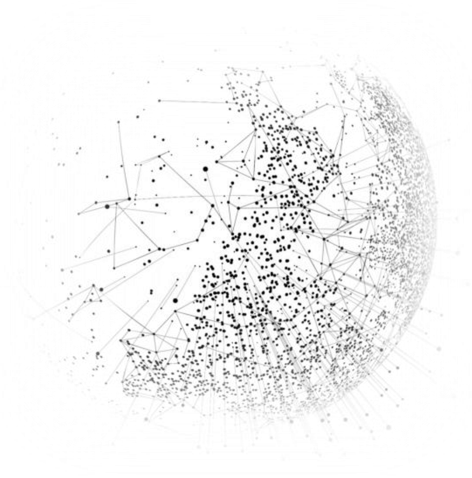

七、贝叶斯分类器

主讲教师：周志华

贝叶斯决策论 (Bayesian Decision Theory)

概率框架下实施决策的基本理论

给定 N 个类别，令 λ_{ij} 代表将第 j 类样本误分类为第 i 类所产生的损失，则基于后验概率将样本 $\boldsymbol{x}$ 分到第 i 类的条件风险为：

$$R(c_i \mid \boldsymbol{x}) = \sum_{j=1}^{N} \lambda_{ij} P(c_j \mid \boldsymbol{x})$$

贝叶斯判定准则 (Bayes decision rule)：

$$h^*(\boldsymbol{x}) = \underset{c \in \mathcal{Y}}{\arg\min}\, R(c \mid \boldsymbol{x})$$

- h^* 称为贝叶斯最优分类器(Bayes optimal classifier)，其总体风险称为贝叶斯风险 (Bayes risk)
- 反映了学习性能的理论上限

判别式 vs. 生成式

$P(c \mid \boldsymbol{x})$ 在现实中通常难以直接获得

从这个角度来看，机器学习所要实现的是基于有限的训练样本尽可能准确地估计出后验概率

两种基本策略：

判别式 (Discriminative) 模型

思路：直接对 $P(c \mid \boldsymbol{x})$ 建模

代表：

- 决策树
- BP 神经网络
- SVM

生成式 (Generative) 模型

思路：先对联合概率分布 $P(\boldsymbol{x}, c)$ 建模，再由此获得 $P(c \mid \boldsymbol{x})$

$$P(c \mid \boldsymbol{x}) = \frac{P(\boldsymbol{x}, c)}{P(\boldsymbol{x})}$$

代表：贝叶斯分类器

注意：贝叶斯分类器 ≠ 贝叶斯学习 (Bayesian Learning)

极大似然估计

先假设某种概率分布形式，再基于训练样例对参数进行估计

假定 $P(\boldsymbol{x} \mid c)$ 具有确定的概率分布形式，且被参数$\boldsymbol{\theta}_c$ 唯一确定，则任务就是利用训练集 D 来估计参数 $\boldsymbol{\theta}_c$

$\boldsymbol{\theta}_c$ 对于训练集 D 中第 c 类样本组成的集合 D_c 的似然(Likelihood)为

$$P(D_c \mid \boldsymbol{\theta}_c) = \prod_{\boldsymbol{x} \in D_c} P(\boldsymbol{x} \mid \boldsymbol{\theta}_c)$$

连乘易造成下溢，因此通常使用对数似然 (Log-Likelihood)

$$LL(\boldsymbol{\theta}_c) = \log P(D_c \mid \boldsymbol{\theta}_c) = \sum_{\boldsymbol{x} \in D_c} \log P(\boldsymbol{x} \mid \boldsymbol{\theta}_c)$$

于是，$\boldsymbol{\theta}_c$ 的极大似然估计为 $\hat{\boldsymbol{\theta}}_c = \underset{\boldsymbol{\theta}_c}{\arg\max}\ LL(\boldsymbol{\theta}_c)$

估计结果的准确性严重依赖于所假设的概率分布形式是否符合潜在的真实分布

朴素贝叶斯分类器 (Naïve Bayes Classifier)

$$P(c \mid \boldsymbol{x}) = \frac{P(c)\boxed{P(\boldsymbol{x} \mid c)}}{P(\boldsymbol{x})}$$

主要障碍：所有属性上的联合概率难以从有限训练样本估计获得

组合爆炸；样本稀疏

基本思路：假定属性相互独立？

$$P(c \mid \boldsymbol{x}) = \frac{P(c)\, P(\boldsymbol{x} \mid c)}{P(\boldsymbol{x})} = \frac{P(c)}{P(\boldsymbol{x})} \prod_{i=1}^{d} P(x_i \mid c)$$

d 为属性数，x_i 为 x 在第 i 个属性上的取值

$P(\boldsymbol{x})$ 对所有类别相同，于是

$$h_{nb}(\boldsymbol{x}) = \underset{c \in \mathcal{Y}}{\arg\max}\; P(c) \prod_{i=1}^{d} P(x_i \mid c)$$

朴素贝叶斯分类器

- 估计 $P(c)$：$P(c) = \frac{|D_c|}{|D|}$
- 估计 $P(\boldsymbol{x}|c)$：

• 对离散属性，令 D_{c,x_i} 表示 D_c 中在第 i 个属性上取值为 x_i 的样本组成的集合，则

$$P(x_i \mid c) = \frac{|D_{c,x_i}|}{|D_c|}$$

• 对连续属性，考虑概率密度函数，假定 $p(x_i \mid c) \sim \mathcal{N}(\mu_{c,i}, \sigma_{c,i}^2)$

$$p(x_i \mid c) = \frac{1}{\sqrt{2\pi}\sigma_{c,i}} \exp\left(-\frac{(x_i - \mu_{c,i})^2}{2\sigma_{c,i}^2}\right)$$

一个例子

（青绿；稍蜷；浊响；清晰） - 好瓜 or 坏瓜？

编号	色泽	根蒂	敲声	纹理	好瓜
1	青绿	蜷缩	浊响	清晰	是
2	乌黑	蜷缩	沉闷	清晰	是
3	乌黑	蜷缩	浊响	清晰	是
4	青绿	蜷缩	沉闷	清晰	是
5	浅白	蜷缩	浊响	清晰	是
6	青绿	稍蜷	浊响	清晰	是
7	乌黑	稍蜷	浊响	稍糊	是
8	乌黑	稍蜷	浊响	清晰	是
9	乌黑	稍蜷	沉闷	稍糊	否
10	青绿	硬挺	清脆	清晰	否
11	浅白	硬挺	清脆	模糊	否
12	浅白	蜷缩	浊响	模糊	否
13	青绿	稍蜷	浊响	稍糊	否
14	浅白	稍蜷	沉闷	稍糊	否
15	乌黑	稍蜷	浊响	清晰	否
16	浅白	蜷缩	浊响	模糊	否
17	青绿	蜷缩	沉闷	稍糊	否

P(青绿|好瓜) = 3/8　P(青绿|坏瓜) = 3/9

P(稍蜷|好瓜) = 3/8　P(稍蜷|坏瓜) = 4/9

P(浊响|好瓜) = 6/8　P(浊响|坏瓜) = 4/9

P(清晰|好瓜) = 7/8　P(清晰|坏瓜) = 2/9

P(青绿|好瓜) P(稍蜷|好瓜) P(浊响|好瓜) P(清晰|好瓜) P(好瓜=yes) = 3/8 x 3/8 x 6/8 x 7/8 x 8/17

P(青绿|坏瓜) P(稍蜷|坏瓜) P(浊响|坏瓜) P(清晰|坏瓜) P(好瓜=no) = 3/9 x 4/9 x 4/9 x 2/9 x 9/17

好瓜！