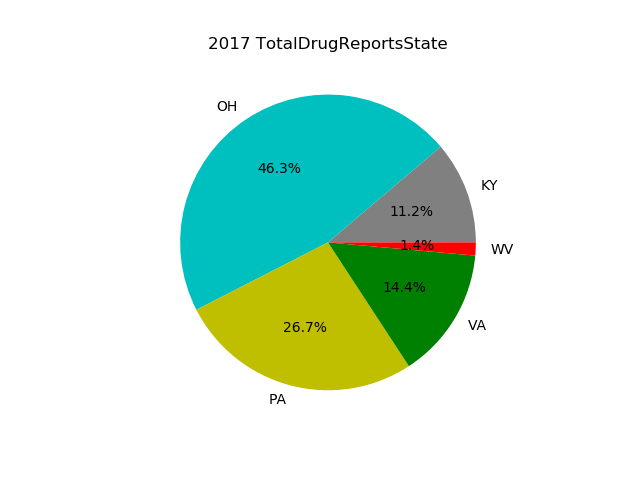
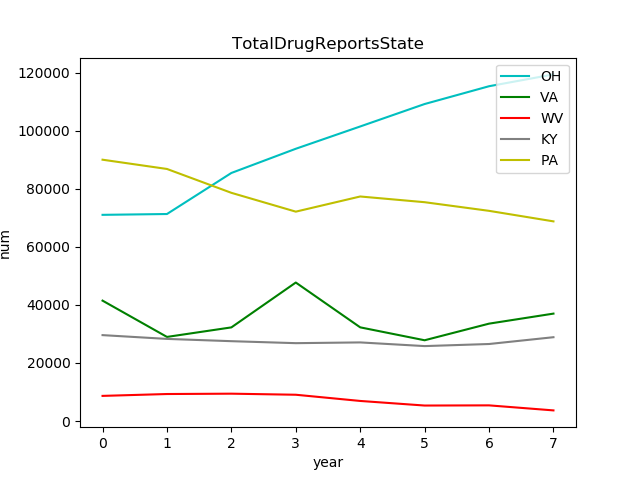


图[1] 2010年各州TotalDrugReportsState占比数饼图



图[2] 2017年各州TotalDrugReportsState占比数饼图

根据图[1]和图[2]，我们发现这7年，OH州药物事件占比数猛增，各州药物事件占比数变化很大，而如果各州影响因子均相同且独立（即五个州同时符合某一线性或非线性的变化），这一药物事件占比数的饼图中，占比数从大到小的排名位次不应该发生改变。



图[3] 各州TotalDrugReportsState量随时间变化图

图[3]中，我们能够更明显看出这一变化，造成变化的因子使这几个州的毒品事件数发生改变。而影响各州的这一变化因子可能包括人口原因，经济原因，地理原因，社会组成原因，政策原因等。由于我们使用的数据集MCM\_NFLIS\_Data中并未包含这些原因，因此我们把这些原因带来的影响使用超参数h来表达。

根据这个数据集，我们能够用来建模的数据只有：

DrugReports TotalDrugReportsCounty TotalDrugReportsState

我们把它们命名为x1,x2,x3。

超参数h作用于x3上，使得:next(x3) = h(x3)。

我们假设：

对于每个CITY当年的DrugReports（我们把它设为y），它前一年的DrugReports，所在COUNTY前一年的DrugReports，所在STATE前一年的DrugReports都对其有作用，这一作用因子我们设为w。

我们可以得出这样一个式子:

y = w\*x + b(b为偏置项bias)

这里超参数h产生的效果我们可以近似于w3,b3。

根据我们上面的三张图可以知道，由于h对于每个州产生的效果不同，我们把x3删去，增加新的x3,x4,x5,x6,x7,作为每个州的 TotalDrugReportsState放入我们的式子中，此时h的作用被等效于w3,w4,w5,w6,w7,b3,b4,b5,b6,b7。

为了引用非线性关系，我们引入tanh作为激活函数，并在后面再增加一个激励因子s，s的作用是控制我们的映射关系落在tanh函数的线性区还是非线性区。

x1: 前一年的DrugReports

x2: 前一年的TotalDrugReportsCounty

x3~x7:当前state前一年的TotalDrugReportsState，若不是当前state补0

y:该county当年的DrugReports

w,b:权重和偏置项

s:激励因子

我们使用了一个浅层的神经网络去训练这个式子，使用了随机梯度下降优化器作为优化器。