方骏-2020年8月9日-读书报告

1. 自己提出的问题的理解：
2. 别人提出的问题的理解：
3. 读书计划

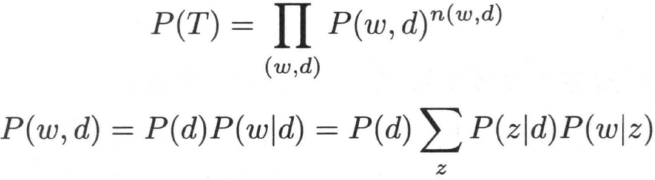
1、本周完成的内容章节：《统计机器学习》第十八章

2、下周计划：《统计机器学习》第十九章

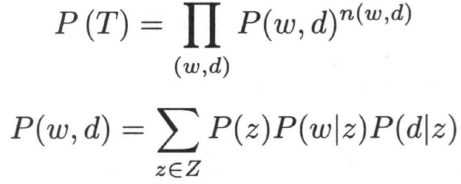
四、读书摘要及理解

1、概率潜在语义分析是利用概率生成模型对文本集合进行话题分析的方法。概率潜在语义分析受潜在语义分析的启发提出，两者可以通过矩阵分解关联起来。概率潜在语义分析的模型有生成模型，以及等价的共现模型。其学习策略是观测数据的极大似然估计，其学习算法是EM 算法。

2、生成模型表示文本生成话题，话题生成单词，从而得到单词-文本共现数据的过程:假设每个文本由一个话题分布决定，每个话题由一个单词分布决定。单词变量ω与文本变量d 是观测变量话题变量z 是隐变量。生成模型的定义如下：



3、共现模型描述文本单词共现数据拥有的模式。共现模型的定义如下：



4、概率潜在语义分析的学习通常采用EM算法。通过迭代学习模型的参数，和，而可直接统计得出。