

机器学习复习大纲

理解概念最重要！不用知道推导过程！必要的计算一定要会！

6~8 道大题，每道大题 3~5 问。只有简答。一道大题考一个分支，小题 3~5 分，大题 15~20 分

一、评判指标

准确度、召回率的定义

如何根据表格计算准确率和召回率

F 评判指标的含义

P@K 怎么计算

MAP 怎么计算

DCG 怎么计算

NDCG 怎么计算

MRR 怎么计算

二、线性回归模型

目标函数

参数的定义方法是什么

模型的数值形式是什么

目标函数的推导

什么是梯度下降

了解使用梯度下降求解

★学习率对于求解的影响

(为什么减是求最小值?)

为什么要进行空间的转换

了解偏置和方差 (考试不考推导 知道即可)

三、逻辑回归

目标函数的定义

为什么叫做逻辑回归? 和线性回归的区别?

求偏导的方式

如何更新参数

为什么要进行正则化?

四、决策树

给定一个数据集, 如何构造决策树 ✓

决策树解决的任务 (分类) ✓

决策树的基本思路 (总结即可) ✓

信息增益怎么计算 ✓

基尼值和基尼指数是什么 怎么算 ✓

信息熵和条件熵是怎么定义的 ✓

增益率是怎么算的 ✓

为什么要提出增益率 ✓

预剪枝和后剪枝是什么 (思想) ✓

五、支持向量机

支持向量机的目标

支持向量机的基本思想

如何计算出来点到给定线的距离

支持向量机的目标函数

知道如何求解目标函数

等高线和向量场的意义是什么

如何用拉格朗日求解目标函数（拉格朗日乘数）

理解什么叫作支持向量

支持向量的作用

支持向量机的优点和缺点

核函数的功能

核函数的种类及其功能（如高斯核函数）

引入核函数的原因

六、KNN

KNN 是用来做什么任务的

KNN 的优缺点

KNN 的敏感度

七、集成学习

集成学习的动机是什么

集成学习的经典算法

Bagging 算法是什么？

Boosting 算法是什么？

随机森林不需要掌握

无监督聚类的集成学习的基本思想



Soft Correspondence 的目标函数的求解 ?

八、★PCA

一定要了解 PCA !

PCA 的目的是什么 ?

PCA 的假设是什么 ?

PCA 的目标是什么 ? (对应的最优化问题的理解)

PCA 的推导一定要了解清楚 !

最小化误差的推导方式

最大化方差的推导方式

(奇异值分解可以不用管它) ?

它的一些应用和缺陷

九、EM 算法

EM 算法是什么 ?

EM 算法的参数 (观察变量 隐变量 模型参数) ✓

E 步和 M 步分别做什么 ✓

EM 算法的推导过程 ✓

(EM 算法的隐变量用在哪里) ✓

EM 算法不能获得全局最优的原因 ?

EM 算法的应用 (如何结合高斯混合聚类) ✓

十、聚类

C-means 聚类

K-means 聚类是什么？

K-means 的目标函数和原理

K-means 的步骤

K-means 的终止条件

层级聚类的目标函数

层级聚类的思想

层级聚类的步骤

高斯混合聚类的步骤

高斯混合聚类的均值和方差是怎么更新的

十一、推荐系统

推荐系统的目的

推荐系统的应用

推荐系统的步骤

如何基于内容进行分解 基于内容推荐（适合新闻）
用户协同滤波（找相似用户，根据用户评分矩阵计算）
每个商品分数计算
冷启动是什么 优缺点：协同滤波有冷启动的问题

协同过滤的优缺点

如何求解相似用户（相似度）

如何计算相似商品的打分

如何评估商品的好坏

推荐系统的评判指标

（基于检索的评判指标 或者是前述提到的评判指标）

十二、反向传播神经网络

激活函数及其种类

激活函数的意义

目标函数是什么

模型的参数是怎么更新的

输入信号是怎么传播的

误差是怎么传回的

梯度的更新是如何更新的

动量法的含义（为什么要加一个动量）

反向传播神经网络的应用

（公式不需要去记忆，但需要知道公式的含义）

十三、自动编码器

无监督的方法

目标函数是怎么定义的？

它的网络结构是如何定义的？

损失函数是怎么定义的？

自动编码器的两种类型

PCA 和自动编码器的关系

自动编码器的变种算法及其功能原理

用自动编码器设计一个分类器