事件数有明显效果的社会经济因素。然后根据该模型，对毒品事件较为猖獗的OH州CUYAHOGA县进行分析，并以减少毒品事件数为目标给出了对社会经济因素的建议。我们使用神经网络模型给出的建议和可视化数据模型进行了比对，得到了一致的建议，并完成了相互验证。

模型的优势

我们的模型使用了神经网络的结构，对模型的表达性优良，能够给各个county提出较为准确地建议，也能够根据各个county的数据预测未来的发展。

模型的劣势

由于使用了神经网络结构，对模型的解释性很低，深层次的内容无法做进一步分析，也不清楚各层网络是学习到了什么。所以我们会依赖具体案例分析和可视化数据模型进行一个辅助分析。