

# 贝叶斯网络

## 1 半朴素贝叶斯分类器

### 1.1 独依赖性估计 (ODE)

贝叶斯公式需要计算属性的联合概率 $P(c | x)$ ，朴素贝叶斯分类器进行了属性条件独立性假设，但这个假设在现实中往往很难成立。为了解决这一问题，人们尝试对属性条件独立性假设进行松弛，由此产生了**半朴素贝叶斯分类器**。

半朴素贝叶斯的特点是适当考虑一部分属性间的相互依赖信息，最常用的策略是“独依赖性估计” (One – Dependent Estimator, ODE)，即假设每个属性在类别之外最多仅依赖于一个其他属性，即

$$P(c | x) \propto P(c) \prod_{i=1}^d P(x_i | c, pa_i)$$

其中 $pa_i$ 为属性 $x_i$ 所依赖的属性，称为 $x_i$ 的父属性。如果能够确定父属性，则能够根据数据集估计条件概率 $P(x_i | c, pa_i)$ ，然而如何确定父属性是件困难的工作。

### 1.2 超父独依赖估计 (SPODE)

假设所有属性都依赖于同一个属性，称为“超父”。然后根据CV等模型选择方法来确定超父属性。如图所示：

### 1.3 TAN

TAN是基于最大带权生成树的算法，其过程如下：

由于属性条件互信息刻画了属性间的关联，所以TAN实际上通过最大带权生成树仅仅保留了强相关属性之间的依赖性。

### 1.4 AODE

AODE是SPODE算法的集成，通过将每个属性作为超父来构建SPODE，设定一个阈值 $m'$ ，将所有达到阈值的SPODE集成起来：

$$P(c | x) \propto \sum_{i=1, |D_{x_i}| \geq m'}^d P(c) \prod_{i=1}^d P(x_i | c, pa_i)$$

### 1.5 KDE

通过独依赖假设可能获得泛化性能的提升，如果将独依赖ODE扩展为多依赖KDE直观上能进一步提升泛化能力。然而随着 $k$ 的增加，估计条件概率 $P(x_i | y, pa_i)$ 所需要的训练样本数量也将呈指数级增长。因此，如果训练数据不足，很可能又会遭遇高阶联合概率的问题。

## 2 贝叶斯网

贝叶斯网即信念网，是一个有向无环图，用于描述属性间的依赖关系，并使用条件概率表（*CPT*）来描述属性的联合概率分布。其中，贝叶斯网B的