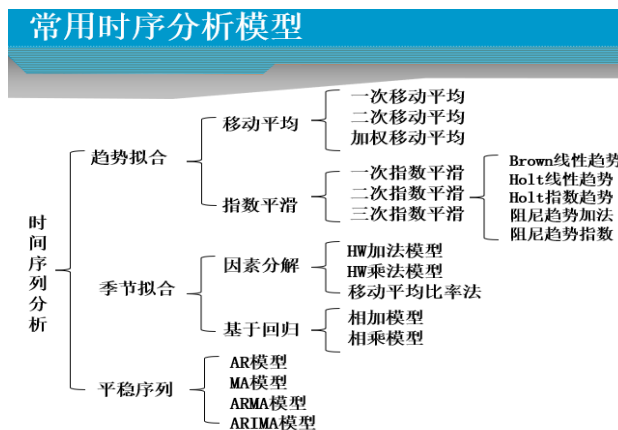


Python 数据建模（时序模型篇）

【课程目标】

本课程为中级课程《大数据建模》的第二篇：时序篇

本课程主要讲解如何利用 Python 进行时间序列的数据建模，帮助学员构建系统全面的预测建模思维，提升学员的数据建模综合能力。



本系列课程从实际的业务需求出发，结合行业的典型应用特点，围绕实际的商业问题，对数据预测建模的过程进行了全面的介绍（从模型选择，到属性选择，再到训练模型，评估模型以及优化模型），通过大量的操作演练，帮助学员掌握数据建模的思路、方法、技巧，以提升学员的数据建模的能力，支撑运营决策的目的。

通过本课程的学习，达到如下目的：

- 1、掌握数据建模的标准流程。
- 2、掌握时序预测建模的基本思想，理解因素分解的思路。
- 3、掌握常用的趋势拟合模型。
- 4、掌握常用的季节预测模型，能够进行季节周期性的时序建模。
- 5、深刻理念平稳序列的概念，以及平稳性检验。
- 6、掌握平稳序列的模型识别，以及模型定阶。
- 7、掌握时序预测模型的评估，以及优化。
- 8、掌握高级时序模型的训练与建模。

【授课时间】

2 天时间

(要根据学员的实际情况调整重点内容及时间)

【授课对象】

业务支撑部、运营分析部、数据分析部、IT 系统部、大数据系统开发部等相关技术人员。

【学员要求】

- 1、 每个学员自备一台便携机(必须)。
- 2、 便携机中事先安装好 Python 3.9 版本及以上。
- 3、 安装好 Numpy,Pandas,statsmodels,sklearn,scipy,keras 等扩展库。

注：讲师现场提供分析的代码和数据源。

【授课方式】

建模流程+ 案例演练 + 开发实践 + 可视化呈现

采用互动式教学，围绕业务问题，展开数据分析过程，全过程演练操作，

让学员在分析、分享、讲授、总结、自我实践过程中获得能力提升。

【课程大纲】

第一部分：预测建模基础

1、数据建模六步法

- 选择模型：基于业务选择恰当的数据模型

- 属性筛选：选择对目标变量有显著影响的属性来建模
- 训练模型：采用合适的算法，寻找到最合适的模型参数
- 评估模型：进行评估模型的质量，判断模型是否可用
- 优化模型：如果评估结果不理想，则需要对模型进行优化
- 应用模型：如果评估结果满足要求，则可应用模型于业务场景

2、数据挖掘常用的模型

- 数值预测模型：回归预测、时序预测等
- 分类预测模型：逻辑回归、决策树、神经网络、支持向量机等
- 市场细分：聚类、RFM、PCA 等
- 产品推荐：关联分析、协同过滤等
- 产品优化：回归、随机效用等
- 产品定价：定价策略/最优定价等

3、属性筛选/特征选择/变量降维

- 基于变量本身特征
- 基于相关性判断
- 因子合并 (PCA 等)
- IV 值筛选 (评分卡使用)

- 基于信息增益判断（决策树使用）

4、训练模型及实现算法

- 模型原理
- 算法实现

5、评估模型

- 评估指标
- 评估方法
- 残差评估

6、模型优化

- 优化模型：选择新模型/修改模型
- 优化数据：新增显著自变量
- 优化公式：采用新的计算公式

7、模型应用

- 模型解读
- 模型保存/加载
- 模型应用/预测

8、好模型是优化出来的

第二部分：时序模型评估

1、评估指标

- 判定系数： R^2 和 \hat{R}^2
- 平均误差：MAE
- 根均方差：RMSE
- 平均误差率：MAPE

2、信息准则指标

- 赤池信息准则 (Akaike Information Criterion , AIC)
- 贝叶斯信息准则 (Bayesian Information Criterion , BIC)
- HQIC (Hannan-Quinn Information Criterion , HQIC)

3、评估方法

- 滚动交叉验证法 (cross validation)

4、其它评估

- 残差评估：白噪声评估

第三部分：趋势预测模型

问题：无法找到影响因素，无法回归建模，怎么办？随着业务受季节性因素影响，未来的销量如何预测？

1、时间序列简介

2、时序预测的原理及应用场景

3、常见时序预测模型

- 趋势类预测模型
- 季节类预测模型
- 平稳时序预测模型
- 其它高级模型

4、移动平均

- 应用场景及原理
- 理解滑动窗口
- 移动平均种类
 - ◇ 一次移动平均
 - ◇ 二次移动平均
 - ◇ 加权移动平均

- ◇ 移动平均比率法

- 移动平均关键问题

- ◇ 最佳期数 N 的选择原则

- ◇ 最优权重系数的选取原则

演练：销售额预测模型及评估

演练：快销产品季节销量预测及评估

5、指数平滑

- 应用场景及原理

- 最优平滑系数的选取原则

- 指数平滑种类

- ◇ 一次指数平滑

- ◇ 二次指数平滑 (Brown 线性)

- ◇ 三次指数平滑

演练：煤炭产量预测

6、Holt 趋势模型 (亦称二次指数平滑)

- Holt 线性模型

- Holt 指数模型

- 阻尼线性趋势

- 阻尼指数趋势

第四部分：季节预测模型

1、因素分解思想

2、时间序列的四个构成要素

- 长期趋势 Trend

- 季节变动 Seasonality

- 循环变动 Circle

- 不规则变动 Irregular

案例：时间序列的季节分解

3、Holt-Winters 季节模型

- 三个组成部分

- 三个平滑因子

4、HW 加法模型

- 适用场景

- 计算公式

- 超参优化

- 模型解读

5、HW 乘法模型

6、HW 指数模型

案例：航空飞行里程预测模型

案例：汽车销量预测模型

案例：沃尔玛收益预测模型

7、基于回归的季节模型

- 相加模型

- 相乘模型

- 模型训练及优化

- 模型解读

第五部分：平稳序列模型

1、平稳序列预测模型简介

2、序列平稳性概念

- 恒定的均值

- 恒定的标准差
- 与位置无关的协方差

3、序列平稳性检验

- 折线图法
- ACF/PACF 图
- ADF 检测法

4、特殊平稳序列：白噪声

案例：序列平稳性检验

案例：白噪声检验

5、平稳序列常用拟合模型

- AR(p)自回归模型
- MA(q)移动平均模型
- ARMA(p,q)自回归移动平均模型

6、模型识别

- ACF 图
- PACF 图

7、模型定阶

- 图形定阶 (ACF/PACF)

- 最小信息准则定阶

8、非平稳序列处理

- 平滑法

- 变量变换

- 差分运算：k 步差分与 d 阶差分

9、ARIMA(p,d,q)建模流程

案例：上海证券交易所收盘价建模

10、 SARIMA(p,d,q)(P,D,Q,S)模型

- 图形确定阶数

- 遍历确定阶数

11、 时序模型总结

第六部分：模型质量评估篇

1、 回归模型的评估指标

- 三个基本概念:SSR/SST/SSE

- 两个判定系数: R^2 ,调整 R^2

➤ 三个误差指标:MAE/MAPE/RMSE

◇ 平均绝对误差 MAE

◇ 均方差 MSE/RMSE

◇ 平均误差率 MAPE

2、 模型的评估方法

➤ 原始评估法

➤ 留出法 (Hold-Out)

➤ 交叉验证法 (k-fold cross validation)

➤ 自助采样法 (Bootstrapping)

3、 时间序列的滚动交叉验证

第七部分：高级时序模型

1、 Prophet 模型介绍

➤ 趋势拟合

➤ 季节性预测

➤ 节假日和特殊事件的影响

➤ 离群值分析

案例：销售额时序预测模型

2、LSTM 模型简介

- 数据集构造
- 形状构造
- 滚动预测

3、实战：广告点击量时序建模

结束：课程总结与问题答疑。