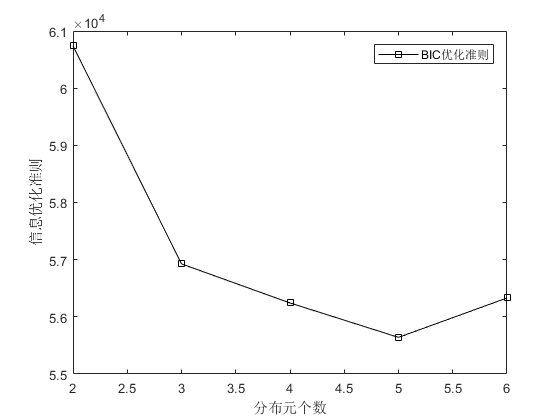
## 仿真及实验结果分析

### 点模式特征分布的分布元信息对性能的影响

某类点模式模型的基数分布服从泊松分布，特征分布服从混合高斯分布，参数为;;,,;,,。通过该模型生成140组点模式数据，将其中的100组数据作为训练集，剩余的40组作为测试集。

图表, 散点图

描述已自动生成

图（一） 图（二）

图（一）为该模型的真实分布，我们通过结合BIC信息准则的参数学习算法对100组训练数据进行参数学习.对分布元个数从2个递增至6个分别进行学习，并且以BIC信息准则作为判断依据，从图（二）的BIC信息准则曲线得，分布元为5时得到最小的BIC值。我们选取所有学习得的模型中分布元为5作为训练集所对应的模型。

我们用该模型的40组测试集数据对BIC信息准则的结果进行验证，分别通过Gibbs采样得的三分布元的参数、四分布元的参数、五分布元的参数和六分布元参数构建似然4个函数。将每个训练集分别代入4个似然函数，若基于BIC信息准则和Gibbs采样的有限混合模型学习算法所得结果合理，那么测试集数据在5分布元处取到最大值的比列应为最高。通过图（三）的实验结果，40组数据中有38组在5分布元处取到最值，说明5分布元的混合模型最符合样本数据，且基于BIC信息准则和Gibbs采样的有限混合模型学习算法的学习准确率高。

图表, 条形图

描述已自动生成 图表

描述已自动生成

图（三） 图（四）

图表, 散点图

描述已自动生成 图表, 条形图

描述已自动生成

图（五） 图（六）

### 点模式数据集分类实验

#### 实验一

第一类基数分布服从泊松分布，参数为，特征分布为高斯混合模型，参数为；，,;, , 。

第二类基数分布服从泊松分布，参数为，特征分布为高斯混合模型，参数为；，,;, , 。

第三类基数分布服从泊松分布，参数为，特征分布为高斯混合模型，参数为；，,;, , 。

每一类生成140个样本，其中100个用作模型学习，40个用作测试学习效果；从图（五）可以看出三类样本在特征空间中重叠严重，单纯使用点模式的特征信息进行分类，无法得到效果较为不错的分类结果；再从图（四）分析，三类点模式在基数分布上存在部分重叠现象，结合基数信息理论上可以提高分类效果。通过实验效果得，我们从图（六）可以看出，结合基数信息的Poisson模型相对于单纯使用特征分布信息的NB模型，在分类效果上存在显著的提升。

#### 实验二

第一类基数分布服从泊松分布，参数为，特征分布为高斯混合模型，参数为；，,;, , 。

第二类基数分布服从泊松分布，参数为，特征分布为高斯混合模型，参数为；，,;, , 。

第三类基数分布服从泊松分布，参数为，特征分布为高斯混合模型，参数为；，,;, , 。

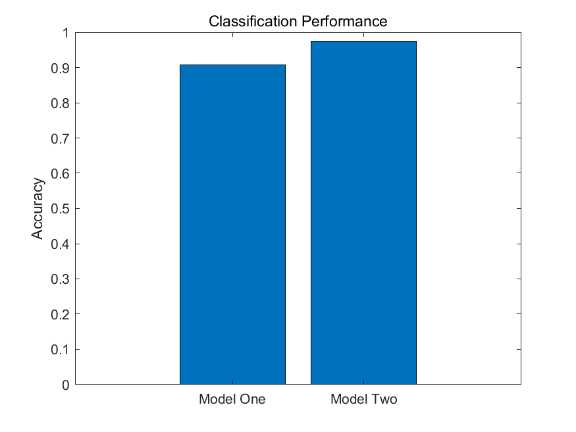
图表, 直方图

描述已自动生成 图表, 散点图

描述已自动生成

图（七） 图（八）

图表, 条形图

描述已自动生成 

图（九） 图（十）

每一类生成140个样本，其中100个用作模型学习，40个用作测试学习效果；从图（七）中可以看出类一与类二在基数信息上重叠，类三可以借助基数分布信息提高自身分类效率；

从图（八）中可以看出，类二和类三在特征分布信息上重叠，类一可以通过特征信息提高自身的分类效果，从图（九）的分类结果中显示结合基数信息的Poisson模型相对于单纯使用特征分布信息的NB模型在分类任务中效果更好。

图（十）中Model Two三类样本的特征分布均选取BIC信息最小混合模型，而Model One除了类二的特征分布，其余两类我们均选取BIC信息最小混合模型，进行对比实验得，Model Two的分类效果优于Model One，由此可以说明分布元信息的确影响分类效果。

## 3 泊松点过程模型的参数估计

学习点过程模型的计算高效算法很重要，因为机器学习与空间统计应用相比通常涉及大数据集，由于学习一般的点过程模型在计算上是非常困难的，而泊松点过程模型通过忽略点之间的相关性，在模型的通用性和模型参数的学习难度之间取得了一个良好的平衡。

泊松点过程模型由它的基数分布和特征分布共同决定，模型的参数表示形式如下

 （3-1）

其中和分别由参数和所决定的基数分布和特征分布，的函数形式为泊松分布，由于高斯混合模型在数据拟合方面有着不俗的表现，所以的函数形式为高斯混合分布，对学习泊松点过程模型的学习过程相当于通过训练集估计参数。

从泊松点过程模型的似然函数形式中我们不难发现，对整体模型参数极大似然估计等价于分别对基数分布参数和特征分布参数进行极大似然估计，估计过程如下：

假设，……，是个独立同分布由某一个泊松点过程模型生成的样本数据集，因此样本数据集关于参数的似然函数如下：

 （3-2）

从似然函数的形式中可以得出，最大化等价于分别最大化似然函数的第二与第三部分，即

 （3-3）

 （3-4）

### 3.1 基数分布的参数估计

基数分布关于参数的似然函数如下：

 （3-5）

为了便于分析与计算，我们定义基数分布关于参数的对数似然函数，

 （3-6）

 （3-7）

容易证明，使对数函数似然值最大的值也是似然函数最大。对关于求导为零得



 （3-8）

### 3.2 特征分布的参数估计

泊松点过程模型的特征分布为有限高斯混合模型可以表示为如下形式：

 （3-9）

其中，维度为维，令；是混合权重且，；是高斯分布密度，，

 （3-10）

称为第个分模型；特征分布的参数。那么特征分布似然函数可以表示为：

 （3-11）

根据式（3-11），高斯混合模型的后延分布可以描述为：

 （3-12）

 （3-13）

其中是正则常量。对于每一个训练样本中的元素,需要定义一个维的的指示变量，表示该元素由哪个分布元产生，称为缺失变量，表示形式如下：



 （3-14）

#### 3.2.1 高斯混合模型的参数先验分布

实现公式（3-12），优先需要考虑如何获取先验分布，在高斯混合分布的情况下，先验参数为