

数据科学项目实践

项目背景

在日常生活中，时间的痕迹无处不在。股票交易、气温、水质、PM25等数据在时间的横轴上不断变化，或有周期性，或无。具体到商业分析、数据科学中，时间序列数据亦无处不在，例如QQ日活变化、商品日销量变化、实体店日访客人数变化等等。本次实践，将从时间序列分析着手，结合真实数据，充分利用统计学习方法、深度学习技术，与同学们一起探索神秘的时间序列世界。

项目内容

为了提升同学们对时间序列预测技术的理解，要求对常见的统计学习方法SVR、~~逻辑回归~~，经济学方法常用分析方法ARIMA、LSTM深度学习方法全面掌握。

项目实践所用数据选取了美国某上市公司真实股票交易数据，在数据中你可以看到7个字段，分别为日期、当日开盘价、最高价、最低价、收盘价、交易量和OpenInt（忽略该字段），如下所示：

| Date | Open | High | Low | Close | Volume | OpenInt |
|------------|--------|--------|--------|--------|---------|---------|
| 1962-01-02 | 0.6277 | 0.6362 | 0.6201 | 0.6201 | 2675579 | 0 |

项目要求

- 分别绘制上述数据当日最高、交易量曲线图
- 分别利用SVR、~~逻辑回归~~、ARIMA和LSTM对上述进行拟合预测
- 计算每种机器学习方法在该数据集中的表现，评估指标MAE和MSE
- 总结本次项目，写出项目实践报告

项目交付

- 项目时间：20200205-20200212
- 项目规划：建议1-3天调研算法和时间序列相关技术，并写出调研报告；建议1-3天完成SVR、逻辑回归和ARIMA方法；建议1-3天完成LSTM和项目实践总结。
- 项目沟通：遇到问题及时通过微信与我沟通。