DM

准确率、精确率、召回率

【机器学习】准确率、精确率、召回率、误报率、漏报率概念及公式_Better Bench的博客-CSDN博客漏报率

ROC曲线和AUC值

机器学习之分类器性能指标之ROC曲线、AUC值 - dzl ML - 博客园

PR曲线

准确率、精确率、召回率

lift曲线

怎么做模型提升度的曲线?

Co-training

机器学习教程 之 半监督学习 Co-training 协同训练 (论文、算法、数据集、代码)_Liangjun_Feng的博客-CSDN博客_协同训练代码(写得太好了)

协同训练算法之co-training_cccloudcloudy的博客-CSDN博客_co-training(课本翻译版本)

PU learning

PU-Learning 原理介绍_一个打码的小年轻的博客-CSDN博客_pu学习

EM算法

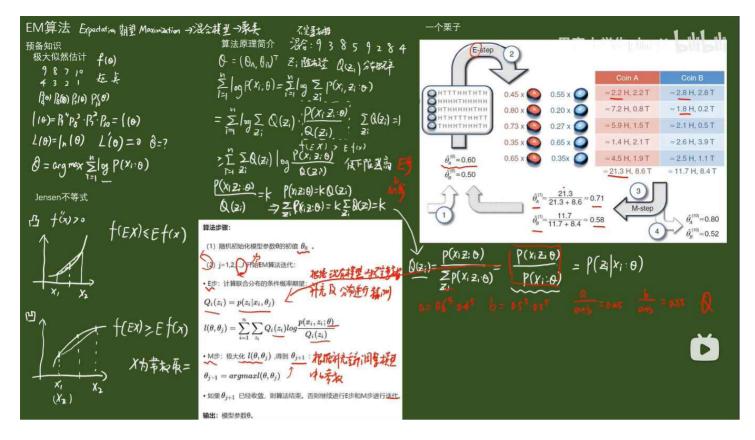
【机器学习】EM——期望最大(非常详细)

快速讲解

什么是 EM 算法(最大期望算法)?【知多少】_哔哩哔哩_bilibili

视频讲解

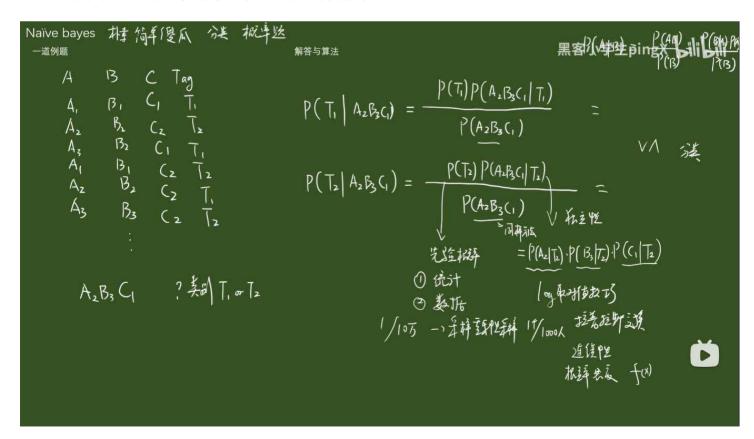
数据挖掘十大算法之十——EM算法 哔哩哔哩 bilibili



3. 使用 EM 的半监督文本分类

朴素贝叶斯分类

二项分布和多项分布 - 小糊涂也学要编程 - 博客园



常见期望和方差

常见分布的 数学期望以及方差公式 二喵君的博客-CSDN博客 期望与方差公式汇总

 $D(X)=E\{[X-E(X)]^2\}$

 $=E\{X^2-2XE(X)+E^2(X)\}$

因为 $E[-2XE(X)]=-2E^2(X)$,所以上式可写成如下:

 $D(X)=E\{X^2-2XE(X)+E^2(X)\}$

 $=E[X^2-2E^2(X)+E^2(X)]$

 $=E[X^2-E^2(X)]$

 $=E(X^2)-E^2(X)$

常见距离和相似度度量

10种距离和相似性度量(简单介绍)

常见的9种距离度量方法-电子发烧友网(优缺点)

全面归纳距离和相似度方法(7种)(包括一些信息论的距离)

SVM

sym原理从头到尾详细推导

【机器学习】支持向量机 SVM(非常详细)

smo:快速理解SMO算法_哔哩哔哩_bilibili

kkt条件KKT条件,原来如此简单 | 理论+算例实践

贝叶斯、极大似然、最大后验

极大似然估计、最大后验估计

拉普拉斯、高斯分布和L1,L2正则

Laplace(拉普拉斯)先验与L1正则化,Gauss先验导出L2正则化 - Sanny.Liu-CV&&ML - 博客园深入理解L1、L2正则化

二项分布和多项式分布

二项分布和多项分布 - 小糊涂也学要编程 - 博客园

多项式分布的理解概率公式的理解 猪逻辑公园的博客-CSDN博客 多项式分布

集成学习

Boosting学习笔记(Adboost、GBDT、Xgboost) - Will的笔记 - 博客园 Bagging和Boosting的总结