**\* KNN和K-Means的具体做法可见第4章课件。**

**\* 贝叶斯分类：用一个例子来说明：**

**下表中的实例显示出如何根据天气、温度、湿度、风力来判断是否打网球。现有一个天气预报状况为：**

**Z=（天气=“晴”，温度=“高”，湿度=“小”，风力=“无”）**

**试采用贝叶斯分类方法判断在Z情况下是否打网球。**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **样本id** | **天气** | **温度** | **湿度** | **风力** | **类别** |
| **X1** | **晴** | **高** | **大** | **无** | **否** |
| **X2** | **晴** | **高** | **大** | **无** | **否** |
| **X3** | **云** | **高** | **大** | **无** | **是** |
| **X4** | **雨** | **中** | **大** | **无** | **是** |
| **X5** | **雨** | **低** | **小** | **无** | **是** |
| **X6** | **雨** | **低** | **小** | **有** | **否** |
| **X7** | **云** | **低** | **小** | **有** | **是** |
| **X8** | **晴** | **中** | **大** | **无** | **否** |
| **X9** | **晴** | **低** | **小** | **无** | **是** |
| **X10** | **雨** | **中** | **小** | **无** | **是** |
| **X11** | **晴** | **中** | **小** | **有** | **是** |
| **X12** | **云** | **中** | **大** | **有** | **是** |
| **X13** | **云** | **高** | **小** | **无** | **是** |
| **X14** | **雨** | **中** | **大** | **有** | **否** |

**根据题意，共14个实例，分为两个类，分别假设为C1和C2，且有：**

**C1=C是={X3,X4,X5,X7,X9,X10,X11,X12,X13}，数量为9**

**C2=C否={X1,X2,X6,X8,X14}，数量为5**

**故有：P(C1)=9/14, P(C2)=5/14**

**原理：Z已经发生了，现在要看Z发生的情况下，Z属于C1类的概率（即P(C1/Z)）与Z属于C2类的概率（即P(C2/Z)）哪个大，就判断Z属于那个类。**

**如上所述，需要比较P(C1/Z)和P(C2/Z)这两个值的大小。**

**根据贝叶斯公式：P(C1/Z)= ，P(C2/Z)= ，其中P(Z)为常数，且对个公式两者都有，故可去掉。这样，只需比较P(Z/C1)P(C1)与P(Z/C2)P(C2)这两个值的大小。**

**先计算P(Z/C1)P(C1)**

**P(Z/C1)P(C1) = P(“天气=晴”/C1)\*P(“温度=高”/C1)\*P(“湿度=小”/C1)\*P(“风力=无”/C1)\*P(C1)**

**= (2/9)\*(2/9)\*(6/9)\*(6/9)\*(9/14)=0.014**

**（注意，只有“天气”、“温度”、“湿度”、“风力”这4个属性的取值相互无关时，上式才成立。**

**P(“天气=晴”/C1)就是“是”这个类（共9个样本）中“天气为晴”的样本（共2个）的概率，即2/9。其余类推。）**

**再计算P(Z/C2)P(C2)**

**P(Z/C2)P(C2) = P(“天气=晴”/C2)\*P(“温度=高”/C2)\*P(“湿度=小”/C2)\*P(“风力=无”/C2)\*P(C2)**

**= (3/5)\*(2/5)\*(1/5)\*(3/5)\*(5/14)=0.010**

**最后比较：**

**因为0.014 > 0.010，即P(C1/Z)>P(C1/Z)，所以分类为“是”，结论时天气Z适合打网球。**