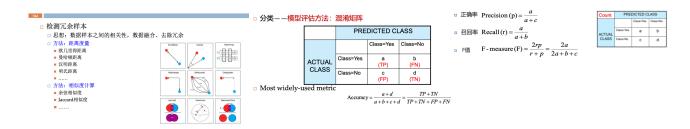
计算题:

・基本概念计算(各种距离、各种相似度、Precision、Recall、ACC、F1)



- ・计算Spearman相关系数、CG、DCG、NDCG
- · PCA和TF-IDF忘了有没有了? , 但也挺重要的, 需要看一看
- ・参数估计---抛硬币
 - ・似然估计
 - ·最大后验估计(Beta分布)



参数估计——最大后验估计

61

- □ 最大后验概率估计(MAP)—理解先验p(θ)
 - □ 扔硬币的例子: 我们期望先验概率(待估计的参数 θ)分布在0.5处取得最大值,也可以选用Beta分布(θ 服从Beta分布)即:

$$p(\theta|\alpha,\beta) \triangleq Beta(\alpha,\beta) = \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

□ 其中,Beta函数是

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

- □ Gamma函数
- $\Gamma(n) = (n-1)!$
- □ Beta分布的随机变量范围是[0,1],不同参数情况下的Beta分布的概率密度函数形式如图



参数估计一最大后验估计

63

- □ 最大后验概率估计(MAP)—理解先验p(θ)
 - □ 扔硬币的例子: 我们期望待估计的参数θ的先验分布在0.5处取得最大值,可以选用 Beta分布(θ服从Beta分布)即:

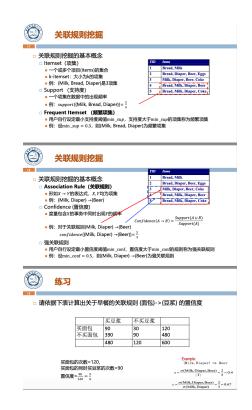
$$p(\theta|\alpha,\beta) \triangleq Beta(\alpha,\beta) = \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

- \square 取 $\alpha = \beta = 5$,使得先验分布Beta分布在0.5处取得最大值
- □ 使用MAP方法求解参数

· 贝叶斯估计



关联规则中的若干概念计算:





A-Priori算法

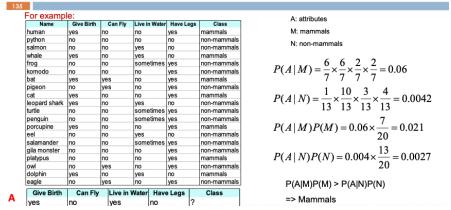
略

决策树(信息熵、信息增益、基尼指数、Error、如何利用这些指标分裂) 略

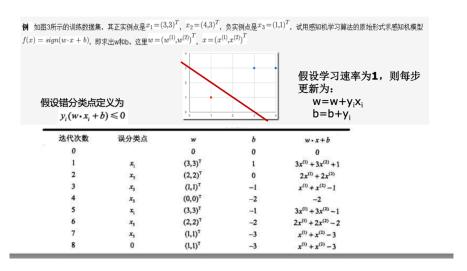
贝叶斯分类器 (类似下图的一道题)



贝叶斯分类器



感知机 (PPT原题♥)



注: ARIMA、ETS、Theta我没印象,疑似是今年新加的内容

简答题:

- ・原题
 - □ 特征工程(**重复迭代**)的流程
 - 1. 对特征进行头脑风暴 深入分析问题,观察数据的基本统计信息,结合问题的相关领域知识和参考其他问题的相关特征工程的方法并应用 到自身的问题中来。
 - **2. 特征的设** —**基础且重要的步骤** 人工设计 存征、自动提取特征,或两者结合,得到模型使用的特征。
 - 3. 特征的选择 使用不同的特征重要性评分方法或者特征选择方法,对特征的有效性进行分析,选出有效的特征
 - 4. 评估模型 利用所选择的特征对测试数据进行预测,评估模型的性能。
 - 5. 上线测试 通过在线。试的效果判断特征是否有效,若不能达到要求,则重复2-5 步骤,直到模型的性能达到要求。

3/6/2025

- ·结构化数据、半结构化数据、非结构化数据的定义与区别
- ・其它忘了

其它:

除了第一章以及太偏机器学习的内容(如SVM)不会考,其它基本概念和方法都可能出现在选择、简答中,不过在我印象中都不难