

认准一手资源 VX: rgzn321 其他均为翻录 需要其他人工智能资源加 V

学习第 3 章 k 近邻

2019-05-25 00:00 至 2019-05-25 23:59

前言：

教学内容完全依据《统计学习方法》一书，每一章的学习计划会详细列出本章需要学习的内容和不做学习要求的内容。

每章会有一个导读视频和两个重难点讲解视频（第三章和第五章内容比较容易理解，第三章没有重难点讲解视频，第五章只有一个重难点讲解视频），重难点讲解视频涉及模型理解、理论推导、算法实现等内容。书中共有 10 个算法，每个算法会在学习当天提供一个示例代码，同学们先自行理解，助教会在每周周末讲解作业和代码。

资料领取：《统计学习方法》电子书请在公众号**深度之眼**后台直接回复关键词【统计】，即可领取。

Week1

Day7

任务简介：学习第 3 章 k 近邻，学习 k 近邻算法在分类问题中的应用，理解 k 近邻法的三要素及模型对应的损失函数。

详细说明：第 3 章讲了如何用 k 近邻算法进行分类。同学们通过学习第 1 节，理解 k 近邻算法的基本思想；通过学习第 2 节，掌握 k 近邻算法如何实现；第 3 节是一个搜索技术，关于如何对某一训练数据快速找到相邻的 k 个示例，这里不做学习要求。

学习目标：

- 1、导读视频



《统计学习方法》-K 近邻



- 1.掌握 k 近邻算法的原理。
- 2.理解 k 近邻算法三要素及模型对应的损失函数。
- 3.掌握 k 近邻算法在分类问题上的求解过程。

作业 3:

1.思考 k 近邻算法的模型复杂度体现在哪里？什么情况下会造成过拟合？打卡进行文字说明。

2.给定一个二维空间的数据集 $T=\{\text{正实例: } (5,4), (9,6), (4,7); \text{负实例: } (2,3), (8,1), (7,2)\}$ ，试基于欧氏距离，找到数据点 $S(5,3)$ 的最近邻 ($k=1$)，并对 S 点进行分类预测。

(1)用“线性扫描”算法自编程实现。

(2)试调用 `sklearn.neighbors` 的 `KNeighborsClassifier` 模块，对 S 点进行分类预测，并对比近邻数 k 取值不同，对分类预测结果的影响。

(3)思考题：思考“线性扫描”算法和“kd 树”算法的时间复杂度。

打卡代码运行结果的截图。

作业答案及代码讲解在下周日公布，助教会进行视频讲解。