

常见面试问题

特征处理、共线性、内生性

小胖

目录

ONE 数值型特征

TWO 类别型特征

THREE 共线性、内生性

数值型特征

面试题目：为什么需要对数值类型的特征做归一化？

精通数据科学：
从线性回归到深度学习

难度：★★★★☆

解答要点：

- 多元函数n阶泰勒展开式的数学公式和推导
- 梯度下降法的算法思路和具体的数学推导
- 多重共线性对模型的影响以及它的分类
- 结构性多重共线性的产生原因以及解决方法

涉及课程：

- 6_1 如何算出模型参数的估计值
- 7_5 多变量的烦恼：多重共线性

精通数据科学：
从线性回归到深度学习

精通数据科学：
从线性回归到深度学习

精通数据科学：
从线性回归到深度学习

数值型特征

面试题目：什么是数值特征离散化？为什么需要对数值特征做离散化？

精通数据科学：
从线性回归到深度学习

难度：★★★★☆

解答要点：

- 直接使用数值型特征的隐含假设和陷阱
- 数值型特征离散化的常用方法以及各方法的优缺点
- 数值型特征离散化之后对模型训练的好处

涉及课程：

- 6_1 如何算出模型参数的估计值
- 7_1 如何在模型里使用定性变量
- 7_3 如何更有效地使用定量变量

精通数据科学：
从线性回归到深度学习

精通数据科学：
从线性回归到深度学习

精通数据科学：
从线性回归到深度学习

精通数据科学：
从线性回归到深度学习

目录

ONE 数值型特征

TWO 类别型特征

THREE 共线性、内生性

类别型特征

面试题目：在对数据进行预处理时，应该如何处理类别型特征？

难度：★★★★☆

解答要点：

- 类别型变量的分类以及在模型里无法直接使用类别型变量的原因
- 虚拟变量的定义、虚拟变量陷阱以及解决方法
- 其他常用的类别型特征编码方式，比如二进制编码、Helmert Contrast、Sum Contrast、Polynomial Contrast

涉及课程：

- 7_1 如何在模型里使用定性变量
- 7_5 多变量的烦恼：多重共线性

目录

ONE 数值型特征

TWO 类别型特征

THREE 共线性、内生性

共线性、内生性

面试题目：线性回归模型对数据的关键假设有哪些？这些假设不满足时，会如何影响模型效果？相应的解决方法？

难度：★★★★☆

解答要点：

- 线性回归模型对数据的假设
- 多重共线性的定义、对模型的影响以及相应的解决方法
- 内生性的定义、对模型的影响以及相应的解决方法

涉及课程：

- 4_3 统计眼中的线性回归模型
- 7_5 多变量的烦恼：多重共线性
- 7_7 Y的变化来自何处：内生性

THANK YOU

精通数据科学：
从线性回归到深度学习