朴素贝叶斯分类器

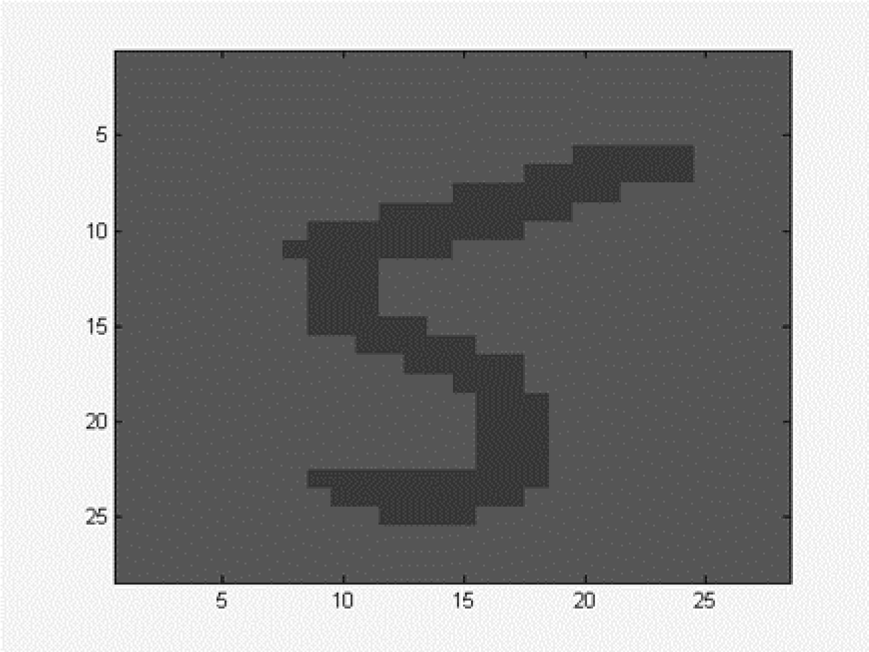
该课程主要为大家讲授如下的内容：

- 贝叶斯公式
- 朴素贝叶斯公式
- 用参数估计替代概率统计

1. 贝叶斯公式
   我们可以使用贝叶斯公式计算每一个标签类别的概率,然后选择概率最高的标签作为预测标签。

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

   数字识别案例：
   想要识别输入数字是5还是6即可。输入是一张30 * 30的图片,为简化问题,假定图片的每个像素只有两种取值(0和1)。需要分别计算P(Y=5|X1,…,XN)和P(Y=6|X1,…,XN)这两个概率. 然后选取概率较大的标签类作为输出即可。为此需要找到X相对于Y的条件概率:$\boldsymbol{P(X \mid Y)}$和先验概率:$\boldsymbol{P(Y)}$, $\boldsymbol{P(X)}$。而这些值的计算则是在训练过程中完成的。贝叶斯分类器的训练过程实际上是根据大数定律,使用分布的抽样(也就是训练数据)来估计这些概率。



$$P(Y=5|X_1,\ldots,X_n) = \frac{P(X_1,\ldots,X_n|Y=5)P(Y=5)}{P(X_1,\ldots,X_n|Y=5)P(Y=5)+P(X_1,\ldots,X_n|Y=6)P(Y=6)}$$
$$P(Y=6|X_1,\ldots,X_n) = \frac{P(X_1,\ldots,X_n|Y=6)P(Y=6)}{P(X_1,\ldots,X_n|Y=5)P(Y=5)+P(X_1,\ldots,X_n|Y=6)P(Y=6)}$$

2. 朴素贝叶斯公式
   首先引入一个假设：特征之间是相互独立的, 为模型引入这个条件之后,就可以计算之前无法计算的概率。当特征 x1 x2 ...之间相互独立时, 其联合条件概率等于各个概率的乘积。
   这种条件虽然理论上存在下次，但是实际应用中效果不错。
   朴素贝叶斯分类器模型的训练过程其实是一个统计过程,它会分别统计 x 和y 的先验概率以及($\boldsymbol{X} \mid \boldsymbol{Y}$)的条件概率。这些概率都是通过计数的方式去计算的。在模型预测的时候使用下面的公式计算所有可能标签的概率,选取概率最大的标签作为预测标签，这就是朴素贝叶斯分类器。
3. 用参数估计替代概率统计
   可以假设概率符合一种特定的分布,比如说高斯分布(这里分布的选择要根据数据特点来决定)。对于高斯分布，知道它是由两个参数来控制的,也就是均值和方差。只需要使用训练数据来估计这两个参数就可以得到 $\boldsymbol{P}(\boldsymbol{X}_i \mid \boldsymbol{Y})$的概率分布了。这里的参数估计可以使用极大似然估计或

者是其他方式来进行。

$$P(x_i|c) = \frac{1}{\sqrt{2\pi\sigma_{c,i}^2}} exp(\frac{-(x_i - \mu_{c,i})^2}{2\sigma_{c,i}^2})$$