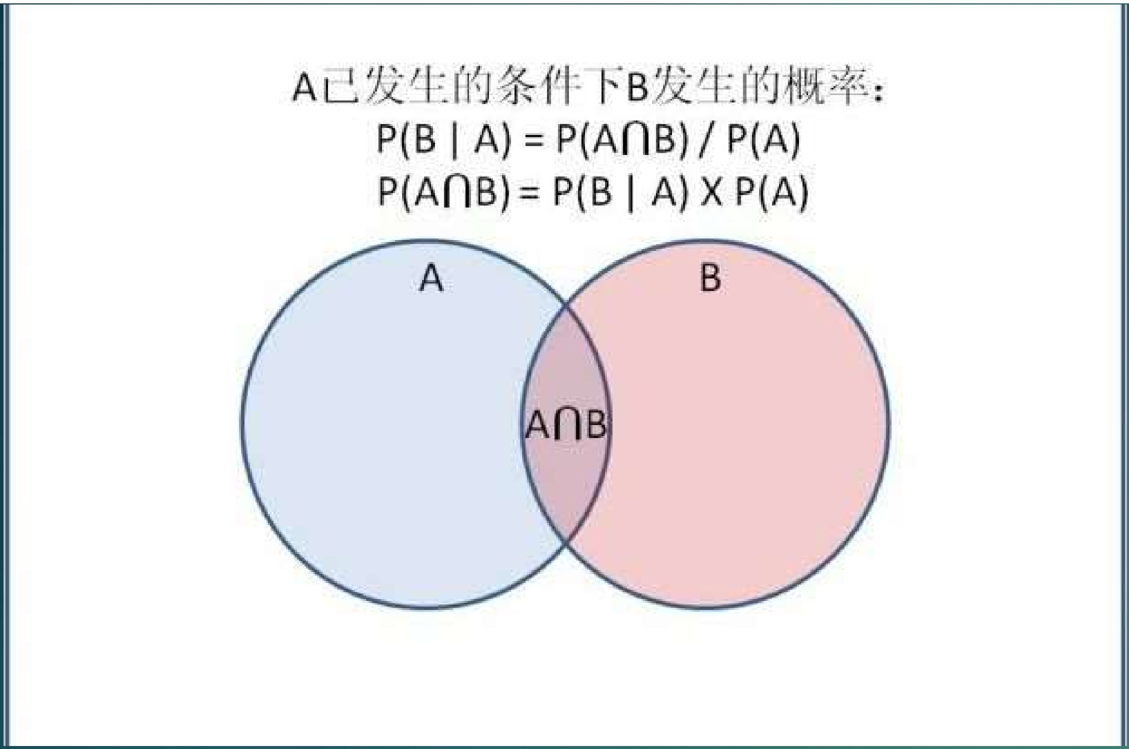
3-5 贝叶斯公式

张霖

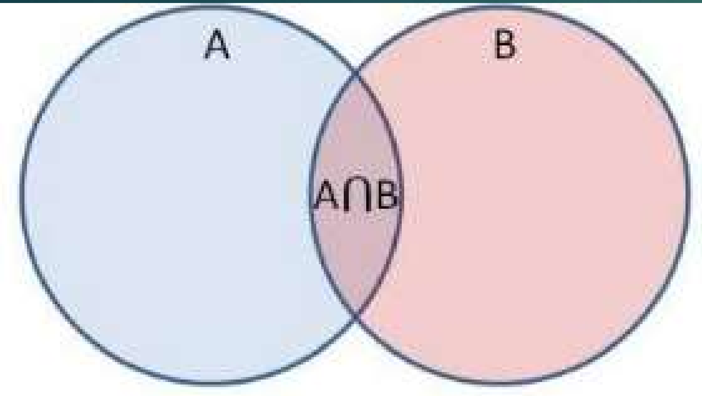


$P(A)$指的是事件A发生的概率，这个称为先验概率，即不考虑任何其它事件的影响

$P(B|A)$指事件A发生条件下，B发生的概率，这个称为后验概率

- 有一座别墅在过去的20年里一共发生改过2次被盗，别墅的主人有一条狗，狗平均每周晚上叫3次，在盗贼入侵时狗叫的概率被估计为0.9，问题：在狗叫的时候发生入侵的概率是多少？

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



设狗叫为事件A，家中被盗为事件B

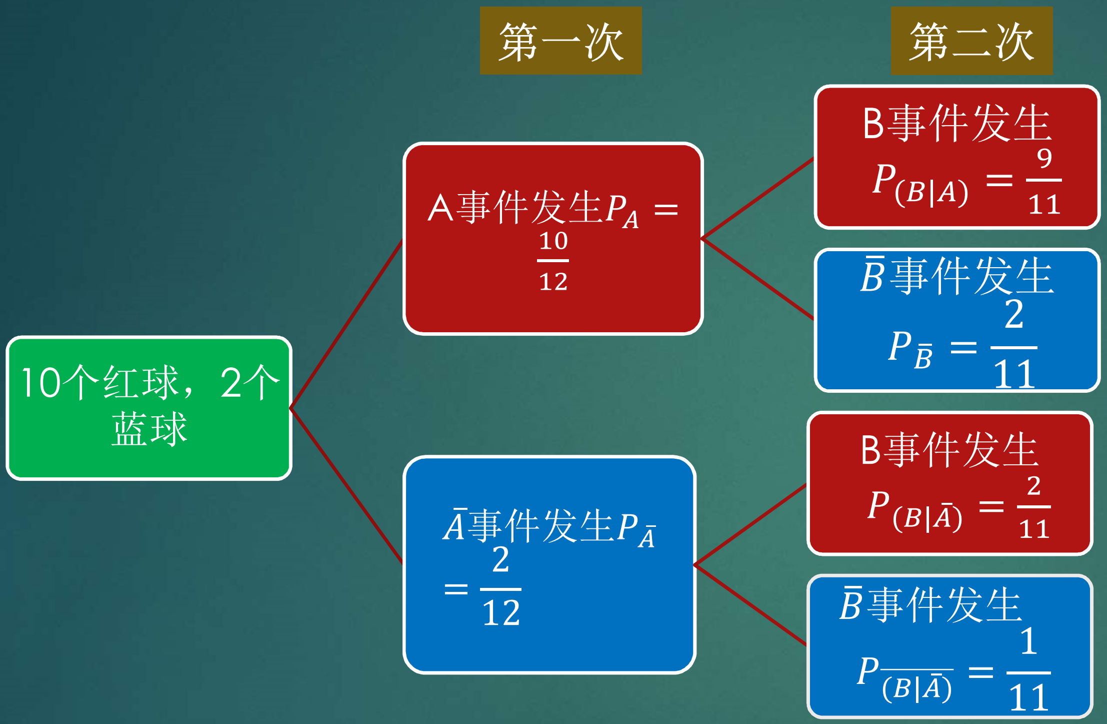
$$P(A) = \frac{3}{7}，P(B) = \frac{2}{20 \times 365} = 0.00027$$

$$P(A|B) = 0.9$$

$$P(AB) = P(A|B) \times P(B) = 0.00025$$

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{0.00025}{0.00027} = 0.926$$

- 口袋里面有12个球，10个红球，2个蓝球，从口袋中先取一个球，不看颜色且不放回，问第二个球取到红球的概率。



全概率公式

$$P(A) = \sum_{i=1}^{N} P(A|B_i) \cdot P(B_i)$$

假设第一次抓到红球为事件A，则第一次抓到蓝球为$\bar{A}$，第二次抓到红球为事件B，则第二次抓到蓝球为$\bar{B}$。

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A})$$

```
def calculateClassProb(input_data,train_Summary_by_class):
  prob = {}
  for class_value, summary in train_Summary_by_class.items():
      prob[class_value] = 1
      for i in range(len(summary)):
          mean,var = summary[i]
          x = input_data[i]
          p = calculateProb(x,mean,var)
      prob[class_value] *=p
  return prob

input_vector = testset[1]
input_data = input_vector[:-1]
train_Summary_by_class = summarizeByClass(trainset)
class_prob = calculateClassProb(input_data,train_Summary_by_class)

#{'Setosa': 3.3579279836005993,
# 'Versicolour': 1.5896628317396685e-07,
# 'Virginica': 5.176617264913899e-12}
```