



# 期末复习重点

选择题28分+简答题72分

小样本、持续、自监督**不考**

无监督、贝叶斯**无大题**

原则：不需要精确计算，只需要相对比较

大作业DDL：7.31

概论（概念多，理解概念）

模型评估与选择（概念多，理解概念）

线性模型

决策树（肯定重要，做好心理准备）

KNN

高斯混合模型（比较重要，多多掌握）

支持向量机（重要）

PCA（重要）

贝叶斯分类器

集成学习（比较简单）

无监督学习

神经网络（必考内容）

## 概论（概念多，理解概念）

机器学习系统——选择题：定义、组成、过程、理论基础

归纳偏好

统计学习的基本概念：统计量、凸函数、高斯分布 etc.

## 模型评估与选择（概念多，理解概念）

概念：统计方法三要素、损失函数、过拟合与欠拟合

模型选择、评估（指标）、数据集划分

偏差-方差分解

## 线性模型

概念：线性可分

多元线性回归（不考计算、证明等）

对数几率回归（不会有复杂的公式推导与计算，考试会给信息）

线性判别分析LDA（花时间自己掌握与学习，比较重要）

## 决策树（肯定重要，做好心理准备）

最优划分属性：如何选择、信息增益、增益率、基尼指数

剪枝：为什么、如何做（预剪枝、后剪枝及其优缺点、实际问题如缺失值与连续空间离散化）

## KNN

K的选择

最近邻分类器（k=1）：泛化的理论分析

## 高斯混合模型（比较重要，多多掌握）

多变量高斯分布的概率密度函数、统计分布意义等

混合高斯分布：计算比例系数和分布参数

EM算法（不要求理论推导和证明）

## 支持向量机（重要）

线性模型 → 最大化最小间隔 → 支持向量机

margin 的计算

拉格朗日乘子法与KKT条件

非线性SVM

kernel trick 定义、限制条件以及得到的 kernel SVM

kernel function, classification

多类支持SVM (比较简单)

## PCA (重要)

原因：维度灾难

方法：线性投影

公式、推导与求解

子空间选择

核化PCA

## 贝叶斯分类器

掌握基础概念与基础公式

## 集成学习 (比较简单)

原理：“好而不同”

Bagging

原理：有放回采样

并行

特点：降低方差，误差由基学习器的 bias 决定

代表算法：随机森林

Boosting

串行

思想：后一个纠正前一个

代表算法：AdaBoost (算法解释不用花时间深究)

GBDT (omitted)

学习法：Stacking (简单了解与掌握知识点)

# 无监督学习

- 不考大题，比较简单

不需要花太多时间复习具体算法，只需要掌握具体概念，如何做，为何这样做

小知识：高斯混合聚类，与EM算法很有关系，**知识串联**

## 神经网络（必考内容）

起源：MP

单层感知机 → 多层感知机

BP（推导仅了解）

卷积计算

动机：降低参数量

基本知识

池化（下采样）

作用：减少参数、避免过拟合、扩大感受域

基本知识

深度学习技巧

标准化（若干种）

经典网络架构（有一点点了解就好，常见神经网络架构如resnet, transformer的基本思想）