概率图模型的一些提问与回答

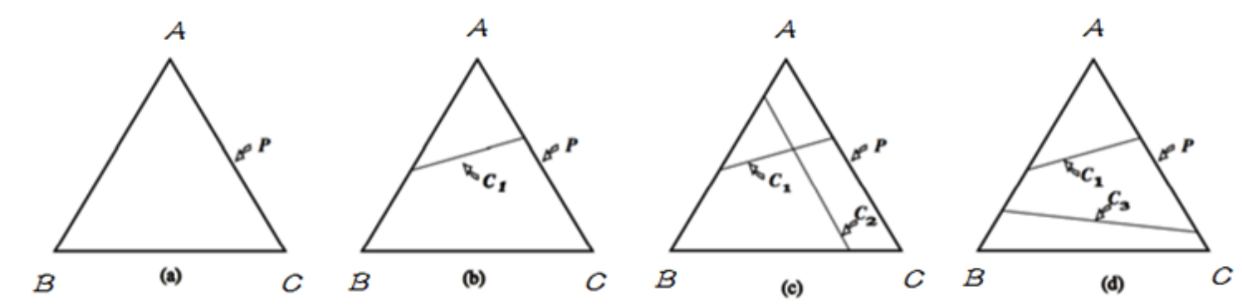
吴锐 李林 李震

**最大熵**

熵增原理：熵越大模型越稳定。

最大熵原理认为：要选择的概率模型首先必须要满足已有的事实，即约束条件。

1. 这个约束条件在最大熵模型中可以利用特征函数关于经验分布的期望与特征函数关于模型P(Y|X)与经验分布的期望值相等来建立约束条件，然后求熵最大。
2. 在这里求熵最大，一般思路是利用求导，但是求导不能求解有约束的最值问题，所以利用对偶，将有约束问题转化为无约束问题，再进行求解。



1. 特征函数

1 ， 当 厶 黟 满 足 某 一 事 实 ， 
0 ， 不 满 足 该 事 实 ． 

1. 约束条件
2. 优缺点：

最大熵模型的优点有：

* + 1. 最大熵统计模型获得的是所有满足约束条件的模型中信息熵极大的模型,作为经典的分类模型时准确率较高。
    2. 可以灵活地设置约束条件，通过约束条件的多少可以调节模型对未知数据的适应度和对已知数据的拟合程度

最大熵模型的缺点有：

* + 1. 由于约束函数数量和样本数目有关系，导致迭代过程计算量巨大，实际应用比较难。

**HMM**

1. 隐马尔科夫的隐代表什么

指：状态之间的转移过程不可观察

HMM有什么缺陷：

定义观察序列和标记序列的联合分布意味着所有可能的观察序列必须是可枚举的，如果在观察序列中存在长距离依赖，这个任务将是十分困难的。因此，为了便于问题的处理，生成模型必须要给出一个严格的独立性假设。例如在隐马尔可夫模型中，我们假设 t 时刻的观察值只依赖于 t 时刻的状态，这确保了每个观察值被作为独立的单元对待，并且独立于序列中所有其他观察值。但是事实上，绝大多数的数据序列并不能完全地表示为一组独立的单元。当序列中的数据元素存在长距离依赖时，允许这种长距离依赖并且使观察序列可以表示为非独立的交叉特征的模型才是比较合适的。

**MEMM：**

条件模型的提出克服了生成模型所要求的严格的独立假设，它定义了一个在给定观察序列 x 的条件下，状态序列 s 的条件分布 p(s|x)

1. HMM与MEMM有什么不同？
   1. HMM是根据联合概率P(S,O)来求P(S|O)
   2. MEMM将观测序列当做是条件序列，而不是由状态生成的。
2. MEMM为什么比HMM与crf的计算效率要高？
3. MEMM存在什么问题?

最大熵马尔可夫模型和其他“判别式”有限状态模型一样，在特定情形下都存在一个共同的标注偏置问题。

标记偏置问题会对自然语言的处理结果产生不良影响。例如，在词性标记任务中，有一个位置的词总是对应于某个特定的词性标记。当某个句子中该词的词性是其很少使用的一种词性时，由于标记偏置现象，该词将被标注为更为常用的那个标记，而并不考虑余下句子中的上下文信息，这是“判别式”模型的固有缺陷

**CRF**

1. HMM与crf中的序列标注问题有什么区别和联系？
2. CRF中的矩阵形式中的M与HMM中的A有什么区别和联系？
   1. 在HMM的矩阵A中行的和为1，M没有要求。
   2. A中的状态转移概率不随位置t的变化而变化，M会随着t的变化而变化，即M由多个组成，每一个状态转移对应一个，这样M的参数会更多，模型会更加灵活。
3. 结构化感知机和CRF，在做序列标注的时候有什么异同？

相同点：1.特征函数相同。2权重向量相同。3打分函数相同。4预测算法相同。5同数据结构化学习。

不同点：感知机算法属于在线学习：每次参数更新只使用一个训练实例。由于没有考虑到整个数据集，所以在线学习难免顾此失彼。而CRF对数释然函数及其梯度定义在整个数据集之上，每次参数更新都是全盘考虑。