第十二周读书报告 5.11-5.17

61518426 周之遥

·读书进度

学习统计学习方法第18章：概率潜在语义分析。

概率潜在语义分析（PLSA）可以通过矩阵分解与LDA建立联系。

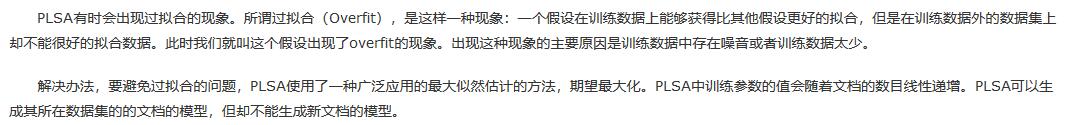
PLSA有生成模型和共现模型，学习策略是极大似然估计的EM算法。

·提出问题

1. 生成模型和共现模型有什么区别和联系？应该如何选择使用哪一个模型？

解答：两种模型是等价的，只是表现形式不同。由不同的条件概率组成。这就导致生成模型是文本到话题到单词。而贡献模型是由话题到单词，话题到文本。学习思路应该就是导致对应的生成关系不同。

1. PLSA为什么可以防止过拟合？



·下周计划

完成19章马尔科夫链蒙特卡罗法。