

Probability Graph 01 Background

Chen Gong

23 November 2019

机器学习的重要思想就是，对已有的数据进行分析，然后对未知数据来进行预判或者预测等。这里的图和我们之前学习的数据结构中的图有点不太一样，俗话说有图有真相，这里的图是将概率的特征引入到图中，方便我们进行直观分析。

1 概率的基本性质

我们假设现在有一组高维随机变量， $p(x_1, x_2, \dots, x_n)$ ，它有两个非常基本的概率，也就是条件概率和边缘概率。条件概率的描述为 $p(x_i)$ ，条件概率的描述为 $p(x_j|x_i)$ 。

同时，根据这两个基本的概率，我们可以得到两个基本的运算法则：Sum Rule 和 Product Rule。

Sum Rule: $p(x_1) = \int p(x_1, x_2) dx_2$ 。

Product Rule: $p(x_1, x_2) = p(x_1)p(x_2|x_1) = p(x_2)p(x_1|x_2)$ 。

根据这两个基本的法则，我们可以推出 Chain Rule 和 Bayesian Rule。

Chain Rule:

$$p(x_1, x_2, \dots, x_N) = \prod_{i=1}^N p(x_i|x_1, x_2, \dots, x_{i-1}) \quad (1)$$

Bayesian Rule:

$$p(x_2|x_1) = \frac{p(x_1, x_2)}{p(x_1)} = \frac{p(x_1, x_2)}{\int p(x_1, x_2) dx_2} = \frac{p(x_2|x_1)p(x_1)}{\int p(x_2|x_1)p(x_1) dx_2} \quad (2)$$

2 条件独立性

条件独立性这个词是一个看似好像很熟，实际上一点也不熟的词吧，哈哈哈！我们来想一想，为什么要引入条件独立性，这个很 Fashion 的词呢？

首先，我们想想高维随机变量所遇到的困境，也就是维度高，计算复杂度高。大家想想，当维度较高时，这个 $p(x_1, x_2, \dots, x_N) = \prod_{i=1}^N p(x_i|x_1, x_2, \dots, x_{i-1})$ 肯定会算炸去。所以，我们需要简化运算，之后我们来说我们简化运算的思路。

1. 假设每个维度之间都是相互独立的，那么我们有 $p(x_1, x_2, \dots, x_N) = \prod_{i=1}^N p(x_i)$ 。比如，朴素贝叶斯就是这样的设计思路，也就是 $p(x|y) = \prod_{i=1}^N p(x_i|y)$ 。但是，我们觉得这个假设太强了，实际情况中的依赖比这个要复杂很多。所以我们像放弱一点，增加之间的依赖关系，于是我们提出了马尔科夫性质 (Markov Property)。

2. 假设每个维度之间是符合马尔科夫性质 (Markov Propert) 的。所谓马尔科夫性质就是，对于一个序列 $\{x_1, x_2, \dots, x_N\}$ ，第 i 项仅仅只和第 $i - 1$ 项之间存在依赖关系。用符号的方法我们可以表示为：

$$X_j \perp X_{i+1} | x_i, j < i \quad (3)$$

在 HMM 里面就是这样的齐次马尔可夫假设，但是还是太强了，我们还是要想办法削弱。自然界中经常会出现，序列之间不同的位置上存在依赖关系，因此我们提出了**条件独立性**。

3. 条件独立性：条件独立性假设是概率图的核心概念。它可以大大的简化联合概率分布。而用图我们可以大大的可视化表达条件独立性。我们可以描述为：

$$X_A \perp X_B | X_C \quad (4)$$

而 X_A, X_B, X_C 是变量的集合，彼此之间互不相交。

3 概率图算法分类

概率图的算法大致可以分为三类，也就是，表示 (Representation)，推断 (Inference) 和学习 (Learning)。

3.1 Representation

知识表示的方法，可以分为有向图，Bayesian Network；和无向图，Markov Network，这两种图通常用来处理变量离散的情况。对于连续性的变量，我们通常采用高斯图，同时可以衍生出，Gaussian Bayesian Network 和 Guassian Markov Network。

3.2 Inference

推断可以分为精准推断和近似推断。所谓推断就是给定已知求概率分布。近似推断中可以分为确定性推断 (变分推断) 和随机推断 (MCMC)，MCMC 是基于蒙特卡罗采样的。

3.3 Learning

学习可以分为参数学习和结构学习。在参数学习中，参数可以分为变量数据和非隐数据，我们可以采用有向图或者无向图来解决。而隐变量的求解我们需要使用到 EM 算法，这个 EM 算法在后面的章节会详细推导。而结构学习则是，需要我们知道使用那种图结构更好，比如神经网络中的节点个数，层数等等，也就是现在非常热的 Automate Machine Learning。