

课程概述

您好！欢迎来到数据分析中级课程，本课程是中级部分的第二个案例：新三板公司融资金额分析。本案例所使用的数据集与案例一中的数据集为同一套数据。

1. 内容简介

1) 问题界定

通过对“新三板”根本任务出发，不在拘于对于交易发生的预测分析，而是从融资金额的角度，构造主营业务问题。

2) 数据准备

通过对重要字段的检视与处理，提出适用于本案例的数据改造方案；
并根据数据改造方案完成数据预处理及数据分割等操作，形成满足数据建模要求的数据集。

本案例数据准备中的内容与上个案例大致相同，但也增加了一些新的内容。

3) 数据建模与数据可视化

使用线性回归模型，通过网格搜索超参数、模型评估等方法构建最终模型；
输出模型结果并结合业务进行较为细致的分析。

2. 学习目标

学习完本次案例，你能够达到以下目标：

- 1) 能够掌握流水线结构，优化逻辑回归中数据预处理的代码；
- 2) 能够使用线性回归模型（套索回归）进行预测分析；
- 2) 能够从逻辑上，梳理出有监督模型的一系列建模过程；
- 3) 能够从代码实现上，梳理出有监督模型建模的代码结构。

业务问题构造

新三板作为我国非上市股份有限公司的股权交易平台，最根本任务是帮助中小微公司解决融资难的问题。本次案例将围绕“新三板”部分公司数据，挖掘公司融资金额背后的相关因素。

1. 业务问题

上一个案例中通过对公司这一群体的行为信息特征及企业背景进行分类分析，帮助投资人更加了解“新三板”市场，发现优质公司及预测公司未来走向，而在本案例中对于“新三板”市场的中小企业，他们的主要目标是**融资**。

融资是一个企业的资金筹集的行为与过程，是通过各种方式到金融市场上筹措或发放资金的行为，而一个公司的融资能力可以理解为获得贷款的能力，它是反映公司多方面的业务水平，也是银行等贷款方对公司的评估依据之一。

因此本案例基于公司数据，不再拘于对是否交易进行预测，而是预测公司的融资金额，为公司可以有效把握银行等债权方提供预测支持。

基于此，本次案例的主业务问题是：

围绕新三板部分公司数据，挖掘公司融资金额背后的相关因素。

对于金额这类连续型数据，常用的预测方法有一大类是回归模型，本次案例我们选择线性回归模型。