

## 第 14 章

1:

频率方法与贝叶斯方法不同点在于频率方法从自然角度出发, 试图直接为事件本身建模, 即事件  $A$  在独立重复试验中发生的频率趋于极限  $p$ , 那么这个极限就是该事件的概率。

贝叶斯方法并不从试图刻画事件本身, 而从观察者角度出发。贝叶斯方法并不试图说事件本身是随机的, 或者世界的本体带有某种随机性, 这套理论根本不言说关于世界本体的东西, 而只是从观察者知识不完备这一出发点开始, 构造一套在贝叶斯概率论的框架下可以对不确定知识做出推断的方法。

2:

先验分布: 是先于结果, 确定原因得概率分布

后验分布: 是在已知结果的情况下, 由结果估计原因的的概率分布

3:

MAP 估计与最大似然估计通常是相同的, 因为数据决定先验知识, 当有大量满足似然得信息时, 先验由似然决定, 因此期望后验估计最大值与极大似然估计近似。

MAP 分类要求我们发现: 问题在于可能产生维度灾难; 为所有可能存在的  $X$  变量的组合  $p(X_1=x_1, X_2=x_2, \dots, X_m=x_m|\theta) p(\theta)$ . 此时会产生组合空间爆炸问题. 因此需要找到一种方法削减问题的搜索空

间问题.

**8:**

平衡数据集由不同分类提供不同的 MAP 估计, 在数据平衡时会缩小出现在训练数据中目标类的差距

分类决策会受数据集中占绝对优势的非流失数据的影响.平衡数据的方法:1 重采样稀疏记录 2 取消部分非稀疏记录.

**9:**

如果  $P(X,Y|Z)=P(X|Z)P(Y|Z)$ , 或  $P(X|Y,Z) =P(X|Z)$ , 则称事件 X,Y 对于给定事件 Z 是条件独立的。

$$m \log \frac{p(\theta_e)}{p(\bar{\theta}_e)} + \sum_{i=1}^m \log \left( \frac{p(X_i = x_i | \theta)}{p(X_i = x_i | \bar{\theta})} \right)$$

**10:**

在条件独立性假设成立时, 朴素贝叶斯分类与 MAP 分类的结果相同。从优化的角度看, 条件独立性假设成立时可以减少整个分类器误差率。

当条件独立性假设有效时,朴素贝叶斯分类与 MAP 分类相同.在条件独立性假设有效时,计算工作会更加简单容易.此外,在条件独立性假设成立时,朴素贝叶斯分类与 MAP 分类可得到同样的结果.从减小整个分类器误差率的意义上考虑,朴素贝叶斯方法更优秀.在实践中,条件独立性假设会限制贝叶斯分类器的优化能力.不能盲目使用条件

独立性假设.例如,预测变量的关系性会违背条件独立性假设.

**15:**

朴素贝叶斯的假设需满足条件独立性,而贝叶斯网络允许采用联合条件独立定义子集。条件独立性假设可以如下表示:因此朴素贝叶斯分类为  $\theta_{NB} = \arg \max_{\theta} \prod_{i=1}^m p(X_i = x_i | \theta) p(\theta)$ . 当条件独立性假设有效时,朴素贝叶斯分类与 MAP 分类相同.在使用朴素贝叶斯分类时,需要估计的概率数量相对少得多,仅为  $k \cdot m$ ,而不是 MAP 分类器的  $km$ ,仅需要预测变量数量乘以目标变量的不同值.此外,在条件独立性假设成立时,朴素贝叶斯分类与 MAP 分类可得到同样的结果,而从减小整个分类器误差率的意义上考虑,朴素贝叶斯方法更优.不过,在实践中,条件独立性假设通常会限制贝叶斯分类器的优化能力.不能盲目地使用条件独立性假设,例如预测变量的关联性会违背条件独立性假设.

$$p(X_1 = x_1, X_2 = x_2, \dots, X_m = x_m | \theta) = \prod_{i=1}^m p(X_i = x_i | \theta)$$

**17:**

第一：感兴趣的变量之间的关系是什么？

第二：哪些是有关联的局部概率

第三：如何刻画彼此之前的关系和概率模型