了解机器学习中特征工程和模型评估

2019-06-25 00:00 至 2019-06-26 23:59

前言：

人工智能技术正在对社会结构、职场、教育等带来革命性的变化。 未来几年是人工智能技术全面普及化的时期，也是该技术的相关人才最为稀缺的时期。我们希望通过面试刷题班，帮助对人工智能和机器学习感兴趣的朋友更加深入地了解这个领域的基本技能，帮助已经有计算机技术基础的同行们，成为AI领域的offer收割机。

知识要求：(学习的过程中可以遇到问题后再复习）

了解高等数学、概率论、线性代数基础知识，若数学基础薄弱可向任意深小享回复【数学基础】学习数学基础训练营内容。

了解基础机器学习算法，若基础机器学习算法不熟悉建议可参考西瓜书进行适当复习，若完全不了解可向任意深小享报名西瓜书训练营。

资料下载：

《百面机器学习：算法工程师带你去面试》：

电子版链接：https://share.weiyun.com/5k9AYXR 密码：swqwd9

本日任务简介：

1、学习绪论视频，了解预备知识

2、学习《百面机器学习》1.1-1.7，2.1-2.7

今天学习任务比较简单，因此只有书籍阅读任务，无补充图文或视频教程。之后的难点任务，均会有老师录制成讲解视频。并且每周六老师会对当周疑难问题进行直播讲解。

本节将会向大家介绍特征工程和模型评估当中的一些知识点。特征工程是对原始数据进行一系列工程处理，将其提炼为特征，作为输入供算法和模型使用。做过项目或者竞赛的都应当了解特征工程的重要性，面对不在一个数量级的特征，类别性特征，高维特征，特征组合等等场景，我们应该怎么做。除此之外，我们还需要对自己领域内，比如自然语言处理，图像处理等有哪些特征工程的技巧有所了解。

模型评估是构建一个模型重要的一环，分类问题、排序问题、回归问题往往需要使用不同的指标进行评估。在这里我们将着重攻克常见的指标，如精确率，召回率，F1值，P-R曲线，ROC曲线和AUC。通过指标我们观察到欠拟合和过拟合时，该如何处理。

打卡要求：说明对于不同的特征该如何进行特征工程？模型评估中不同的指标用在什么场景中？

打卡内容：不少于150字

打卡截止提交时间：6/26

说明对于不同的特征该如何进行特征工程？模型评估中不同的指标用在什么场景中？