本周学习：朴素贝叶斯

贝叶斯定理：

1. 条件概率

条件概率是指在事件B发生的情况下，事件A发生的概率，用http://latex.codecogs.com/gif.latex?P(A\mid&space;B)表示。

1. 全概率公式

http://latex.codecogs.com/gif.latex?P(B)=P(B\mid&space;A)P(A)+P(B\mid&space;%7bA%7d%27)P(%7bA%7d%27)

含义是：如果http://latex.codecogs.com/gif.latex?A和http://latex.codecogs.com/gif.latex?%7bA%7d%27构成样本空间的一个划分，那么事件B的概率，就等于http://latex.codecogs.com/gif.latex?A和http://latex.codecogs.com/gif.latex?%7bA%7d%27的概率分别乘以B对这两个事件的条件概率之和。

1. 贝叶斯推断

C:\Documents and Settings\Administrator\桌面\gif.gif

推导过程：

条件概率公式 C:\Documents and Settings\Administrator\桌面\b64543a98226cffcb1b6da18b8014a90f703eaf8.jpg

P（AB）=P（A|B）P（B）=P（B|A）P(A)代入上式即可

其中http://latex.codecogs.com/gif.latex?P(A)称为先验概率，即在事件B发生之前，我们对事件A的概率的一个判断。http://latex.codecogs.com/gif.latex?P(A\mid&space;B)称为后验概率，即在事件B发生之后，我们对事件A的概率的重新估计。http://latex.codecogs.com/gif.latex?\frac%7bP\left&space;(&space;B\mid&space;A&space;\right&space;)%7d%7bP\left&space;(&space;B&space;\right&space;)%7d称为调整因子，使得预估概率更接近真实概率。

贝叶斯推断告诉我们，先预估计一个“先验概率”，然后加入实验结果，看这个实验到底是增强还是削弱了“先验概率”，由此得到更接近事实的“后验概率”。

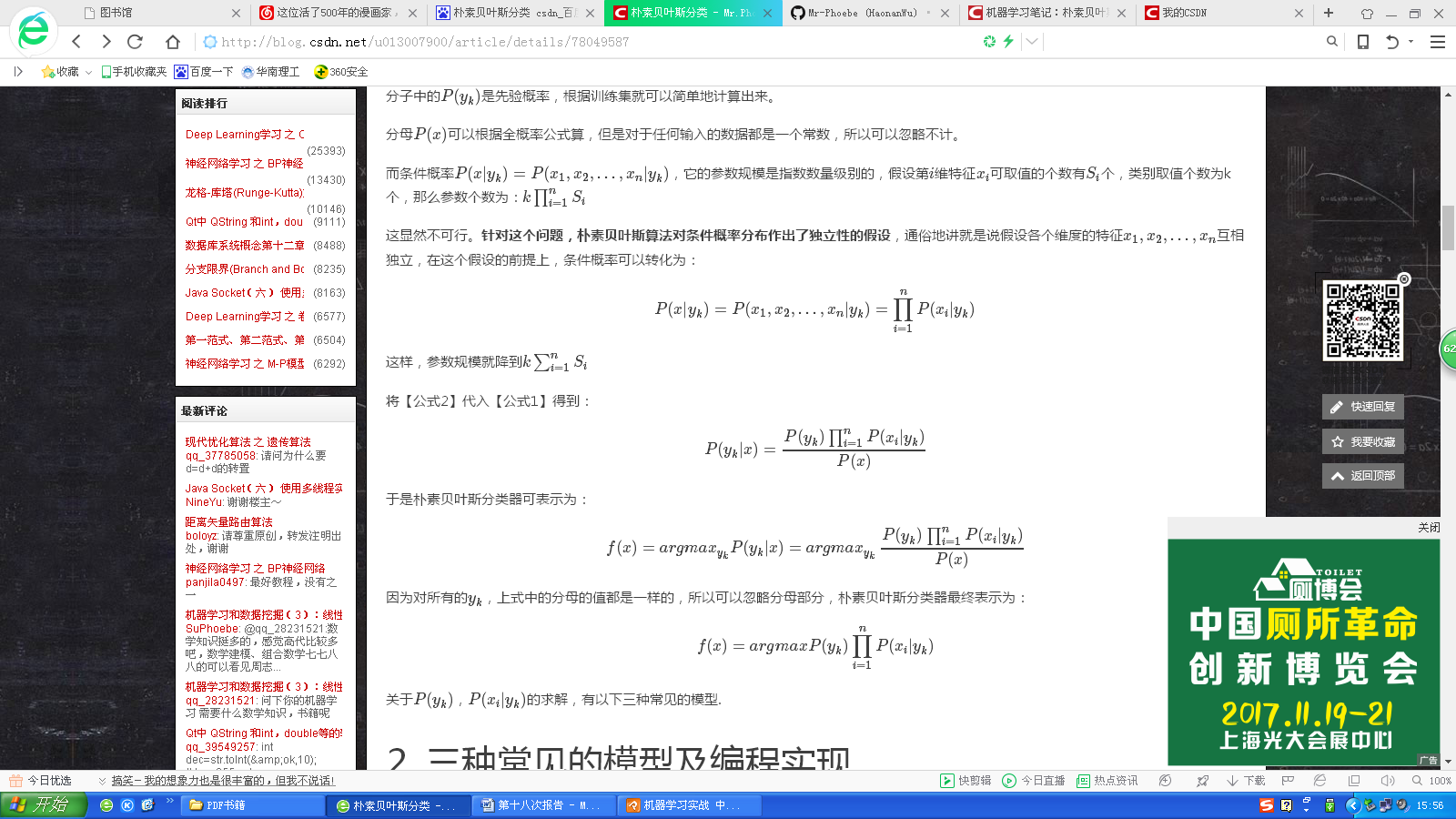
朴素贝叶斯：

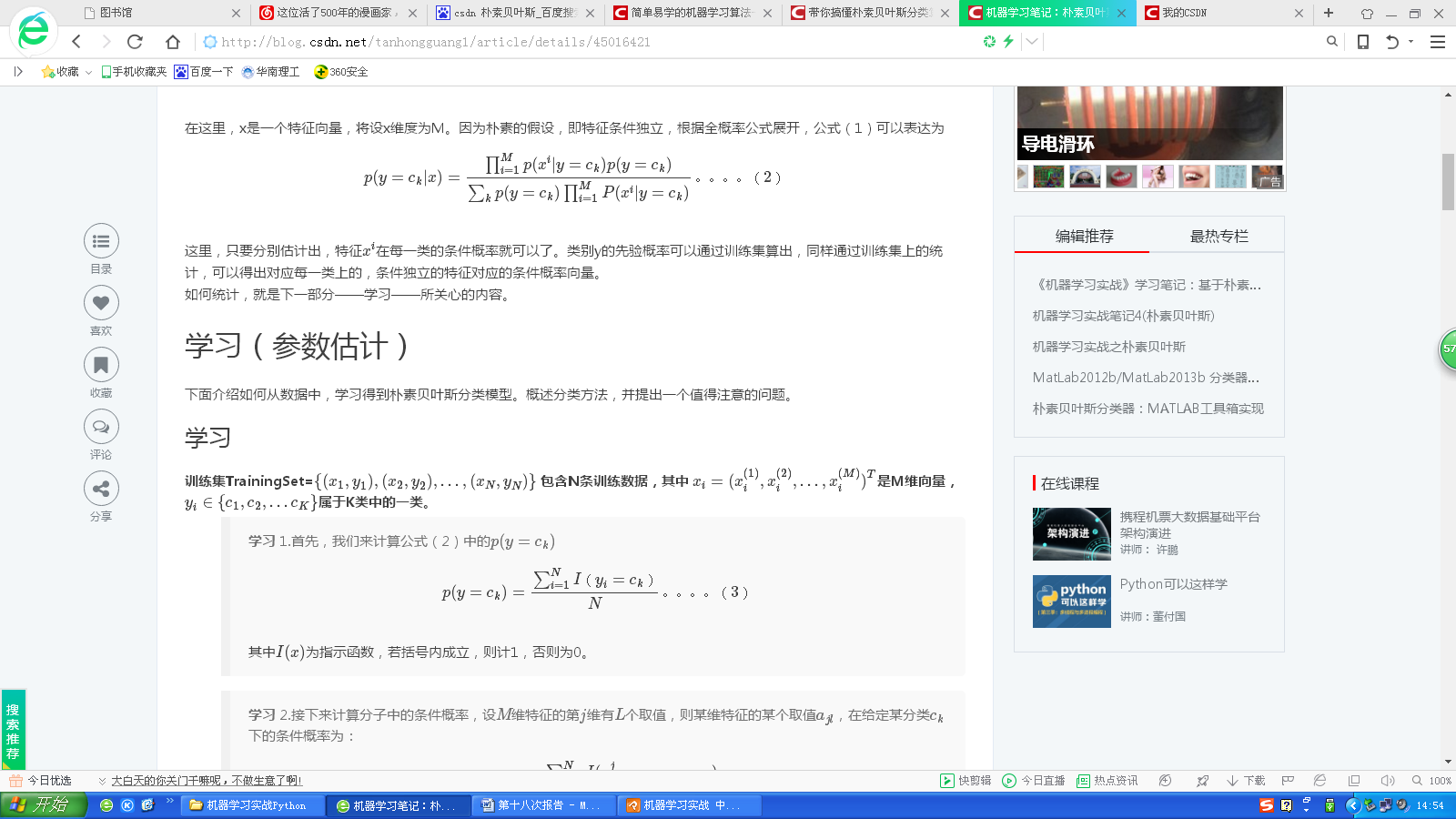
朴素：特征条件独立

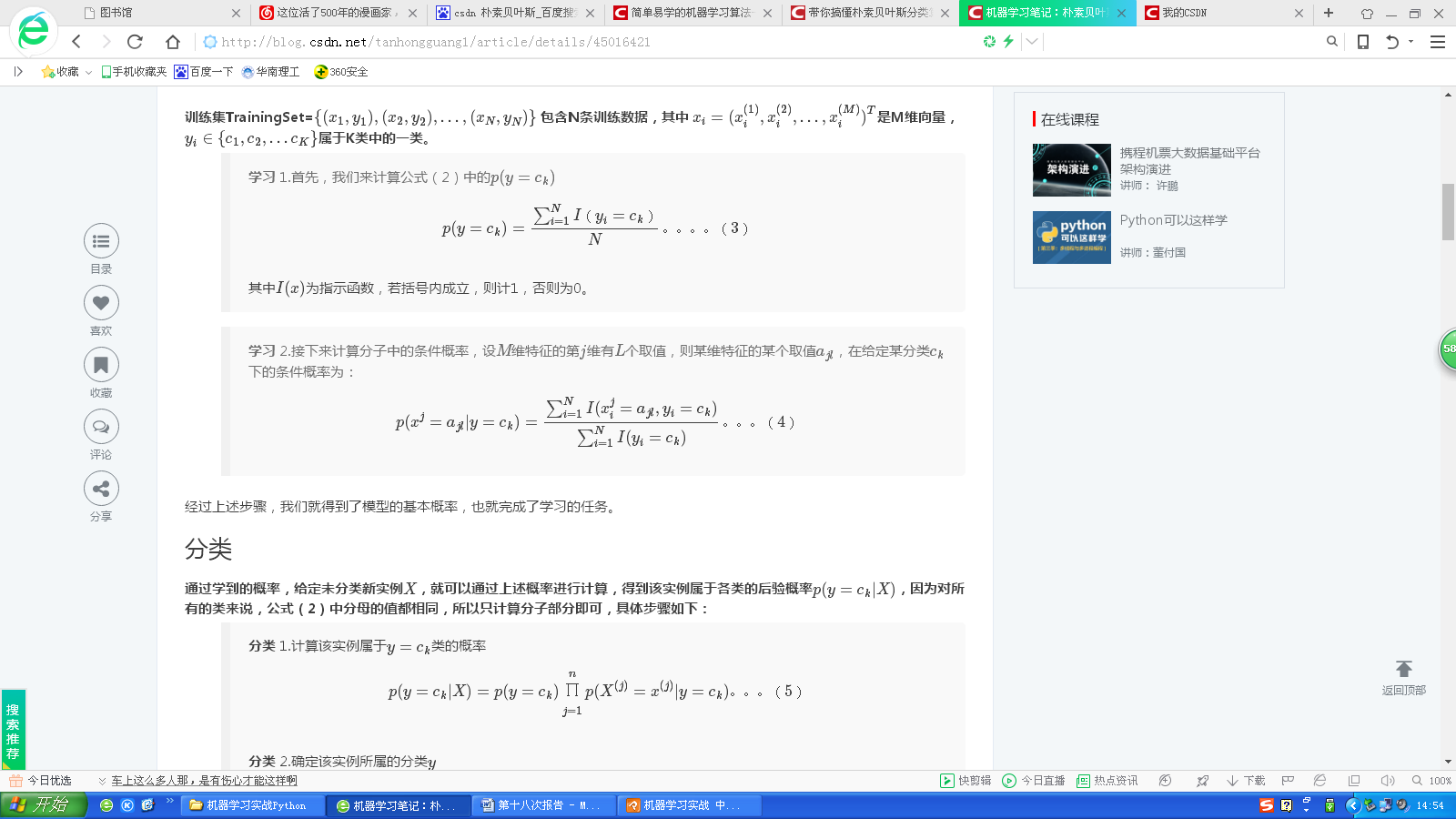
贝叶斯：基于贝叶斯定理

数据学习过程：

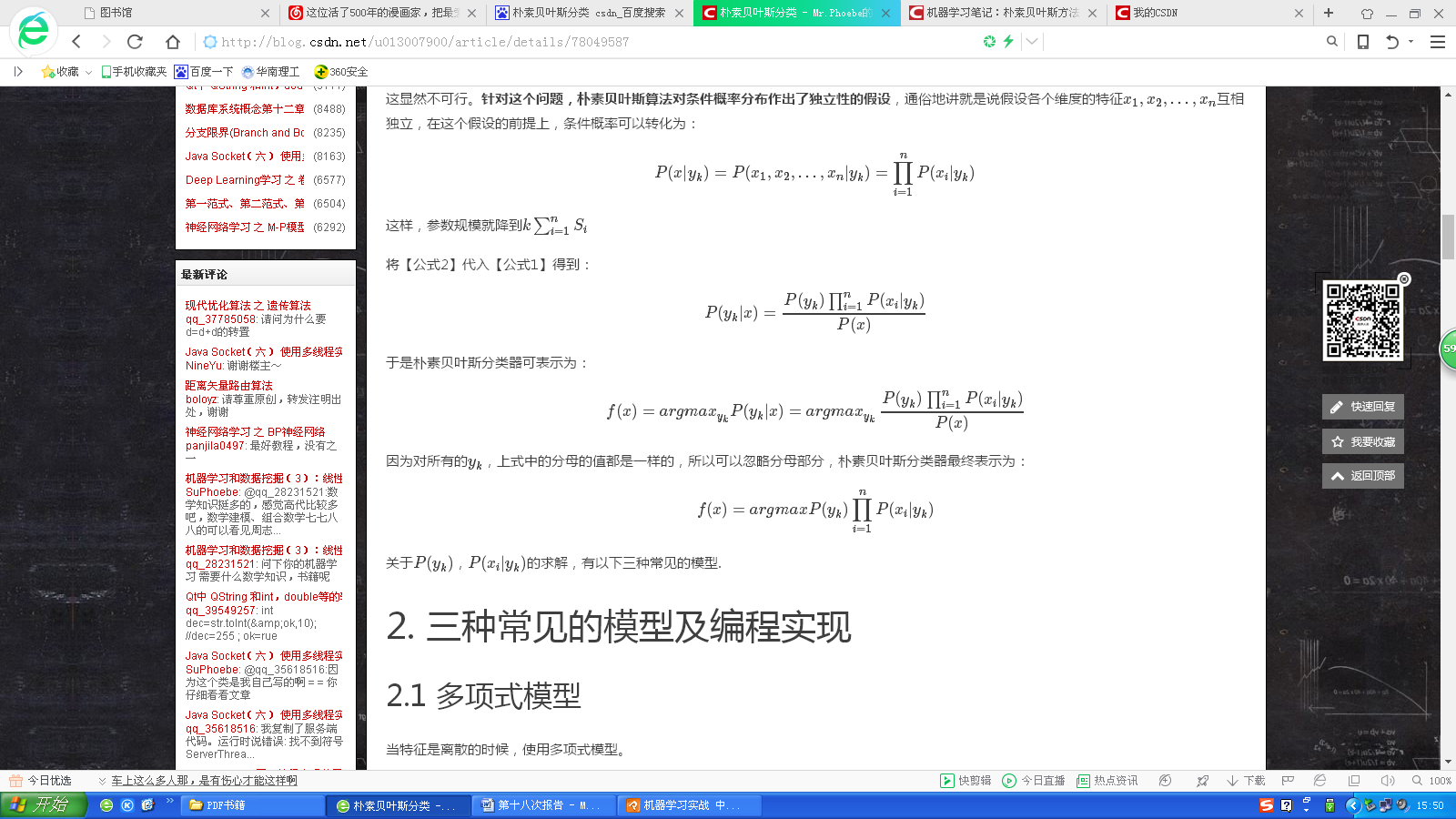
训练集TrainingSet={(x1,y1),(x2,y2),...,(xN,yN)} 包含N条训练数据，其中 xi=(x(1)i,x(2)i,...,x(n)i)T是n维向量，yi∈{c1,c2,...cK}属于K类中的一类。那么公式就可以变成：



其中某些式子的计算如下:  




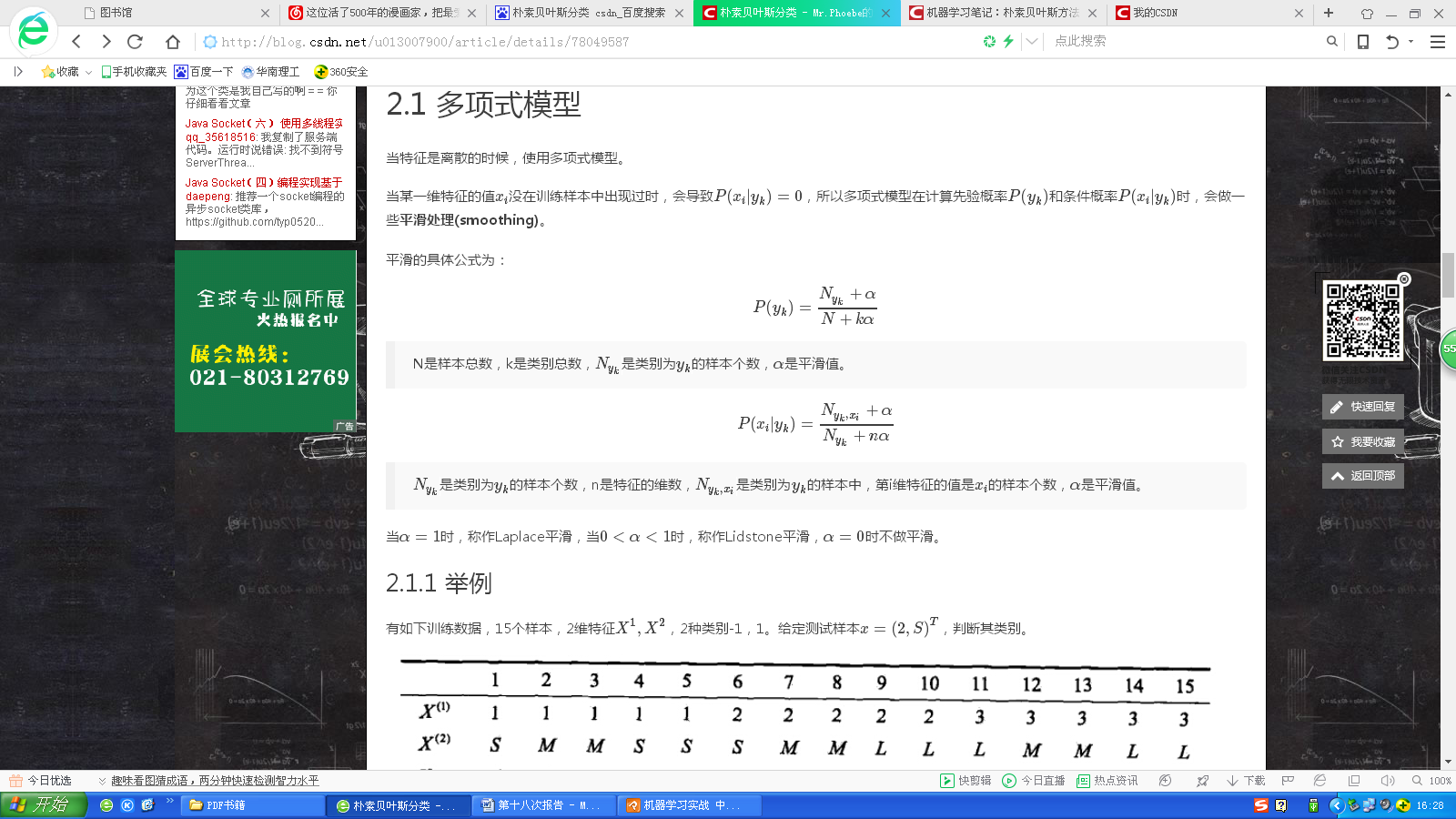
分类过程：



使用模型：

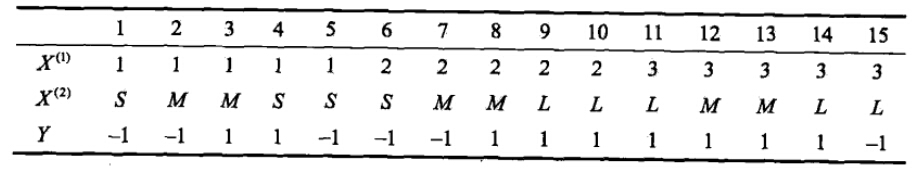
多项式模型：适用于数据是离散的情况

当某一维特征的值xi没在训练样本中出现过时，会导致P(xi|yk)=0，所以多项式模型在计算先验概率P(yk)和条件概率P(xi|yk)时，会做一些平滑处理(smoothing)



例子：

如下训练数据，15个样本，2维特征X1,X2，2种类别-1，1。给定测试样本x=(2,S)T，判断其类别。



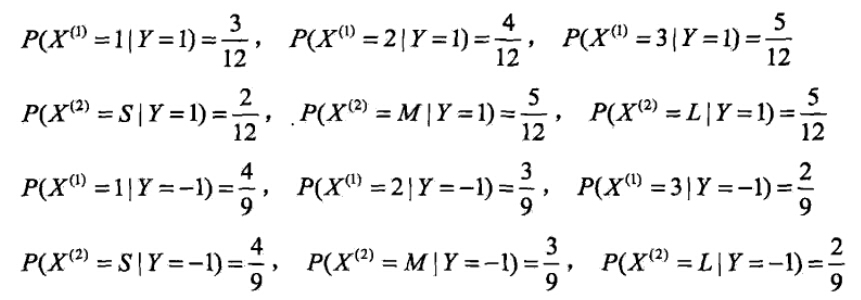
求解过程：

令α=1，有K=2，n=2

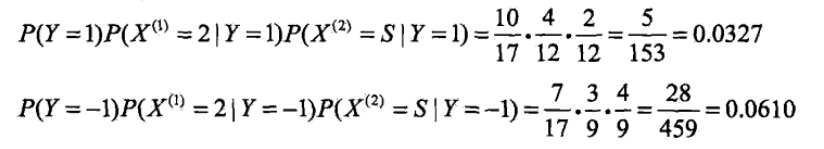
则先验概率为

http://img.blog.csdn.net/20150909085100191

各种条件概率为



对于给定的x=(2,S)T，计算：



取概率较大者y=-1

参考资料：

[1] CSDN博主 [**SuPhoebe**](http://my.csdn.net/u013007900) <http://blog.csdn.net/u013007900/article/details/78049587>

[2] CSDN博主 [**Journey-Go**](http://blog.csdn.net/tanhongguang1)<http://blog.csdn.net/tanhongguang1/article/details/45016421>

[3] CSDN博主 [**zhiyong\_will**](http://my.csdn.net/google19890102)<http://blog.csdn.net/google19890102/article/details/25592833>

[4] 机器学习实战 第四章 基于概率论的分类方法——朴素贝叶斯