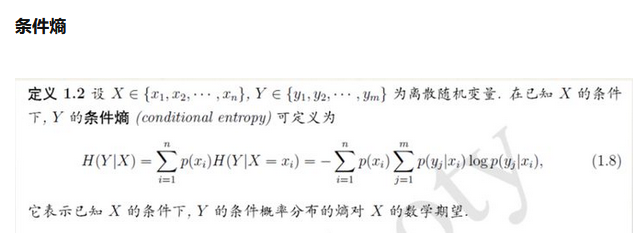
基础概念

拉格朗日乘子法，用于将带约束的极值问题转化为无约束极值问题。

熵(entropy)是热力学中的概念,由香浓引入到信息论中。在信息论和概率统计中,熵用来表示随机变量不确定性的度量。

H(x)依赖于X的分布,而与X的具体值无关。H(X)越大,表示X的不确定性越大。



还是使用经验分布对概率进行近似

重要约束：特征f的期望应该和从训练数据中得到的特征期望相同。

从训练数据中抽取了n个特征,相应的便有n个特征函数以及n个约束条件

主要思路：

最大熵模型的学习过程就是求解最大熵模型的过程。求解约束最优化问题所得的解就是最大熵模型学习的解。

利用拉格朗日乘子法将最大熵模型由一个带约束的最优化问题转化为一个与之等价的无约束的最优化问题,它是一个min max问题。

因为按公式无法得到解析解，所以只能用梯度下降来做。

参考了

李航. 统计学习方法

吴军. 数学之美

https://zhuanlan.zhihu.com/p/29978153