

# 第六章 概率图模型

## 1. 概率图模型的联合概率分布

参考书P.118

## 2. 概率图表示

朴素贝叶斯模型的原理：

朴素贝叶斯模型通过预测指定样本属于特定类别的概率 $p(y_i|x)$ 来预测该样本的所属类别

$$p(y_i|x) = p(x_1|y_i) * p(x_2|y_i) * p(x_3|y_i) \dots p(x_m|y_i) * p(y_i)$$

$x = \{x_1, x_2, \dots, x_m\}$ ，假设特征间相互独立

其中 $p(x_1|y_i)$ ， $p(x_2|y_i)$ ， $p(x_3|y_i)$ 以及 $p(y_i)$ 可以通过训练样本统计获得

最大熵模型的原理：

熵是不确定性的度量，熵越大，不确定性也越大。最大熵原理是概率模型学习的一个准则，指导思想是在满足约束条件的模型集合中选取熵最大的模型，即不确定性最大的模型。

## 3. 生成式模型和判别式模型

哪些是生成式模型，哪些是判别式模型？

假设可观测到的变量集合为 $X$ ，需要观测的变量集合为 $Y$ ，其他的变量集合为 $Z$ ，

生成式模型是对联合概率分布 $P(X, Y, Z)$ 进行建模，判别式模型是直接对条件概率分布 $P(Y, Z|X)$ 进行建模

生成式模型：朴素贝叶斯，贝叶斯网络，pLSA, LDA，隐马尔科夫等都是先对联合概率分布进行建模，然后再通过计算边缘概率分布得到对变量的预测

判别式模型：最大熵模型，条件随机场，直接对条件概率分布进行建模

## 4. 马尔科夫模型

假设一个随机过程中， $t_n$ 时刻的状态 $x_n$ 的条件分布，仅仅与其前一个状态 $x_{n-1}$ 有关，即 $P(x_n | x_1, x_2, \dots, x_{n-1}) = P(x_n | x_{n-1})$ ，则将其称为马尔科夫过程，时间和状态的取值都是离散的马尔科夫过程也称为马尔科夫链。

隐马尔科夫模型是对含有未知参数（隐状态）的马尔科夫链进行建模的生成模型

(P.127)。在隐马尔科夫模型中，隐状态 $x_i$ 对于观测者而言是不可见的，观测者只能观测到的只有每个隐状态 $x_i$ 对应的输出 $y_i$ ，而观测状态 $y_i$ 的概率分布仅仅取决于对应的隐状态 $x_i$ 。

在隐马尔科夫模型中，参数包含了隐状态间的转移概率、隐状态套观测状态的输出概率、隐状态 $x$ 的取值空间、观测状态 $y$ 的取值空间以及初始状态的概率分布。

## 5.主题模型

本章主要涉及NLP的技术

常见的主题模型：pLSA, LDA