**朴素贝叶斯**

> 参考《机器学习实战》第四章

1、贝叶斯决策：概率越大就属于哪一类

是某个实例的特征。

2、条件概率：

3、贝叶斯交换准则：交换条件和结果。

**用朴素贝叶斯进行文本分类**

1、朴素的含义：

* 假设1：每个词出现在一个词后面的概率是相同的。
* 假设2：每个词（特征）的重要性相同。

2、特征定义

* 这里的特征是来自文本的词条（token），一个词条是字符的任意组合。可以把词条想象为单词，也可以使用非单词词条，如URL、IP地址或者任意其他字符串。
* 将每一个文本片段表示为一个词条向量，其中值为1表示词条出现在文档中，0表示词条未出现。