统计学习读书报告

09017244 郑健雄

1. 自己提出的问题的理解：
2. T算法和S算法之间的联系是什么？T算法是否是S算法的一个特例？

讨论后的理解：

T算法是S算法的一个特例，其是为了兼具效率与S算法的简洁性来提出的，它实际上是将S值压到了最小值的S算法。

1. P229公式11.42是如何转变成该形式的？

讨论后的理解：

该公式的推导使用到了P226的公式11.35，其推导的关键是将P(Y|X)进行分解，使其对应到某个位置的概率而不是一个总体的概率，就可以转化为11.42的形式了。

1. 别人提出的问题的理解
2. P229页上的松弛特征该如何理解？为什么可以通过这种方法解决特征数不一致的问题。

自己的理解:

这个问题可以从公式11.37中找到答案。比较11.41，11.40与11.37可以看出来T(x,y)是受到求和符号下标制约的，如果直接使用11.37的形式来求解，那么问题就是得计算所有数据的T(x,y)，这样的话效率会很差。而S算法的T(x,y)默认为常数，一次就能计算出结果。

而松弛应该是因为使用了最大值作为梯度的分母，所以要比直接求解值得梯度更小一点，整体的梯度被松弛了。

1. P231页BFGS算法中的正定对称矩阵B0具体是怎么选取的？

自己的理解：这个问题是牛顿法的初值选择问题。一般情况下，如果没有特殊的限制，应该进行随机选取。

1. CRF的维特比算法类似于HMM。但是有一点疑问：不同于HMM，CRF需要概率归一化，也就是说一个CRF模型在建立的过程中就经历了归一化过程，已经算出了各种y的序列的非规范化概率，在这种情况下为什么不直接比较各序列的概率大小而是采用维特比算法呢？

自己的理解：

我自己的理解是这和HMM类似，还是一个效率的问题，所有的y序列的概率计算是非常慢的，而维持比算法使用了动态规划的思想，这样就只计算了必须计算的部分，并且记录了最优的路径，总体而言效率更高。

1. 算法T是如何解决算法S收敛速度慢的问题的？

自己的理解:

该问题还是探讨到了S算法和T算法的关系，S算法的收敛速度慢是因为其让更新的梯度值变得很小，速度很慢。而T算法考虑到了这个问题，限制了S的值，使其不会太大，这样就能提升计算的效率。

1. 读书计划

1、本周完成的内容章节：17-18章

2、下周计划：19章

四、读书摘要及理解

1. 条件随机场的学习问题:

该模型实际上是对数线性模型，可以使用IIS算法和拟牛顿法。

IIS算法通过优化对数似然函数来求解问题，使用CRF自身的公式，给出了11.37的更新方程。而考虑到不同数据的特征数不匹配的问题，提出了改进的S算法，其将特征数固定为一个常数，从而使得计算变得简单。而S算法又有着收敛速度非常慢的问题，所以又提出了其特例T算法，T算法将S值限定到最小，从而提高了求解问题的速度。

另一种学习方法就是常用的拟牛顿法。

1. 条件随机场的预测算法:

预测算法实际上和HMM一样，是为了解决标注问题提出的，其与HMM有着非常类似的结构，所以使用了和HMM一样的维持比算法，其思路也是计算每一步最大可能的序列并且记录路径点，然后一直计算到结束，反向得出最优路径。

其特征在于仅考虑了非规范化概率，而不用计算规范化的概率，因为只是需要比大小，所以这样做可以提升计算的效率并保证结果正确。

思考:CRF虽然与HMM并不是同一个模型，但是其有着类似的结构，并且都可以用于处理标注问题和顺序问题。此外，CRF模型实际上也是一种最大熵模型，所以其可以使用最大熵模型的算法进行学习，而因为其与HMM结构上的相似性，在进行预测算法学习时，也可以使用维持比算法，提升计算的效率。CRF可以被看作是最大熵马尔可夫模型在标注问题上的推广，可以更好的处理标注问题。