NJUAI 南京大学 人工智能学院
SCHOOL OF ARTIFICIAL INTELLIGENCE, NANJING UNIVERSITY
南京大學 NANJING UNIVERSITY 1902

01 绪论

时间序列的基本概念和基本分析思路

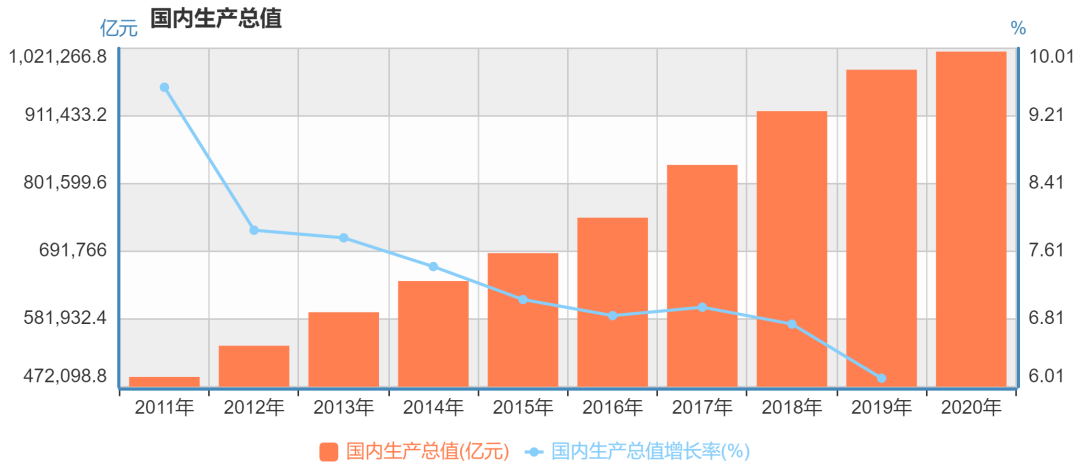
本章概要

1. 时间序列研究的主要思路

2. 统计方法和机器学习方法

3. 时间序列的发展历史和流派

时间序列分析的可能思路

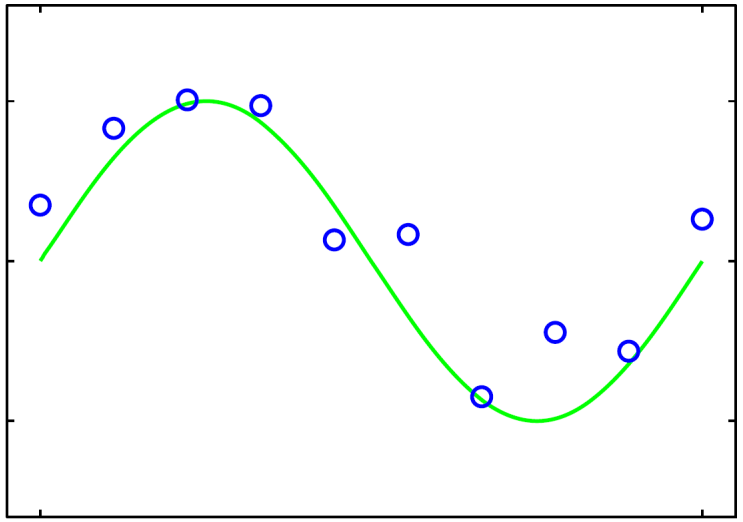


画图观察 → 描述性时间序列分析

提取特征 → 统计时间序列分析

时间序列分析和回归

通过回归方法
分析时间序列



回归对样本有独立性假设，时间序列分析利用数据相关性（惯性）分析预测

二者在某些方面有相似处，如时间序列分析中也会使用最小二乘方法进行参数估计

时间序列与传统机器学习的假设

传统机器学习

- 给定示例（属性）$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$，以及其对应的标记$y_1, y_2, \ldots, y_n$，需要学习示例到标记的映射

$$\hat{y} = f(\boldsymbol{x})$$

  - $\boldsymbol{x}_i \in \mathbb{R}^d$
  - 假设样本独立同分布，

$$(\boldsymbol{x}_i, y_i) \sim \Pr(\boldsymbol{x}, y)$$

- 如果标记y为**实数**，则为回归问题
- 如果$\boldsymbol{x}$是一个时间序列，则转化为时间序列分类问题

时间序列（预测）

- 给定时间序列上下文属性$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$，以及其对应的行为属性$y_1, y_2, \ldots, y_n$，以及$\boldsymbol{x}_{n+1}$，需要对$y_{n+1}$进行预测，即学习

$$\hat{y}_{n+1} = f\left(\boldsymbol{x}_{n+1}; \begin{matrix} \boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n \\ y_1, y_2, \ldots, y_n \end{matrix}\right)$$

  - $\boldsymbol{x}_i \in \mathbb{R}^d$
  - 样本之间存在依赖关系，后续样本依赖以往样本，基于不同的上下文属性有不同的概率假设

$$(\boldsymbol{x}_i, y_i) \sim \Pr_i(\boldsymbol{x}, y)$$

  - 忽略依赖关系则转化为回归问题

时间序列（预测）分析的“流派”

- 统计时间序列分析（时域分析）
  - 基于统计学方法对时间序列的概率特性进行估计，一般基于（线性）随机过程假设
    - AR模型
    - MA模型
    - ARMA模型
    - ARIMA模型
    - …

- 统计时间序列分析（频域分析）
  - 任何一种无趋势的时间序列都可以分解为若干不同频率的周期波动
  - 可视为对时间序列的特征提取
  - DFT
  - DWT
  - …

- 机器学习方法分析
  - 基于机器学习方法从历史或相关序列中学习时间序列预测的方法
  - 基于近邻的方法
  - RNN/LSTM
  - Transformer
  - TCN
  - Meta-Learning
  - …

本章概要

1. 时间序列研究的主要思路

2. 统计方法和机器学习方法

3. 时间序列的发展历史和流派

时间序列的统计定义

- 随机序列: 按时间顺序排列的一组**随机变量**

$$\cdots, Y_1, Y_2, \cdots, Y_t, \cdots$$

- 观察值序列: 随机序列的n个有序观察值，称之为序列长度为n的观察值序列

$$y_1, y_2, \cdots, y_n$$

- 随机序列和观察值序列的关系
  - 观察值序列是随机序列的一个实现
  - 研究的目的是想揭示随机时序的性质
  - 实现的手段都是通过分析观察值序列的性质，由观察值序列的性质来推断随机序列的性质

一个时间序列预测的例子

- 对小麦价格序列$\{y_t\}$进行预测，

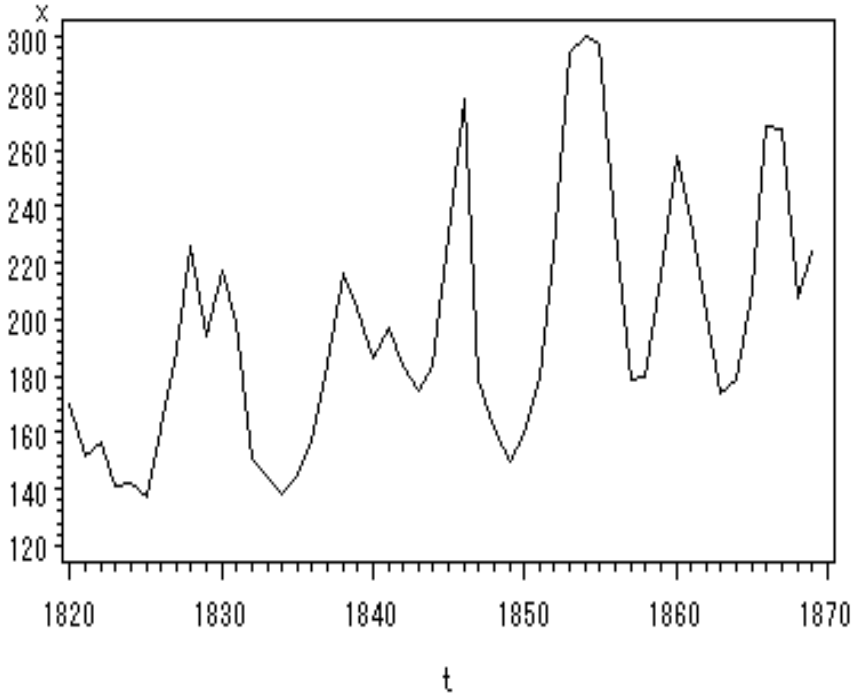
- 给定Y_t，如何预测下一个月的小麦价格?
  - 假定价格**不变化**

$$\hat{y}_{t+1|t} = y_t$$

  - $t+1$月份的小麦真实价格为y_{t+1}，则预测残差为

$$\epsilon_{t+1} = y_{t+1} - \hat{y}_{t+1|t}$$

$$Y_t = Y_{t-1} + \epsilon_t$$



贝弗里奇小麦价格指数序列

一个时间序列预测的例子：随机模型

- 建立小麦价格序列的随机模型（stochastic model）

$$Y_t = Y_{t-1} + \epsilon_t$$

  - 假设$\{\epsilon_t\}$是均值为0互不相关的随机变量（白噪声），则该模型为随机游走（random walk）模型
  - 展开为

$$Y_t = \epsilon_t + \epsilon_{t-1} + \cdots \quad \Rightarrow \quad \text{不稳定（non-stationary）序列}$$

- 对上述模型进行扩展

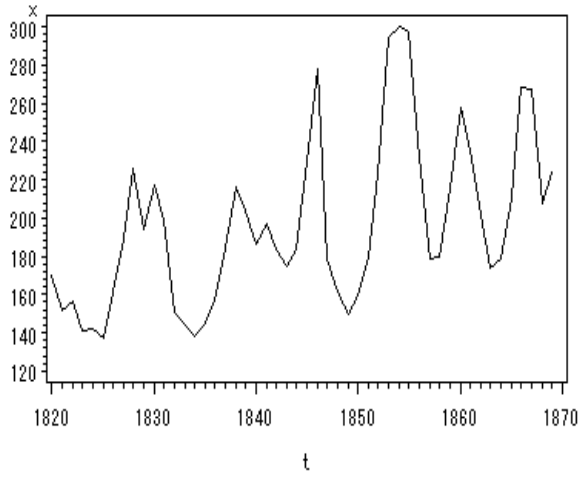
$$Y_t = \varphi Y_{t-1} + \epsilon_t \quad \Rightarrow \quad |\varphi| < 1\text{时为稳定（stationary）序列}$$

  - 该模型为$AR(1)$模型（autoregressive first order），$\varphi = 1$时变化为随机游走模型

一个时间序列预测的例子：季节因素

- 建立小麦价格序列的随机模型

$$Y_t = \varphi Y_{t-1} + \epsilon_t$$



贝弗里奇小麦价格指数序列

- 考虑季节因素影响
  - 环比：基于Y_{t-1}进行预测
  - 同比：基于Y_{t-12}进行预测

$$Y_t = \psi Y_{t-12} + \epsilon_t$$

$$Y_t = \varphi Y_{t-1} + \psi(Y_{t-12} - \varphi Y_{t-13}) + \epsilon_t$$

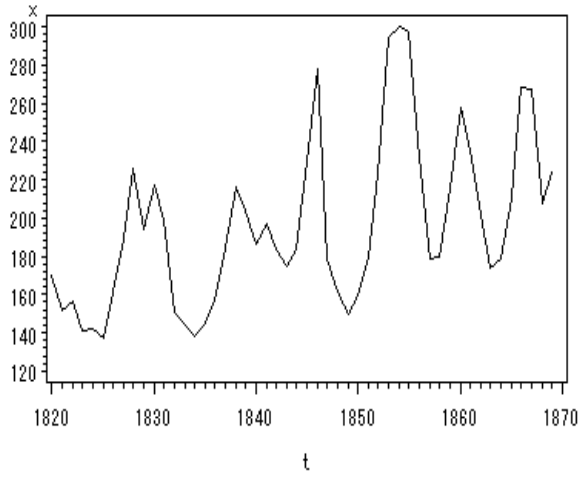
一个时间序列预测的例子：季节因素

- 考虑季节因素影响
  - 同比：基于Y_{t-12}进行预测

$$Y_t = \psi Y_{t-12} + \epsilon_t$$

  - 环比：基于Y_{t-1}进行预测

$$Y_t = \varphi Y_{t-1} + \epsilon_t$$



贝弗里奇小麦价格指数序列

$$Y_t = \varphi Y_{t-1} + \psi(Y_{t-12} - \varphi Y_{t-13}) + \epsilon_t$$

$$(Y_t - \varphi Y_{t-1}) - \psi(Y_{t-12} - \varphi Y_{t-13}) = \epsilon_t$$

⇨ $\psi = 0$，退化为$AR(1)$模型

⇨ $\varphi = 0$，退化为基本季节模型

时间序列分析的主要思路

$$\sum_{i=0}^{N} a_i y[n-i] = \sum_{j=0}^{M} b_j x[n-j]$$

- 银行理财：当前存款为$x[n]$，当前账户总额为$y[n]$，a_i为存款利率，则

$$y[n] = y[n-1] + a_i y[n-2] + x[n]$$

- 对于n时刻的输出仅仅和n之前时刻的输出相关的模型，成为AR (Auto Regressive)模型

- 空谷回声：$x[n]$为当前时刻发出的声音，$y[n]$为当前时刻听到的声音，b_k刻画声音传播过程中的衰减

$$y[n] = x[n] + b_k x[n-k]$$

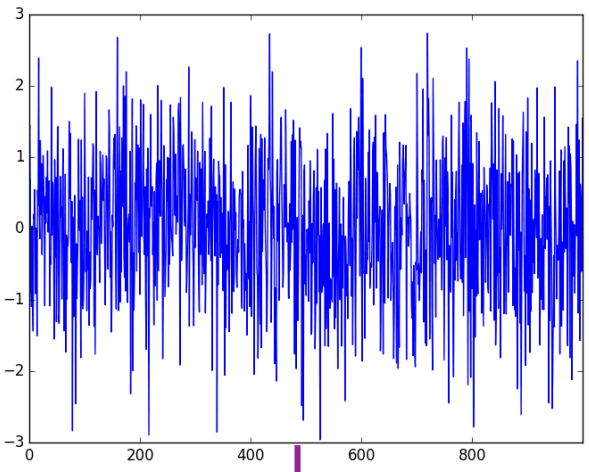
- 对于n时刻的输出仅仅和n之前时刻的输入相关的模型，成为MA (Moving Average)模型

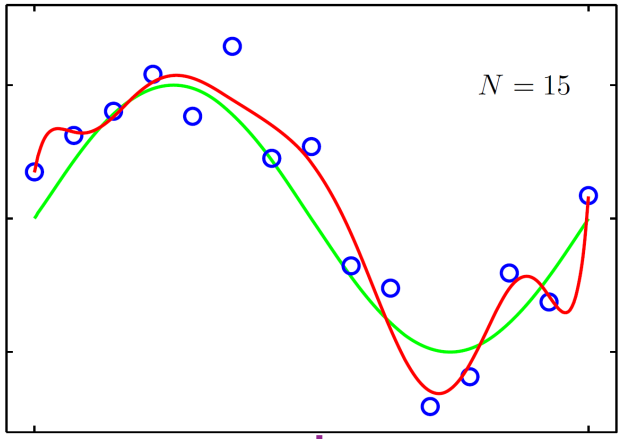
时间序列分析的几个基本问题

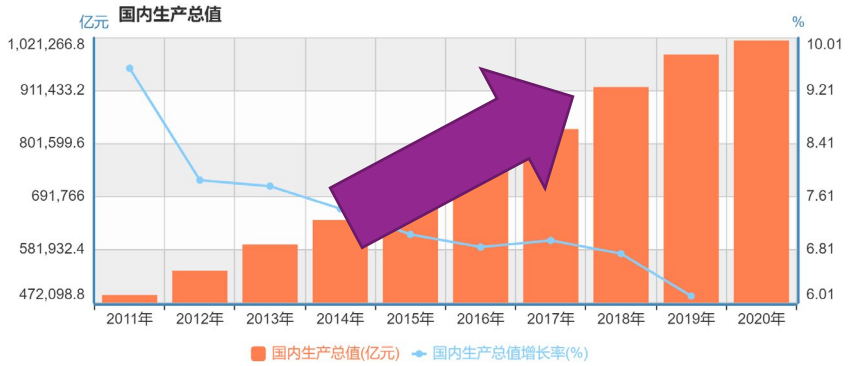
什么样的序列需要分析？

如何才能使得分析可行？

要分析出什么？





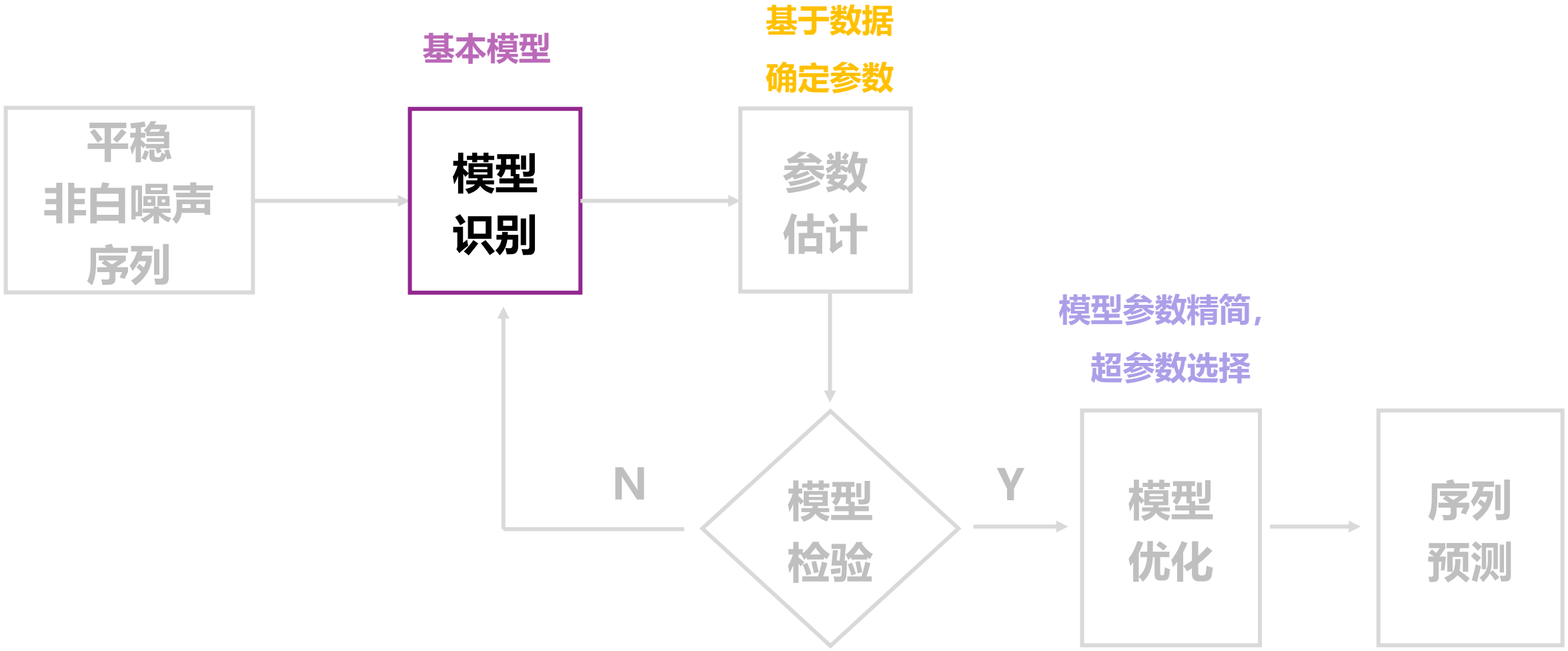


需要鉴别是否
纯随机序列

观测样本丰富，
模型简单

周期、趋势…

时间序列分析的主要思路



机器学习方法的主要思路

- 时间序列（预测）

- 给定时间序列上下文属性$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$，以及其对应的行为属性$y_1, y_2, \ldots, y_n$，以及$\boldsymbol{x}_{n+1}$，需要对$y_{n+1}$进行预测，即学习

$$\hat{y}_{n+1} = f\left(\boldsymbol{x}_{n+1}; \begin{matrix} \boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n \\ y_1, y_2, \ldots, y_n \end{matrix}\right)$$

  - $\boldsymbol{x}_{\boldsymbol{i}} \in \mathbb{R}^d$
  - 样本之间存在依赖关系，后续样本依赖以往样本，基于不同的上下文属性有不同的概率假设

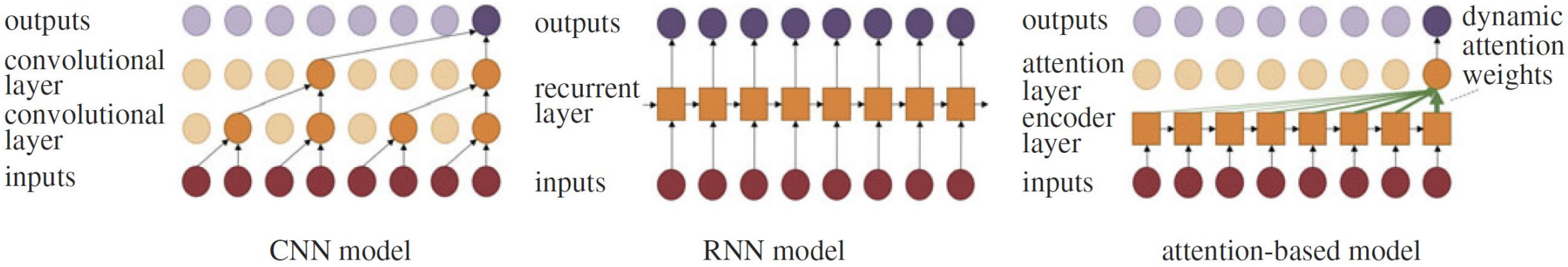
$$(\boldsymbol{x}_i, y_i) \sim \mathrm{Pr}_i\,(\boldsymbol{x}, y)$$

  - 忽略依赖关系则转化为回归问题

关键在于如何从历史数据中学习出f

深度学习实现时间序列分析映射

$$\hat{y}_{n+1} = f\left(\boldsymbol{x}_{n+1};\ \begin{matrix}\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n \\ y_1, y_2, \dots, y_n\end{matrix}\right)$$

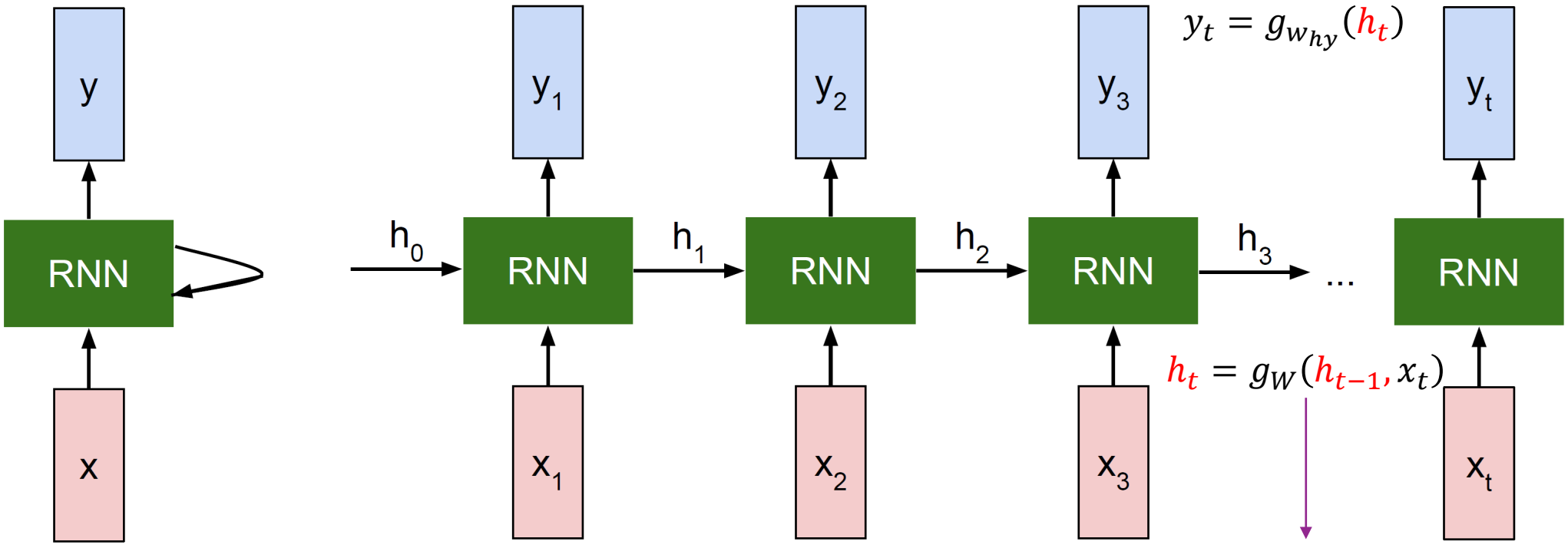


RNN（Recurrent Neural Network）

- RNN：使用隐藏状态h_t“整合”历史信息，隐藏状态随时序更新
  - 在不同时间单位之间状态更新/预测参数W是**共享**的

$$y_t = w_{hy} h_t$$

$$y_t = g_{w_{hy}}(h_t)$$



$$h_t = g_W(h_{t-1}, x_t)$$

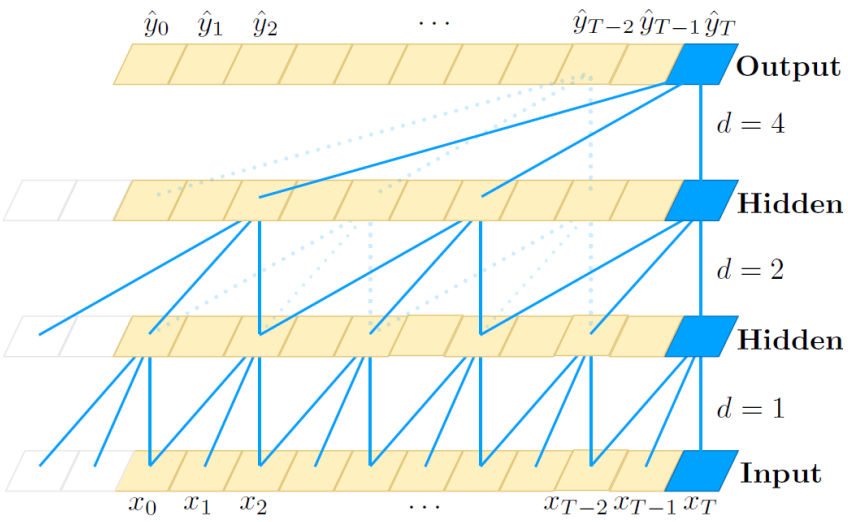
$$h_t = \tanh(W_{hh} h_{t-1} + W_{xh} x_t)$$

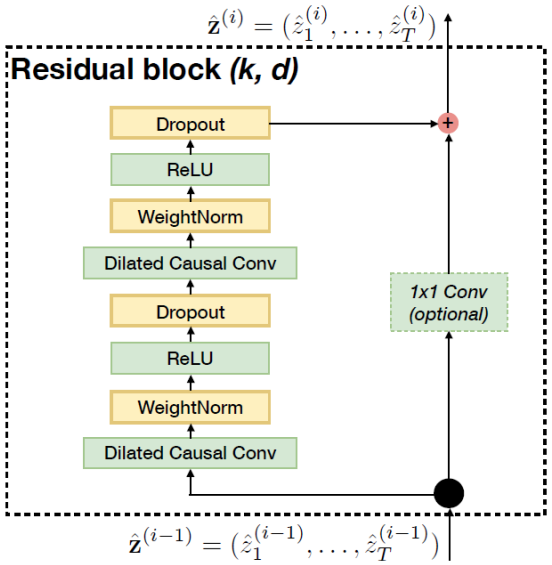
RNN简化表示

RNN展开（unrolled）

Temporal Convolutional Networks (TCN)

- RNN or CNN?
- Temporal Convolutional Networks
  - 因果卷积（Casual Convolution）
  - 变长输入，固定输出：1D卷积
  - 膨胀卷积（Dilated Convolutions）
  - Residual Connections
- 主要优势
  - 并行化（同时处理所有时间节点）
  - 可变receptive field
  - 低内存占用





- 缺点（相比RNN）
  - 测试阶段需要整个历史数据
  - 任务迁移性变弱（不同任务对历史数据需求变化）

Shaojie Bai, J. Zico Kolter, Vladlen Koltun. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. CoRR 2018.

“统计” 和“机器学习”流派

Leo Breiman

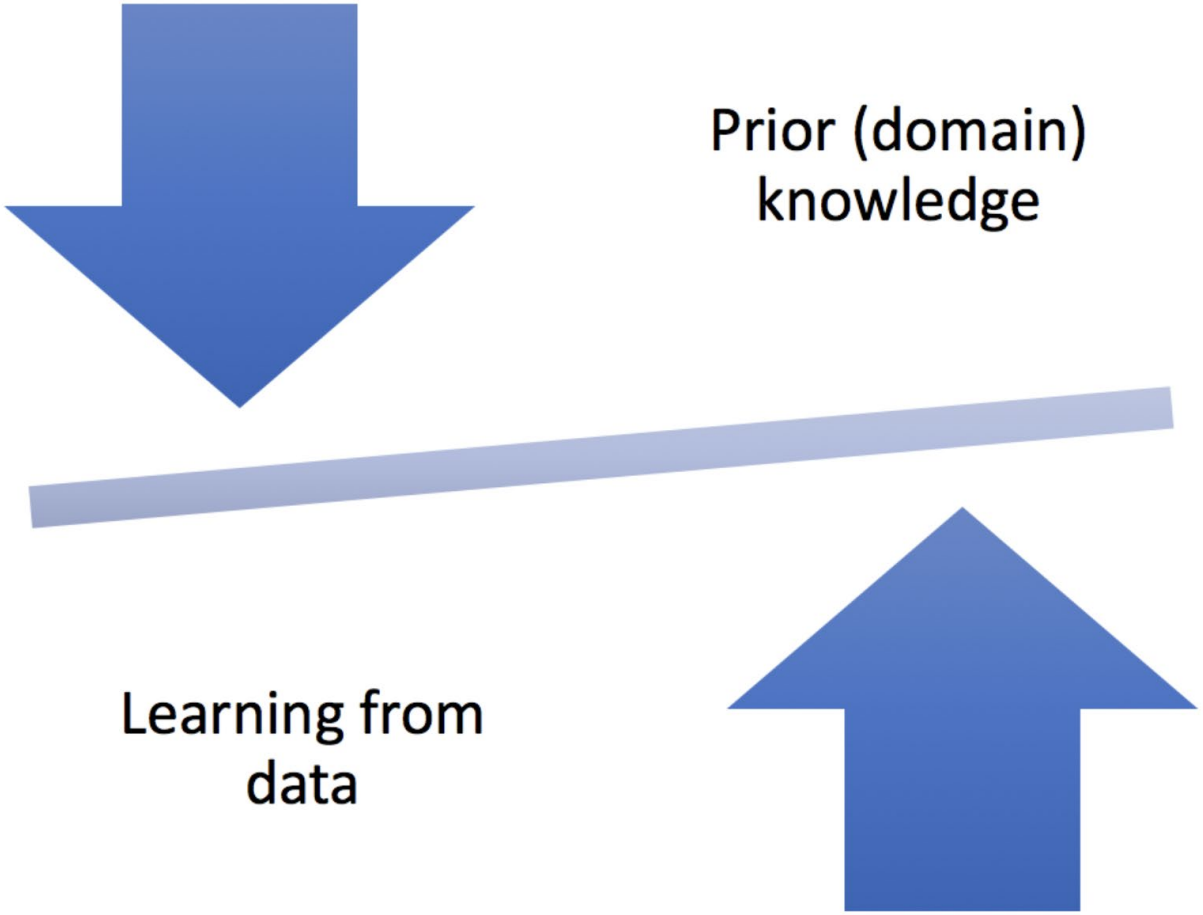
- Statistical Modeling: The Two Cultures
- 统计方法：In-Sample Fit
- 机器学习方法：Out-of-sample Fit

Category	Dimension
Objective	Global vs. Local Methods Probabilistic vs. Point Forecasts Computational Complexity Linearity & Convexity
Subjective	Data-driven vs. Model-driven Ensemble vs. Single Models Discriminative vs. Generative Statistical Guarantees Explanatory/Interpretable vs. Predictive

Tim Januschowski et al. Criteria for classifying forecasting methods. **International Journal of Forecasting** 36 (1), Pages 167-177, 2020.

“统计” 和“机器学习”流派

- 数据或模型驱动，需要一种平衡
- 目标
  - 增加数据利用效率
  - 提高可解释性
  - 更快推断预测



本章概要

1. 时间序列研究的主要思路

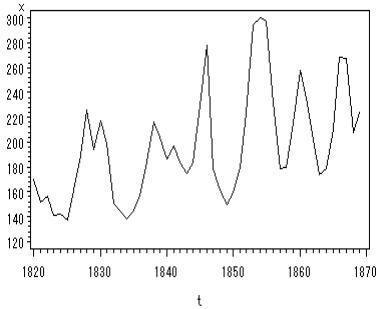
2. 统计方法和机器学习方法

3. 时间序列的发展历史和流派

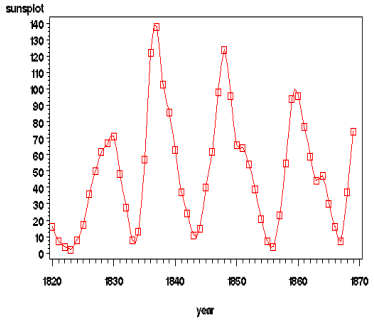
早期时间序列分析

《史记-货殖列传》中记载，公元前500年左右，范蠡认为我国农作物生成存在“六岁穰，六岁旱，十二岁一大饥”的自然规律，提出了我国最早**稳定粮价**的方法：“平粜法”。

J. Herschel，英国天文学家，发现太阳黑子变少时，地球降雨量也会变少，农业生产会和太阳黑子有近似的周期



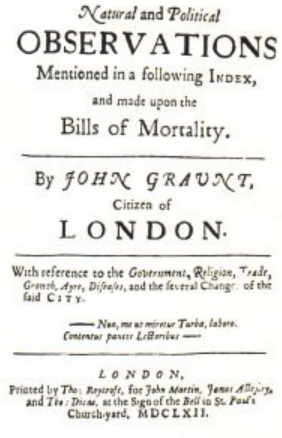
贝弗里奇小麦价格指数序列



太阳黑子序列

John Graunt. 英国经济学家，人口统计学奠基人。1622年出版《关于死亡公报的自然和政治观察》整理了伦敦教会自1604年起每周一次的死亡公报的庞大数据，并对当时有关人口和死亡的原因提出论断。

- 说明统计比率相对于时间和空间的稳定性（某个时间或特性出现的频率随着观察次数的增加趋向于一个固定值）
- 通过序列的一阶差分确定“病态年”
- 通过对原始序列的分析，估计人口占比，对未来情况进行预测和估计



Natural and Political
OBSERVATIONS
Mentioned in a following INDEX,
and made upon the
Bills of Mortality.

By JOHN GRAUNT,
Citizen of
LONDON.

With reference to the Government, Religion, Trade, Growth, Ayre, Diseases, and the several Change of the said CITY.

—— Non, me ut miretur Turba, labore.
Contentus paucis Lectoribus ——

LONDON,
Printed by Tho: Roycroft, for John Martin, James Allestry, and Tho: Dicas, at the Sign of the Bell in St. Paul's Church-yard, MDCLXII.

基于统计学的时间序列分析

英国统计学家Warren M. Persons于1919年在论文《Indices of Business Conditions》中认为所有的序列波动都可以归纳为受到如下四大类因素的综合影响：长期趋势（secular Trends）；循环波动（Cyclical fluctuations）；季节性变化（Seasonal variation）；随机波动 (Irregular variation) 。

George Udny Yule，英国统计学家，1927年对单摆运动和太阳黑子序列的分析中，认为很多时间序列变量不是和时间相关，而是和同一序列中的滞后变量相关，且观察值的差分也和滞后变量的差分相关。提出了 AR(2), AR(4)模型。

Herman Wold，瑞典统计学家，1938年在博士论文《A study in the Analysis of Stationary time Series》给出平稳随机过程的定义，ARMA模型；同时提出了Wold分解定理，证明任何平稳序列都可以分解为确定性序列和随机序列之和。

George Box，英国统计学家。1970年Box和Jenkins出版了《Time Series Analysis Forecasting and Control》一书，提出ARIMA模型，并系统地阐述了ARIMA模型的识别、估计、检验及预测的原理及方法。

机器学习/神经网络的预测方法

ELSEVIER

International Journal of Forecasting

Volume 14, Issue 1, 1 March 1998, Pages 35-62

Forecasting with artificial neural networks:: The state of the art

Guoqiang Zhang, B. Eddy Patuwo, Michael Y. Hu

Journal of FORECASTING

Research article | Full Access

How effective are neural networks at forecasting and prediction? A review and evaluation

Monica Adya, Fred Collopy

First published: 04 December 1998 |
https://doi.org/10.1002/(SICI)1099-131X(1998090)17:5/6<481::AID-FOR709>3.0.CO;2-Q | Citations: 212

- 1969 Weather forecasting with adaptive linear neurons (Hu)
- 1986 Backpropagation (Rumelhart et al.)
- 1988 NNs using backpropagation applied to forecasting; positive results (Werbos)
- 199x Many authors applying mostly feed-forward models to various forecasting problem (single time series)
- 1998 Review articles: "The outcome of all of these studies has been somewhat mixed"; "While ANNs provide a great deal of promise, they also embody much uncertainty."
- 2000 M3 competition – simple methods declared the winner
- 200x Less work on NN-based forecasting methods
- 2012 AlexNet wins ImageNet competition – start of the Deep Learning revival (Krizhevsky et al.)
- 2014 Generating Sequences With RNNs (Graves); seq2seq architecture (Sutskever et al.)
- 2014- Modern deep learning techniques (RNNs, CNNs) get applied to forecasting (across time series)
- 2018 M4 competition: combination of NNs and classical techniques wins

[Jan et al. 2021]