贝叶斯网络

Machine Learning Engineer

机器学习工程师

讲师：加号

稀牛学院 網易 NETEASE

目 录

CONTENTS

01

朴素贝叶斯

1.1 条件概率公式

$$P(A|B)=\frac{P(AB)}{P(B)}$$

1.2 贝叶斯方程

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

朴素贝叶斯的定义

1. 设 $x=\{a_1,a_2,\ldots,a_m\}$ 为一个待分类项，而每个a为x的一个特征属性。
2. 有类别集合 $C=\{y_1,y_2,\ldots,y_n\}$。
3. 计算 $P(y_1|x),P(y_2|x),\ldots,P(y_n|x)$。
4. 如果 $P(y_k|x)=\max\{P(y_1|x),P(y_2|x),\ldots,P(y_n|x)\}$ $x \in y_k$

1.4 朴素贝叶斯的例子

日期	天气	湿度	风级	打球
1	晴	高	弱	否
2	晴	高	强	否
3	阴	高	弱	是
4	雨	高	弱	是
5	雨	正常	弱	是
6	雨	正常	强	否
7	阴	正常	强	是
8	晴	高	弱	否
9	晴	正常	弱	是
10	雨	正常	弱	是
11	晴	正常	强	是
12	阴	高	强	是
13	阴	正常	弱	是
14	雨	高	强	否

要点总结

要点1

贝叶斯公式的定义与理解

要点2

朴素贝叶斯的应用

02 贝叶斯网络与有向分离

2.1 贝叶斯网络

2.2 有向分离

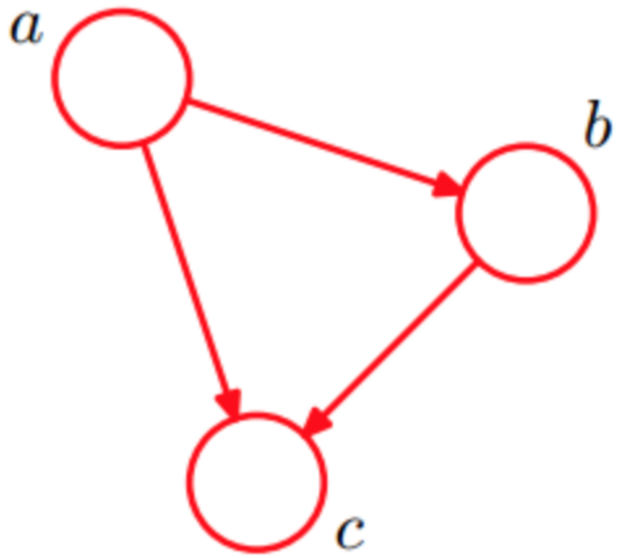
2.1 贝叶斯网络

定义：**贝叶斯网络(Bayesian network)**，又称信念网络(Belief Network)，或有向无环图模型(directed acyclic graphical model)，是一种概率图模型，于1985年由Judea Pearl首先提出。它是一种模拟人类推理过程中因果关系的不确定性处理模型，其网络拓朴结构是一个有向无环图(DAG)。

贝叶斯网络的有向无环图中的节点表示随机变量{x1, x2, x3…xn}，它们可以是可观察到的变量，或隐变量、未知参数等。认为有因果关系（或非条件独立）的变量或命题则用箭头来连接。若两个节点间以一个单箭头连接在一起，表示其中一个节点是“因(parents)”，另一个是“果(children)”，两节点就会产生一个条件概率值。

总而言之，连接两个节点的箭头代表此两个随机变量是具有因果关系，或非条件独立。

2.1 贝叶斯网络



因为a导致b，a和b导致c，所以有:

$$P(a,b,c)=P(c|a,b)*P(b|a)*P(a)$$

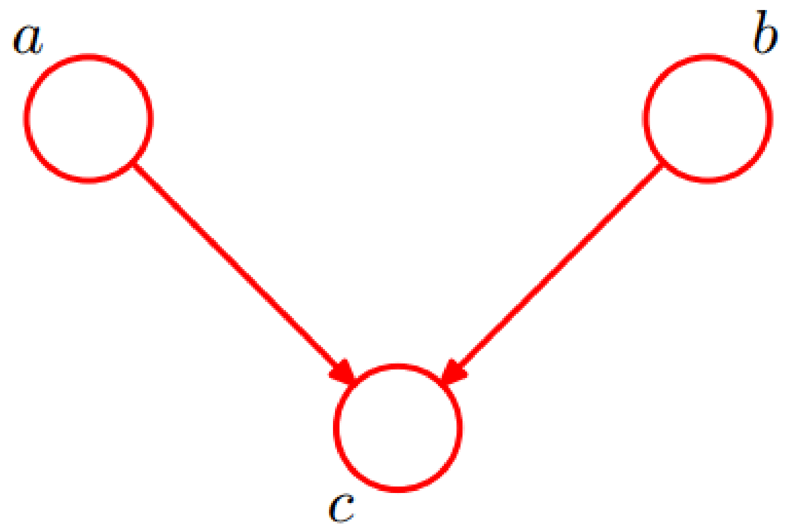
2.2 有向分离

有向分离（D-Separation）是一种用来判断变量是否条件独立的图形化方法。换言之，对于一个DAG(有向无环图)E，D-Separation方法可以快速的判断出两个节点之间是否是条件独立的。

我们根据贝叶斯网络的三种形式来一一讲解：

有向分离

形式1：head-to-head



所以有： $P(a,b,c)=P(a)*P(b)*P(c|a,b)$化简后可得：

$$\sum_c P(a,b,c)=\sum_c P(a)*P(b)*P(c|a,b)$$
$$\Rightarrow P(a,b)=P(a)*P(b)$$

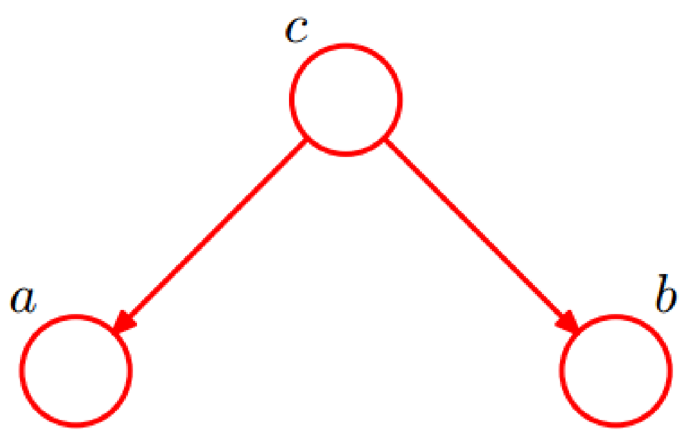
即在c未知的条件下，a、b被阻断(blocked)，是独立的，称之为head-to-head条件独立。

2.2 有向分离

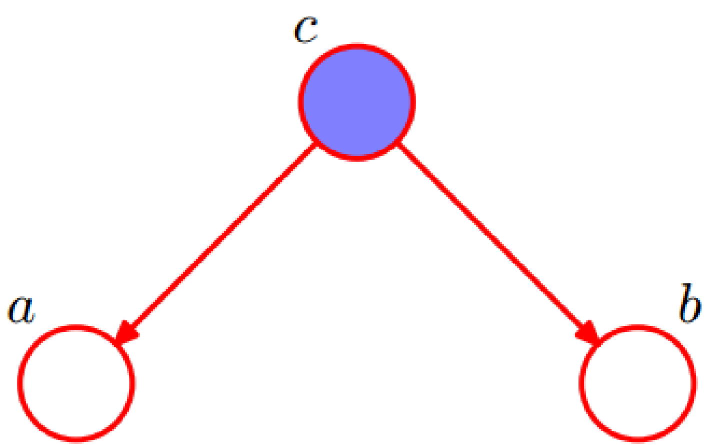
形式2：tail-to-tail

考虑c未知，跟c已知这两种情况：

在c未知的时候，有：$P(a,b,c)=P(c)*P(a|c)*P(b|c)$，此时，没法得出 $P(a,b)=P(a)P(b)$，即c未知时，a、b不独立。



在c已知的时候，有：P(a,b|c)=P(a,b,c)/P(c)，然后将P(a,b,c)=P(c)*P(a|c)*P(b|c)带入式子中，得到：P(a,b|c)=P(a,b,c)/P(c) = P(c)*P(a|c)*P(b|c) / P(c) = P(a|c)*P(b|c)，即c已知时，a、b独立。



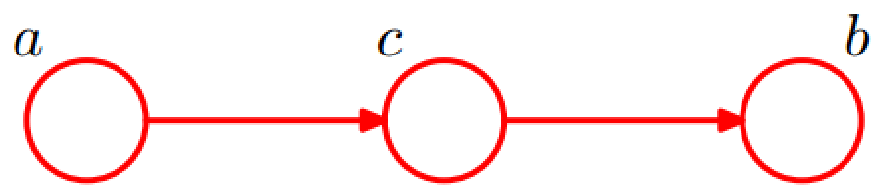
所以，在c给定的条件下，a，b被阻断(blocked)，是独立的，称之为tail-to-tail条件独立。

2.2 有向分离

形式3：head-to-tail

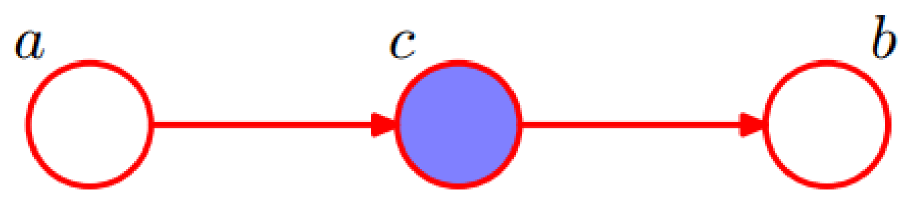
还是分c未知跟c已知这两种情况：

c未知时，有 $P(a,b,c)=P(a)*P(c|a)*P(b|c)$ ，但无法推出P(a,b) = P(a)P(b)，即c未知时，a、b不独立。



c已知时，有：P(a,b|c)=P(a,b,c)/P(c)，且根据P(a,c) = P(a)*P(c|a) = P(c)*P(a|c)，可化简得到：

$$
\begin{aligned}
& P(a,b|c) \\
= & P(a,b,c)/P(c) \\
= & P(a)*P(c|a)*P(b|c) \;/\; P(c) \\
= & P(a,c)*P(b|c) \;/\; P(c) \\
= & P(a|c) * P(b|c)
\end{aligned}
$$



所以，在c给定的条件下，a，b被阻断(blocked)，是独立的，称之为head-to-tail条件独立。

要点总结

要点1

贝叶斯网络定义

要点2

有向分离的概念

03 马尔科夫模型

3.1 马尔科夫的定义

3.2 马尔科夫的应用

3.1 马尔科夫链的定义

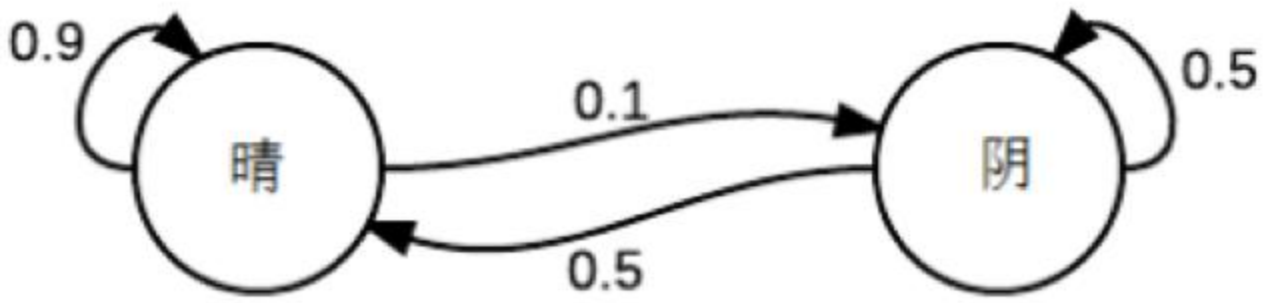
马尔可夫链，因安德烈·马尔可夫（A.A.Markov，1856－1922）得名，是指数学中具有马尔可夫性质的离散事件随机过程。
在给定当前知识或信息的情况下，过去（即当前以前的历史状态）对于预测将来（即当前以后的未来状态）是无关的。

3.1 马尔科夫链的定义

每个状态的转移只依赖于之前的n个状态，这个过程被称为l个n阶的模型，
其中n是影响转移状态的数目。最简单的马尔科夫过程就是一阶过程，
每一个状态的转移只依赖于其之前的那一个状态。
用数学表达式表示就是下面的样子：

$$P\left(X_{n+1}=x \middle| X_1=x_1, X_2=x_2, \ldots, X_n=x_n\right)=P\left(X_{n+1}=x \middle| X_n=x_n\right)$$

假设天气服从马尔科夫链



转移矩阵

$$P=\begin{pmatrix}0.9 & 0.1\\ 0.5 & 0.5\end{pmatrix}$$

从今天（晴或阴）开始，在遥远的未来的某天，阴晴的概率分布是什么？

$$q=\lim_{n\to\infty}\begin{pmatrix}1 & 0\end{pmatrix}P^n=\begin{pmatrix}0.833 & 0.167\end{pmatrix}$$

马尔科夫链的应用

$$P^{\infty}=\begin{pmatrix}0.833 & 0.167\\0.833 & 0.167\end{pmatrix}$$

$$\begin{pmatrix}1 & 0\end{pmatrix}P^{\infty}=\begin{pmatrix}0 & 1\end{pmatrix}P^{\infty}$$

不管今天阴晴，很多天之后的阴晴分布收敛到一个固定分布

这个固定分布叫 **稳态分布**

很久的未来，每一天天气都是 q = (0.833 0.167)的一个样本点

3.2 马尔科夫链的应用

至此，我们就为上面的一阶马尔科夫过程定义了以下三个部分：

1. 状态：晴天、阴天。
2. 初始向量：定义系统在时间为0的时候的状态的概率。
3. 状态转移矩阵：每种天气转换的概率。所有的能被这样描述的系统都是一个马尔科夫过程。

要点总结

要点1

马尔科夫链的定义

要点2

应用马尔科夫链于实际问题

04

实战案例

4.1 详见随堂附加课件

THANK YOU !

Machine Learning Engineer
机器学习工程师微专业