人工智能之机器学习

朴素贝叶斯（Naive Bayes）

主讲人：李老师

思考:

引子：三门问题

参赛者会看见三扇关闭了的门，其中一扇的后面有一辆汽车，选中后面有车的那扇门可赢得该汽车，另外两扇门后面则各藏有一只山羊。当参赛者选定了一扇门，但未去开启它的时候，节目主持人开启剩下两扇门的其中一扇，露出其中一只山羊。主持人其后会问参赛者要不要换另一扇仍然关上的门。问题是：换另一扇门会否增加参赛者赢得汽车的机率？

思考:

1. 最开始的时候，我们对这三扇门之后有什么一无所知，所以我们最好的做法是公平对待三扇门，我们假设 $A_n, n = 1, 2, 3$ 为第n个门之后有汽车，那么我们有$P(A_n) = 1/3$。
2. 假设我们选择门1，主持人打开了门2，这时根据我们打开的门之后是否有汽车，主持人打开的门的概率是会有变化的：如果门1后有汽车，对于一般人（精神正常的人）来说，主持人打开门2和门3的概率基本上应该是一致的，为1/2；如果门2后有汽车，主持人打开门2的概率是0，如果门3后有汽车，主持人打开门2的概率是1。
3. 我们设B为主持人打开了门2，那么我们可以得到：$P(B|A_1) = 1/2$，$P(B|A_2) = 0$，$P(B|A_3) = 1$，也就是2的概率解释。那么我们计算$P(A_1|B)$，这个式子表示我们在得到主持人打开了门2，后面没有汽车这个事实之后，对于$P(A_1)$这个概率的调整：

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B)}$$

，而$P(B)$可以通过全概率公式计算：

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) = 1/2$$

4. 计算得到$P(A_1|B) = 1/3$，这个的含义就是，当我们得到事实B时，我们对先验概率$P(A_n)$的值调整为了后验概率$P(A_n|B)$。当然如上所见，1门后有汽车的整体概率仍然没有变化，其实变化的是$P(A_2|B)$与$P(A_3|B)$，$P(A_2) = 1/3$变成了$P(A_2|B) = 0$，$P(A_3) = 1/3$变成了$P(A_3|B) = 2/3$，提高的概率足够令我们改变自己的决策。

数学回顾：乘法公式、全概率公式、贝叶斯公式

- 条件概率
  - 设A，B为任意两个事件，若P(A)>0，我们称在已知事件A发生的条件下，事件B发生的概率为条件概率，记为P(B|A)，并定义

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

- 乘法公式
  - 如果P(A)>0，则P(AB)=P(A)P(B|A)
  - 如果$P(A_1 \ldots A_{n-1})>0$，则$P(A_1 \ldots A_n)= P(A_1)\, P(A_2|A_1)\, P(A_3|A_1A_2) \ldots P(A_n|A_1 \ldots A_n)$

条件概率及三个公式（乘法公式、全概率公式、贝叶斯公式）

- 全概率公式
  - 如果 $\bigcup_{i=1}^{n} A_i = \Omega$，$A_i A_j = \phi$(对一切$i \neq j$)，$P(A_i)>0$，则对任一事件B，有

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B \mid A_i)$$

  - 全概率公式是用于计算某个“结果” B发生的可能性大小。如果一个结果B的发生总是与某些前提条件Ai 相联系，那么在计算P(B)时，我们就要用Ai 对B作分解，应用全概率公式计算P(B)，我们常称这种方法为**全集分解法。**
  - 根据小偷们的资料，计算村子今晚失窃概率的问题：$P(A_i)$表示小偷 i 作案的概率，$P(B|A_i)$表示小偷 i 作案成功的概率，那么P(B)就是村子失窃的概率

条件概率及三个公式（乘法公式、全概率公式、贝叶斯公式）

- 贝叶斯公式（又称逆概公式）
  - 如果 $\bigcup_{i=1}^{n} A_i = \Omega$，$A_i A_j = \phi$(对一切$i \neq j$)，P(Ai)>0，则对任一事件B，只要P(B)>0，有

$$P(A_j \mid B) = \frac{P(A_j B)}{P(B)} = \frac{P(A_j)P(B \mid A_j)}{\sum_{i=1}^{n} P(A_i)P(B \mid A_i)} (i, j = 1,2,\ldots,n)$$

  - 如果在B发生的条件下探求导致这一结果的各种“原因” A_j 发生的可能性大小$P(A_j|B)$，则要应用贝叶斯公式
  - 若村子今晚失窃，计算哪个小偷嫌疑最大的问题（嫌疑最大就是后验概率最大）

贝叶斯公式

- 假设小偷1和小偷2在某村庄的作案数量比为3:2，前者偷窃成功的概率为0.02，后者为0.01，现村庄失窃，求这次失窃是小偷1作案的概率。
- 【分析】A_1={小偷1作案}，A_2={小偷2作案}，B={村庄失窃}

$$P(A_1)=3/5,\quad P(A_2)=2/5$$

$$P(B\mid A_1)=0.02,\quad P(B\mid A_2)=0.01$$

$$P(A_1\mid B)=\frac{P(A_1)P(B\mid A_1)}{\sum_{i=1}^{2}P(A_i)P(B\mid A_i)}=\frac{3/5\times 0.02}{3/5\times 0.02+2/5\times 0.01}=3/4$$

$$P(A_2\mid B)=\frac{P(A_2)P(B\mid A_2)}{\sum_{i=1}^{2}P(A_i)P(B\mid A_i)}=\frac{2/5\times 0.01}{3/5\times 0.02+2/5\times 0.01}=1/4$$

朴素贝叶斯直观理解

- 肤色x1={黑，黄}， 发型x2={卷，直}； 地区label={亚，非}
- 比如告诉你一个人，其肤色=黑，发型=卷，那么你会预测这个人的地区为亚洲还是非洲?

朴素贝叶斯直观理解

- 模型构建：根据资料计算模型参数
  - 亚洲人的比例
  - 非洲人的比例
  - 亚洲人中肤色=黑的比例
  - 亚洲人中肤色=黄的比例
  - 非洲人中肤色=黑的比例
  - 非洲人中肤色=黄的比例
  - 亚洲人中发型=卷的比例
  - 亚洲人中发型=直的比例
  - 非洲人中发型=卷的比例
  - 非洲人中发型=直的比例

朴素贝叶斯直观理解

- 例子
  - 有一个训练集包含100个人，其中有60个非洲人（黑卷*47, 黑直*1, 黄卷*11, 黄直*1），有40个亚洲人（黑卷*1, 黄卷*4, 黄直*35），请训练朴素贝叶斯模型
- 先计算先验概率：

$$P(\text{非洲}) = \frac{60}{100}, P(\text{亚洲}) = \frac{40}{100}$$

- 再计算每一个特征的条件概率：

$$P(\text{黑}|\text{非洲}) = \frac{48}{60}, P(\text{黄}|\text{非洲}) = \frac{12}{60}, P(\text{直}|\text{非洲}) = \frac{2}{60}, P(\text{卷}|\text{非洲}) = \frac{58}{60}$$

$$P(\text{黑}|\text{亚洲}) = \frac{1}{40}, P(\text{黄}|\text{亚洲}) = \frac{39}{40}, P(\text{直}|\text{亚洲}) = \frac{35}{40}, P(\text{卷}|\text{亚洲}) = \frac{5}{40}$$

朴素贝叶斯直观理解

- 假设新来了一个人【[黑，卷]，地区=？】，请用朴素贝叶斯模型预测这个人的地区。Y表示地区，X表示特征向量，根据贝叶斯公式，并假设特征间独立的假设有：

$$P(Y|X)=\frac{P(Y)\times P(X|Y)}{P(X)} \xrightarrow{\text{条件独立性假设}} P(Y|X)=\frac{P(Y)\times P(X^{(1)}|Y)\times P(X^{(2)}|Y)}{P(X)}$$

- 和特征间独立的假设（朴素），得

$$P(\text{非洲}|\text{黑卷})=P(\text{非洲})P(\text{黑}|\text{非洲})P(\text{卷}|\text{非洲})=\frac{60}{100}\cdot\frac{48}{60}\cdot\frac{58}{60}$$

$$P(\text{亚洲}|\text{黑卷})=P(\text{亚洲})P(\text{黑}|\text{亚洲})P(\text{卷}|\text{亚洲})=\frac{40}{100}\cdot\frac{1}{40}\cdot\frac{5}{40}$$

$$\begin{aligned}P(X=x|Y=c_k)&=P(X^{(1)}=x^{(1)},\cdots,X^{(n)}=x^{(n)}|Y=c_k)\\&=\prod_{j=1}^{n}P(X^{(j)}=x^{(j)}|Y=c_k)\end{aligned}$$

- 根据计算结果，模型会将这个人的地区预测为非洲。

朴素贝叶斯算法推导

$$P(y\mid x_1,x_2,\ldots x_m)=\frac{P(y)P(x_1,x_2,\ldots x_m\mid y)}{P(x_1,x_2,\ldots x_m)}$$

- 特征属性之间是独立的，所以得到： $P(x_i\mid y,x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_m)=P(x_i\mid y)$
- 公式优化得到：

$$P(y\mid x_1,x_2,\ldots x_m)=\frac{P(y)P(x_1,x_2,\ldots x_m\mid y)}{P(x_1,x_2,\ldots x_m)}=\frac{P(y)\prod_{i=1}^{m}P(x_i\mid y)}{P(x_1,x_2,\ldots x_m)}$$

- 在给定样本的情况下，$P(x_1,x_2,\ldots,x_m)$是常数，所以得到：

$$P(y\mid x_1,x_2,\ldots x_m)\propto P(y)\prod_{i=1}^{m}P(x_i\mid y)$$

- 从而：

$$\hat{y}=\arg\max_{y} P(y)\prod_{i=1}^{m}P(x_i\mid y)$$

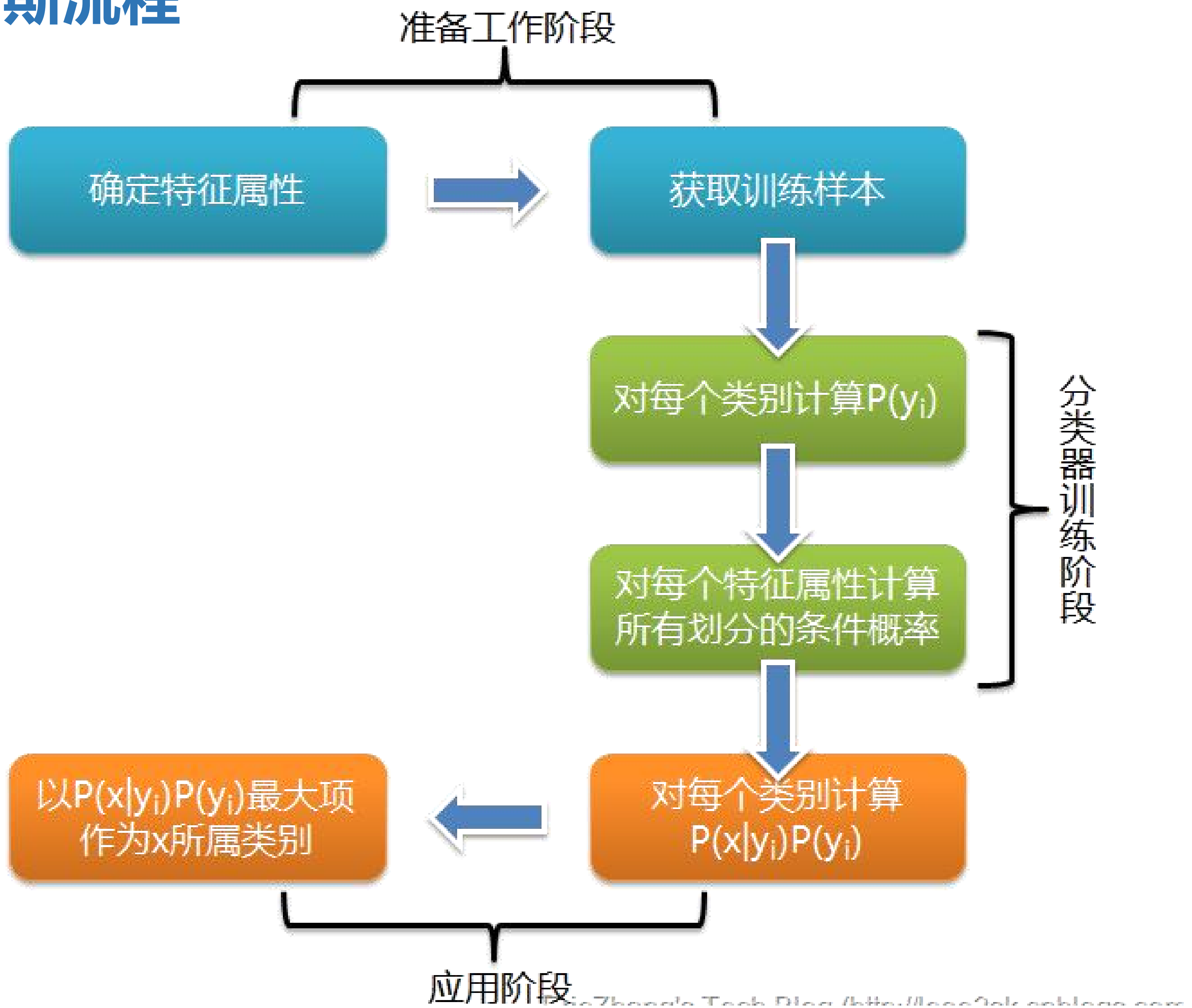
朴素贝叶斯算法流程

- 朴素贝叶斯算法流程/定义如下：
  - 设$x=\{x_1,x_2,\ldots,x_m\}$为待分类项，其中x_i为x的一个特征属性
  - 类别集合为$C=\{y_1,y_2,\ldots,y_n\}$
  - 分别计算$P(y_1|x),P(y_2|x),\ldots,P(y_n|x)$的值（贝叶斯公式）
  - 如果$P(y_k|x)=\max\{P(y_1|x),P(y_2|x),\ldots,P(y_n|x)\}$,那么认为x为$y_k$类型

$$P(y|x_1,x_2,\ldots x_m)=\frac{P(y)P(x_1,x_2,\ldots x_m|y)}{P(x_1,x_2,\ldots x_m)}$$

$$P(y|x_1,x_2,\ldots x_m)\propto P(y)\prod_{i=1}^{m}P(x_i|y)$$

朴素贝叶斯流程



准备工作阶段
确定特征属性
获取训练样本
对每个类别计算P(yi)
对每个特征属性计算所有划分的条件概率
分类器训练阶段
对每个类别计算P(x|yi)P(yi)
以P(x|yi)P(yi)最大项作为x所属类别
应用阶段

高斯朴素贝叶斯

- Gaussian Naive Bayes是指当特征属性为连续值时，而且分布服从高斯分布，那么在计算P(x|y)的时候可以直接使用高斯分布的概率公式：

$$g(x,\eta,\sigma)=\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{(x-\eta)^2}{2\sigma^2}}$$

$$P(x_i \mid y_k)=g\left(x_i,\eta_{i,y_k},\sigma_{i,y_k}\right)$$

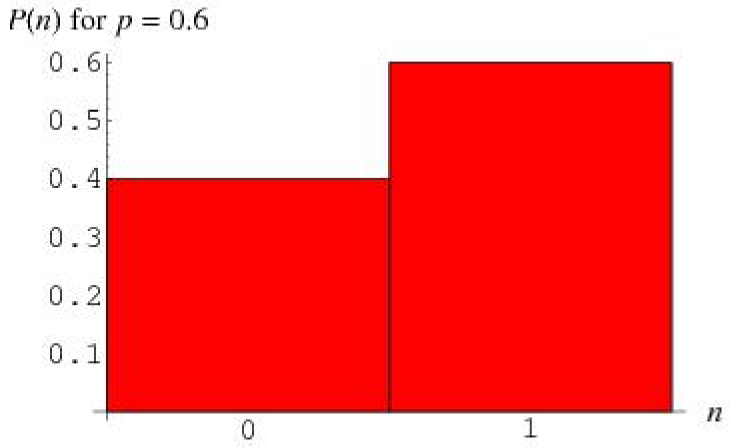
- 因此只需要计算出各个类别中此特征项划分的各个均值和标准差

伯努利朴素贝叶斯

- Bernoulli Naive Bayes是指当特征属性为连续值时，而且分布服从伯努利分布，那么在计算P(x|y)的时候可以直接使用伯努利分布的概率公式：

$$P(x_k \mid y) = P(1 \mid y)x_k + \left(1 - P(1 \mid y)\right)\left(1 - x_k\right)$$

- 伯努利分布是一种离散分布，只有两种可能的结果。1表示成功，出现的概率为p；0表示失败，出现的概率为q=1-p；其中均值为E(x)=p，方差为Var(X)=p(1-p)



多项式朴素贝叶斯

- Multinomial Naive Bayes是指当特征属性服从多项分布(特征是离散的形式的时候)，直接计算类别数目的占比作为先验概率和条件概率。

$$p(y_k)=\frac{N_{y_k}+\alpha}{N+k*\alpha} \qquad p(x_i|y_k)=\frac{N_{y_k,x_i}+\alpha}{N_{y_k}+n_i*\alpha}$$

- N是总样本个数，k是总的类别个数，N_{yk}是类别为yk的样本个数，α为平滑值。
- N_{yk}是类别为yk的样本个数，n_i为特征属性xi的不同取值数目，$N_{yk,xi}$为类别yk中第i维特征的值为xi的样本个数，α为平滑值。
- 当α=1时，称为Laplace平滑，当0<α<1时，称为Lidstone平滑，α=0时不做平滑；平滑的主要作用是可以克服条件概率为0的问题。

朴素贝叶斯有一个问题

- 继续前文的引例，考虑一个这样的问题：
- 假设某人的地区完全依靠其肤色的就能确定，发型是一个对判断地区没有参考价值的特征，假设P(卷|非洲)=0， P(卷|亚洲)=0.001，当来了一个【黑，卷】人的时候，我们算出

$$P(\text{非洲})P(\text{黑}\mid\text{非洲})P(\text{卷}\mid\text{非洲})=0$$

$$P(\text{亚洲})P(\text{黑}\mid\text{亚洲})P(\text{卷}\mid\text{亚洲})=0.00001$$

- 然后被预测为亚洲人，傻了吧?
- 原因：出现某个模型参数为0时，0乘任何数都=0，直接影响到后验概率的计算结果。

拉普拉斯平滑

- 解决这一问题的方法是使用平滑操作，改造先验概率公式：

$$P(\text{非洲}) = \frac{60+\lambda}{100+len\{\text{亚洲，非洲}\}\cdot\lambda} = \frac{60+\lambda}{100+2\cdot\lambda}$$

- 改造每个特征的条件概率公式（这里只列举了2个）：

$$P(\text{黑}\,|\,\text{非洲}) = \frac{48+\lambda}{60+len\{\text{黑,白}\}\cdot\lambda} = \frac{48+\lambda}{60+2\cdot\lambda}$$

$$P(\text{直}\,|\,\text{非洲}) = \frac{2+\lambda}{60+len\{\text{直,卷}\}\cdot\lambda} = \frac{2+\lambda}{60+2\cdot\lambda}$$

- 在随机变量各个取值的频数上赋予一个正数，当时$\lambda=1$，称为拉普拉斯平滑

多项式朴素贝叶斯案例理解

- 对于下列训练数据，使用多项式朴素贝叶斯方式对测试样本(2,M,L)做一个预测判断。

	1	2	3	4	5	6	7	8	9	10
x1	1	1	1	2	2	2	2	3	3	4
x2	S	M	S	L	S	S	L	L	L	S
x3	L	H	L	H	L	M	H	M	H	M
y	-1	1	1	-1	-1	-1	1	1	1	1

$$N = 10$$

$$k = 2 \; n_1 = 4$$

$$n_2 = 3 \; n_3 = 3$$

	x1=1	x1=2	x1=3	x1=4	
y=1	2	1	2	1	6
y=-1	1	3	0	0	4
	3	4	2	1	10

	x2=S	x2=M	x2=L	
y=1	2	1	3	6
y=-1	3	0	1	4
	5	1	4	10

	x3=L	x3=M	x3=H	
y=1	1	2	3	6
y=-1	2	1	1	4
	3	3	4	10

多项式朴素贝叶斯案例理解

$\alpha = 0$

- 先验概率：

$$p(y=1)=6/10=0.6 \qquad p(y=-1)=4/10=0.4$$

- 条件概率：

$$p(x_1=1|y=1)=\frac{2}{6} \qquad p(x_1=1|y=-1)=\frac{1}{4}$$

$$p(x_1=2|y=1)=\frac{1}{6} \qquad p(x_1=2|y=-1)=\frac{3}{4}$$

$$p(x_1=3|y=1)=\frac{2}{6} \qquad p(x_1=3|y=-1)=0$$

$$p(x_1=4|y=1)=\frac{1}{6} \qquad p(x_1=4|y=-1)=0$$

$$p(x_2=S|y=1)=\frac{2}{6} \qquad p(x_2=S|y=-1)=\frac{3}{4}$$

$$p(x_2=M|y=1)=\frac{1}{6} \qquad p(x_2=M|y=-1)=0$$

$$p(x_2=L|y=1)=\frac{3}{6} \qquad p(x_2=L|y=-1)=\frac{1}{4}$$

多项式朴素贝叶斯案例理解

- 条件概率：

$$p(x_3 = L|y=1) = \frac{1}{6} \qquad p(x_3 = L|y=-1) = \frac{2}{4}$$

$$p(x_3 = M|y=1) = \frac{2}{6} \qquad p(x_3 = M|y=-1) = \frac{1}{4}$$

$$p(x_3 = H|y=1) = \frac{3}{6} \qquad p(x_3 = H|y=-1) = \frac{1}{4}$$

- 样本(2,M,L)的预测概率：

$$\alpha = 0$$

$$p(y=1|x) \propto p(y=1)p(x_1=2|y=1)p(x_2=M|y=1)p(x_3=L|y=1) = \frac{6}{10}*\frac{1}{6}*\frac{1}{6}*\frac{1}{6} = \frac{1}{360}$$

$$p(y=-1|x) \propto p(y=-1)p(x_1=2|y=-1)p(x_2=M|y=-1)p(x_3=L|y=-1) = \frac{4}{10}*\frac{3}{4}*0*\frac{2}{4} = 0$$

$$\hat{y} = \arg\max_{y}\{p(y=1|x), p(y=-1|x)\} = 1$$

多项式朴素贝叶斯案例理解

$$\alpha = 1$$

- 先验概率：

$$p(y=1)=(6+1)/(10+2*1)=7/12 \quad p(y=-1)=5/12$$

- 条件概率：

$$p(x_1=1|y=1)=\frac{3}{10} \quad p(x_1=1|y=-1)=\frac{2}{8} \quad p(x_2=S|y=1)=\frac{3}{9} \quad p(x_2=S|y=-1)=\frac{4}{7}$$

$$p(x_1=2|y=1)=\frac{2}{10} \quad p(x_1=2|y=-1)=\frac{4}{8} \quad p(x_2=M|y=1)=\frac{2}{9} \quad p(x_2=M|y=-1)=\frac{1}{7}$$

$$p(x_1=3|y=1)=\frac{3}{10} \quad p(x_1=3|y=-1)=\frac{1}{8} \quad p(x_2=L|y=1)=\frac{4}{9} \quad p(x_2=L|y=-1)=\frac{2}{7}$$

$$p(x_1=4|y=1)=\frac{2}{10} \quad p(x_1=4|y=-1)=\frac{1}{8}$$

多项式朴素贝叶斯案例理解

- 条件概率： $p(x_3=L|y=1)=\frac{2}{9}$ $p(x_3=L|y=-1)=\frac{3}{7}$

$p(x_3=M|y=1)=\frac{3}{9}$ $p(x_3=M|y=-1)=\frac{2}{7}$

$p(x_3=H|y=1)=\frac{4}{9}$ $p(x_3=H|y=-1)=\frac{2}{7}$

- 样本(2,M,L)的预测概率概率：

$$\alpha=1$$

$$p(y=1|x)\propto p(y=1)p(x_1=2|y=1)p(x_2=M|y=1)p(x_3=L|y=1)=\frac{7}{12}*\frac{2}{10}*\frac{2}{9}*\frac{2}{9}=\frac{7}{1215}$$

$$p(y=-1|x)\propto p(y=-1)p(x_1=2|y=-1)p(x_2=M|y=-1)p(x_3=L|y=-1)=\frac{5}{12}*\frac{4}{8}*\frac{1}{7}*\frac{3}{7}=\frac{5}{392}$$

$$\hat{y}=\arg\max_y\{p(y=1|x),p(y=-1|x)\}=-1$$

案例一：鸢尾花数据分类

- 使用高斯朴素贝叶斯API对鸢尾花数据进行分类操作

class `sklearn.naive_bayes.GaussianNB` ¶ [source]

Attributes:

class_prior_ : array, shape (n_classes,)

各个类别的概率

probability of each class.

class_count_ : array, shape (n_classes,)

各个类别的样本数量

number of training samples observed in each class.

theta_ : array, shape (n_classes, n_features)

各个类别中各个特征属性的均值

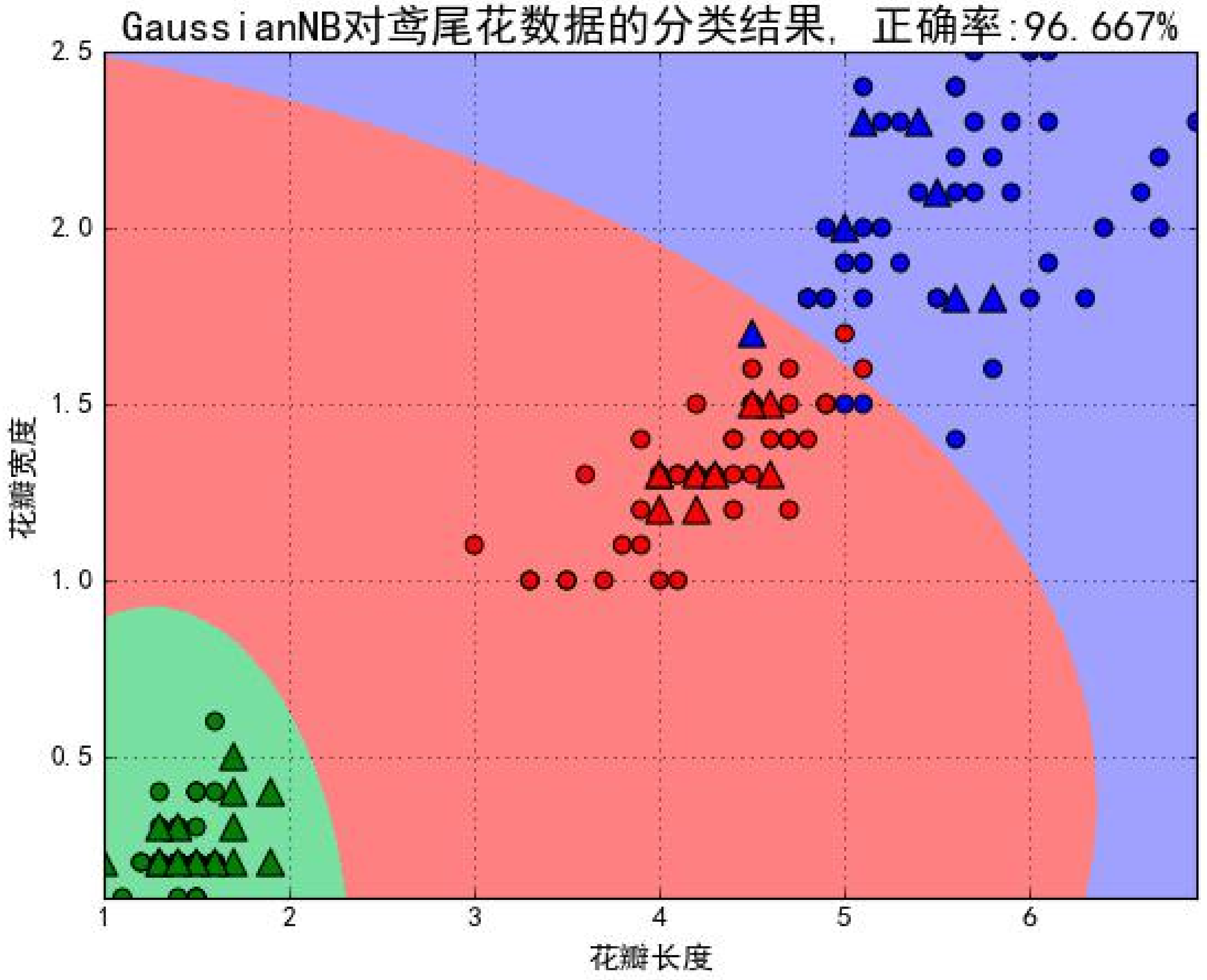
mean of each feature per class

sigma_ : array, shape (n_classes, n_features)

各个类别中各个特征属性的方差

variance of each feature per class

案例一：鸢尾花数据分类



GaussianNB对鸢尾花数据的分类结果, 正确率:96.667%
2.5
2.0
1.5
1.0
0.5
1
2
3
4
5
6
花瓣宽度
花瓣长度

贝叶斯网络

- 把某个研究系统中涉及到的**随机变量**，根据是否条件独立绘制在一个**有向图**中，就形成了贝叶斯网络。
- 贝叶斯网络(Bayesian Network)，又称**有向无环图模型**(directed acyclic graphical model, DAG)，是一种概率图模型，根据概率图的拓扑结构，考察一组随机变量$\{X_1,X_2,\ldots,X_n\}$及其N组**条件概率分布**(Conditional Probabililty Distributions, CPD)的性质

贝叶斯网络

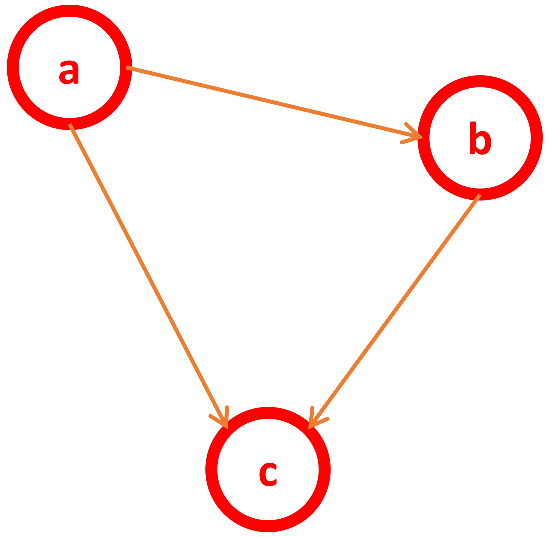
- 当多个特征属性之间存在着某种相关关系的时候，使用朴素贝叶斯算法就没法解决这类问题，那么贝叶斯网络就是解决这类应用场景的一个非常好的算法。
- 一般而言，贝叶斯网络的有向无环图中的节点表示随机变量，可以是可观察到的变量，或隐变量，未知参数等等。连接两个节点之间的箭头代表两个随机变量之间的因果关系(也就是这两个随机变量之间非条件独立)，如果两个节点间以一个单箭头连接在一起，表示其中一个节点是“因”，另外一个是“果”，从而两节点之间就会产生一个条件概率值。

贝叶斯网络

- 贝叶斯网络的关键方法是图模型，构建一个图模型我们需要把具有因果联系的各个变量用箭头连在一起。贝叶斯网络的有向无环图中的节点表示随机变量。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的。
- 贝叶斯网络是模拟人的认知思维推理模式的，用一组条件概率以及有向无环图对不确定性因果推理关系建模

最简单的一个贝叶斯网络

$$P(a,b,c) = P(c \mid a,b)P(b \mid a)P(a)$$

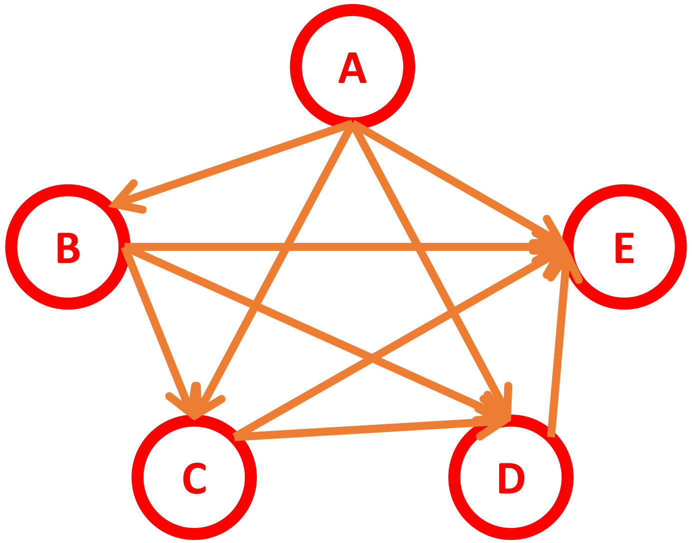


全连接贝叶斯网络

- 每一对节点之间都有边连接

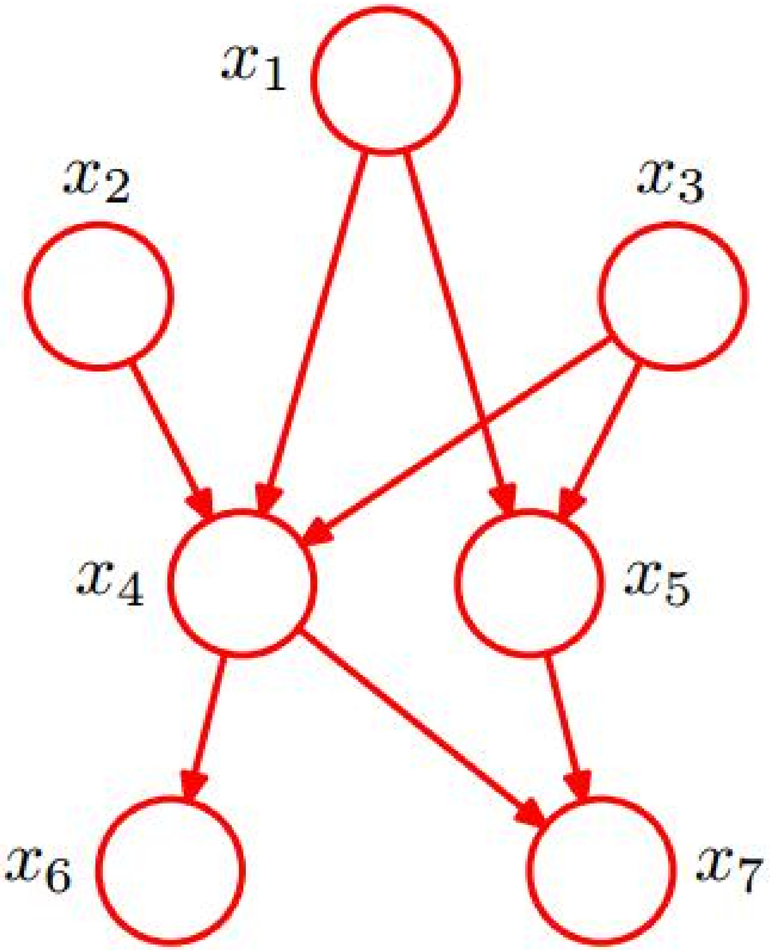
$$P(x_1, x_2, \ldots, x_n) = P(x_n \mid x_1, x_2, \ldots, x_{n-1}) \ldots P(x_2 \mid x_1) P(x_1)$$

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=2}^{n} P(x_i \mid x_1, x_2, \ldots, x_{i-1}) * P(x_1)$$



“正常”贝叶斯网络

- x1,x2,x3独立
- x6和x7在给定条件下独立
- x1,x2,x3…x7的联合分布为



$$p(x_1,x_2,x_3,x_4,x_5,x_6,x_7)=p(x_1)p(x_2)p(x_3)p(x_4\mid x_1,x_2,x_3)p(x_5\mid x_1,x_3)p(x_6\mid x_4)p(x_7\mid x_4,x_5)$$

贝叶斯网络判定条件独立-01

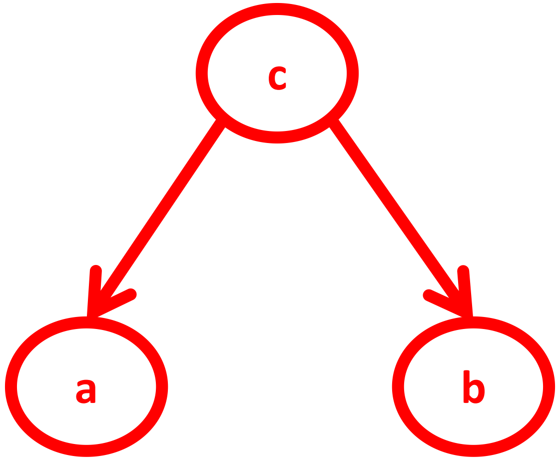
- 在C给定的条件下，a和b被阻断(blocked)是独立的
  - 条件独立：tail - to -tail

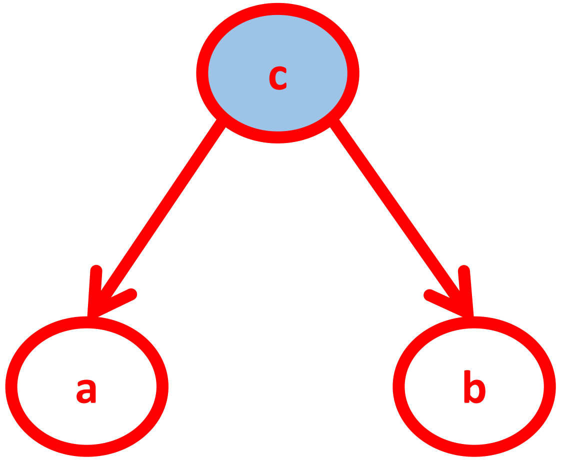
$$P(a,b,c) = P(c)P(b|c)P(a|c)$$

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

$$\because p(a,b|c) = \mathrm{p}(a,b,c)/p(c)$$

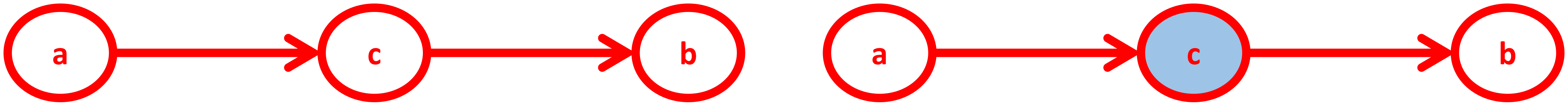
$$\therefore p(a,b|c) = p(a|c)p(b|c)$$





贝叶斯网络判定条件独立-02

- 在C给定的条件下，a和b被阻断(blocked)是独立的
  - 条件独立：head- to -tail

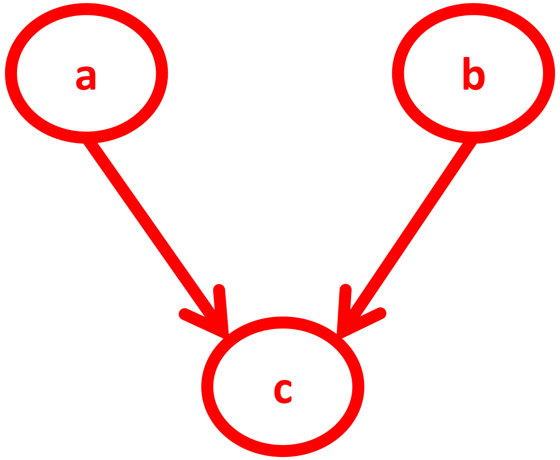


$$P(a,b,c)=P(a)P(c\mid a)P(b\mid c)$$

$$\begin{aligned}
&p(a,b\mid c)\\
&=p(a,b,c)/p(c)\\
&=p(a)*p(c\mid a)*p(b\mid c)/p(c)\\
&=p(a,c)*p(b\mid c)/p(c)\\
&=p(a\mid c)*p(b\mid c)
\end{aligned}$$

贝叶斯网络判定条件独立-03

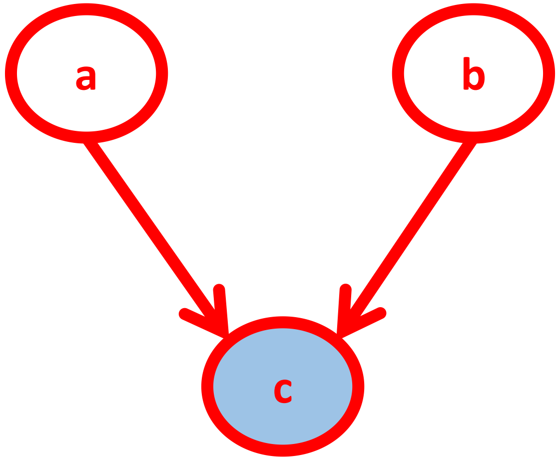
- 在C未知的情况下，a和b被阻断(blocked)，是独立的
  - 条件独立：head - to - head



$$P(a,b,c) = P(\mathrm{a})P(b)P(c \mid a,b)$$

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

$$\Rightarrow P(a,b) = P(a) * P(b)$$



THANKS