

2021机器学习复习指导

0. 关于考试

一、闭卷考试

二、题型：选择、填空、问答或计算题

1. 机器学习概论

一、机器学习的定义

二、可以使用机器学习的三个条件，机器学习的基本术语、工作流程（学习出从输入到输出的映射（函数 g ））

2. 机器学习理论

一、霍夫丁不等式

二、假设空间的dichotomy、break point、VC维与成长函数

三、VC维与模型复杂度及模型误差的关系

3.模型优化与验证方法

一、模型选择方法：留出法、交叉验证法、自助法

二、模型性能度量指标

三、过拟合的影响因素及改进方法

四、特征选择方法：过滤式、包裹式、嵌入式

4. 决策树部分

一、决策树的结构和构成

二、三种典型的决策树算法(ID3,C4.5,CART树)的思想, 优缺点 (不需要记信息增益、信息增益比的公式)

三、掌握CART树的划分思路, 针对表格型数据, 掌握分类问题的基尼指数计算公式以及根节点划分准则 (注意要遍历所有属性)

四、预剪枝和后剪枝的对比, 从树的生成、泛化性能、训练时间角度理解

五、不需要掌握缺失值处理方法以及连续属性处理方法。

5. 集成学习部分

- 一、集成学习的框架组成，有哪几部分组成，各部分的作用
- 二、Bagging和Boosting的思想（不需要记伪代码和算法图，不考包外样本）
- 三、Bagging和随机森林与决策树方法相比的优势（掌握思想即可，不需要记理论探讨公式），随机森林与Bagging算法的比较
- 四、Adaboost、GBDT的特点，以树模型为例，通过什么方式生成一颗新的树
- 五、GBDT的例子
- 五、XGBoost目标函数（每部分的意义和作用），不需掌握目标函数构造过程以及切分节点等。

6. 深度学习部分

- 一、训练神经网络的基本步骤(不需要掌握决策边界问题)
- 二、神经网络中的超参数及其作用（网络宽度、深度、激活函数的选择对网络的影响，目标函数和初始化不需要掌握）
- 三、两层神经网络（输入、隐藏层、输出层）的向量化表示（以Sigmoid为激活函数）
- 四、不需掌握反向传播的部分

7. 卷积神经网络部分

一、卷积神经网络的基本结构

二、卷积神经网络的工作原理（感受野、特征图、Kernel、Padding、通道、池化、Flatten）

三、卷积神经网络输入输出维度变化

四、LeNet和CNN的三个特性及其缺陷

五、AlexNet的优点和特点

六、VGG架构，VGG块原理

七、ResNet中残差连接解决的问题

(注：经典的CNN网络结构LeNet、AlexNet、VGG、ResNet不需要掌握论文实验结果，理解思想)

8. 循环神经网络部分

一、为什么需要循环神经网络

二、经典 (Vanilla RNN) 结构

三、分析随时间反向传播BPTT的推导，为什么会
造成梯度消失和梯度爆炸

四、Seq2seq结构以及基础Attention机制

(PPT28页之后，LSTM、GRU等RNN的变体以
及Transformer在工程作业涉及，不在考点范围)

9. 生成对抗网络部分

- 一、生成对抗网络GAN的基本结构和各部分的作用，生成器和判别器的输入和输出
- 二、GAN的目标函数和各部分的意义
- 三、GAN的训练过程（不需掌握全局最优推导）
- 四、传统GAN的缺陷以及其改进方法（WGAN, WGAN-GP）思想

10. 强化学习

一、强化学习中的基本概念：任务与奖赏、马尔科夫决策过程、强化学习的主要任务及与有监督学习的差异

二、探索和利用、 ϵ 贪心算法过程

三、有模型学习中的值迭代与策略迭代的算法过程及差异

四、免模型学习中的蒙特卡罗算法思想

11. 半监督学习

一、半监督学习的基本概念和假设

二、生成式方法、半监督SVM对应的建模方法、
优化过程及算法

三、约束K-均值聚类算法过程

12. 迁移学习

一、迁移学习问题的概念

二、域适应的概念

三、最大均值差异法的过程（不需要掌握具体的推导）

四、深度迁移学习的方式（掌握概念）

五、例子中的实现形式（掌握思路）

13. 小样本学习

一、小样本学习问题的定义及算法分类

二、基于模型优化的基本思想及类型

三、元学习