P1、各位老师好 我的研究课题是基于模型的点模式学习算法研究

P2、点模式是点或特征的集合或多集合的形式，自然界中许多现象我们都可以用点模式的形式描述，比如药物的分子结构（化学），城市地理布图（地理学），天体的空间分布(天文学)等。而在数据科学中，点模式通常称为“包”（多示例），用于描述对象的特征信息。例如：图像的关键点，词袋，稀疏数据集等。因此，针对点模式对象的学习算法在本质上就是基于多示例的学习算法。

P3、自20 世纪90年代，Dietterich等在对药物活性预测问题的研究中，首次提出多示例学习的概念之后，研究学者提出了许多基于不同使用背景的多示例学习假设和多示例学习算法，大致可以分为基于示例空间、包空间、嵌入空间的多示例算法，但是基于模型的点模式（多示例）学习算法并未受到足够的关注。

P4、我的研究内容打算从以下三点展开：基于MLE和BIC的随机点模式模型学习算法；基于点模式混合模型的无监督学习算法；结构点模式。

P5、独立同分布聚类模型主要由两部分组成：点模式的基数分布，用来描述点模式的基数信息；点模式的特征分布，用来描述点模式的特征信息。当基数分布为泊松分布时，即可得到泊松点过程模型。

P6、基于Gibbs采样的有限混合模型参数学习算法：使用Gibbs采样的方法，逐步收敛于混合模型的参数后验分布，当有限混合模型为高斯混合模型时，算法流程如下

P7、很多参数估计问题都采用似然函数作为目标函数，当训练数据足够多时，以提高模型复杂度为代价可以不断提高模型精度，但会带来一个机器学习中非常普遍的问题——过拟合。所以，模型选择问题必须在模型复杂度与模型对数据集描述能力（即似然函数）之间寻求最佳平衡。贝叶斯信息准则就是通过加入模型复杂度的惩罚项来避免过拟合问题。

(贝叶斯信息准则通过加入复杂度的惩罚项，来平衡模型复杂度与模型对数据集描述能力)

P8、研究内容一针对基于模型的点模式分类问题，提出一种结合模型复杂度估计的模型参数学习方法，首先构建n个模型复杂度不同的独立同分布聚类模型，对右上角训练集关于模型参数做最大似估计，我们发现可以转化为分别关于基数分布参数和特征分布参数做最大似然估计。在对特征分布参数学习时我们用Gibbs参数采样代替最大似然估计算法。模型学习结束后，通过BIC信息准则从备选模型中取得最优模型。

P9、最优模型的性能验证实验：图一图二分别是某个点模式模型生成的140组样本的基数分布和特征分布，其中100组为训练集，40组为测试集。根据图三的BIC曲线，算法选择特征分布5元的为最优模型，从图四训练集数据关于备选模型的对数似然函数的曲线可以看出，曲线后半段对数似然函数随模型复杂度增大提升效果不明显，说明当模型复杂达到一定程度（特征分布元达到5个），即可保证模型的数据拟合能力。·图五，根据每个测试集数据关于备选模型的似然函数值，对备选模型进行投票，结果保证了最优模型的泛化能力。

P10、对比结合复杂度估计和主观判断模型复杂的两种情况，当分布元估计不一致时，本文算法可以一定程度减弱主观判断对分类结果的影响，保证一个较好的分类性能。

P11、研究内容二，建立IID-cluster混合模型，引入基于包空间水平和示例空间水平的缺失变量，尝试将基于示例水平模型的EM算法扩展到基于包水平模型。

P12、研究内容三，通过概率无向图模型来描述点模式各点间的相关性，根据Hammersley-Clifford定理，尝试用无向图中最大团的能量函数描述概率无向图的联合概率。

P13、这是我课题的时间安排 P14、以及目前的研究进程

P15 16、以下是部分参考文献 P17 感谢各位老师的临听，如果有问题请各位老师指正