

# 西瓜书复习笔记07

- 贝叶斯决策论：

- 什么是贝叶斯决策：

对于分类来说，在所有相关概率都已知的情况下，贝叶斯决策论告诉我们如何基于这些概率和误判损失来选择最优的类别标记。

- 贝叶斯最优分类器：

贝叶斯最优分类器是将贝叶斯风险降到最低。

$$h^*(\boldsymbol{x}) = \arg \min_{c \in \mathcal{Y}} R(c|\boldsymbol{x})$$

也就是最小化分类错误率，即使后验概率 $P(c|x)$ 最大。

$$R(c|x) = 1 - P(c|x)$$

$$h^*(x) = \arg \max_{c \in \mathcal{Y}} P(c|x)$$

- 生成式模型和判别式模型：

- 生成式：直接建模预测

$P(x, c)$

- 判别式：对联合概率分布建模预测

$P(c|x)$

- 贝叶斯定理：

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

其中：

联合概率分布： $P(x, c)$ 。

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

其中：

先验： $P(c)$ ，表达样本空间中，各类样本所占比例；

类条件概率： $P(x|c)$ ，是所有属性上的联合概率。

用于归一化的证据因子： $P(x)$

- 如何估计类条件概率：

极大似然估计（MLE），先假定它服从某种概率分布，在机遇训练样本对其参数进行估计。也就是说假设 $p(x|c)$ 在确定的分布形式下被向量 $\theta$ 唯一确定，我们任务是利用训练集 $D$ 估计参数 $\theta$ 。

（概率模型训练过程就是参数估计过程）

也就是说MLE是作用于类条件概率 $P(x|c)$ 的，MAP在MLE基础上再加上先验概率 $P(c)$ 。

似然：

$$P(D_c|\theta_c) = \prod_{x \in D_c} P(x|\theta_c)$$

由于连乘造成下溢，使用对数似然：

$$\begin{aligned} LL(\theta_c) &= \log P(D_c|\theta_c) \\ &= \sum_{x \in D_c} \log P(x|\theta_c) \end{aligned}$$

参数 $\theta_c$ ：

$$\hat{\theta}_c = \arg \max_{\theta_c} LL(\theta_c)$$

估计结果受假设的概率分布形式影响。

- 朴素贝叶斯分类器：

- 为什么要引入NB：

由于后验概率 $P(x|c)$ 的概率分布很难从有限的样本集中直接估计获得。

- 什么是朴素贝叶斯：

朴素贝叶斯分类器采用了属性条件独立性假设，也就是每个属性都是独立存在的，每个属性独立地对分类结果产生影响。

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

$$h_{nb}(x) = \arg \max_{c \in \mathcal{Y}} P(c) \prod_{i=1}^d P(x_i|c)$$

- 训练过程，根据样本D估计先验概率 $P(c)$ ，并为每个属性估计条件概率 $P(x_i|c)$ 。

$$P(c) = \frac{|D_c|}{|D|}$$

对离散属性而言：

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

对连续属性而言：

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

- 平滑：  
避免概率连乘为零。一般用拉普拉斯修正。

$$\hat{P}(c) = \frac{|D_c| + 1}{|D| + N}$$

$$\hat{P}(x_i|c) = \frac{|D_{c,x_i}| + 1}{|D_c| + N_i}$$

其中N表示D中的类别数， $N_i$ 表示第i个属性的可能的取值数。

- 训练方式：
  - 如果数据更替不频繁，可以将所有概率估值计算好，储存起来，新样本预测时直接查表。
  - 如果数据更替频繁，可以使用懒惰学习，先不训练，等新样本来了之后，再对当前数据进行概率估值。

## • 半朴素贝叶斯分类器：

- 为什么要引入：  
朴素贝叶斯的条件独立性假设在现实中很难成立。
- 什么是半朴素贝叶斯：  
既不对联合概率进行计算，又不至于彻底忽略属性间强依赖关系。那么就假设每个属性类别之外最多依赖于一个其他属性。

## • EM算法：

- 隐变量：  
未观测到的变量。之前的似然估计是假设训练样本是完整的，然而现实中往往训练样本是不完整的。比如某个属性的变量值未知。

- 为什么要用EM算法：  
在某种属性的变量未知的情况下，依然能对模型的参数进行估计。
- 什么是EM算法：  
EM算法是一种迭代思想算法。所参数 $\theta$ 已知，求最优隐变量 $Z$ 的值；若隐变量已知，求对 $\theta$ 做最大似然估计。
- EM算法流程：  
EM算法是一种迭代思想算法。
  - 先初始化参数 $\theta$ 的值。
  - E步：基于 $\theta^t$ 推断隐变量的期望 $Z^t$ 。
  - M步：根据观测变量 $X$ 和 $Z^t$ 对参数 $\theta^t$ 做极大似然估计。
- 隐变量的估计：  
是使用EM算法，是一种非梯度优化的方法。