时间序列简介

整合了来自https://zhuanlan.zhihu.com/p/32584136介绍的时间序列简介PPT的要点，纯作个人笔记使用。

1.时间序列算法

2.控制图理论

3.Opprentice系统

4.tsfresh python package

1.时间序列算法

时间序列分析的方法分为：

时域：自相关系数和互相关系数

频域：谱分析和波分析

或：

参数方法

非参数方法

滑动平均 Moving Average

每一个点都是之前w个点的平均数（包括该点）

累计滑动平均 Cumulative Moving Average

每一个点是之前所有点的平均数

加权滑动平均 Weighted Moving Average

每一个点是之前w个点的加权和（包括该点）

指数加权滑动平均 Exponential Weighted Moving Average

Exponential Weighted Moving Average

Suppose $\{Y_t : t \geq 1\}$ is an observed data sequence, the exponential weighted moving average series $\{S_t : t \geq 1\}$ is defined as

$$S_t = \begin{cases} Y_1, & t = 1 \\ \alpha \cdot Y_{t-1} + (1-\alpha) \cdot S_{t-1}, & t \geq 2 \end{cases}$$

- $\alpha \in [0, 1]$ is a constant smoothing factor.
- Y_t is the observed value at a time period t.
- S_t is the value of the EMWA at any time period t.

Moreover, from above definition,

$$\begin{aligned} S_t &= \alpha[Y_{t-1} + (1-\alpha)Y_{t-2} + \cdots + (1-\alpha)^k Y_{t-(k+1)}] \\ &\quad + (1-\alpha)^{k+1} S_{t-(k+1)} \end{aligned}$$

for any suitable $k \in \{0, 1, 2, \cdots\}$. The weight of the point Y_{t-i} is $\alpha(1-\alpha)^{i-1}$.

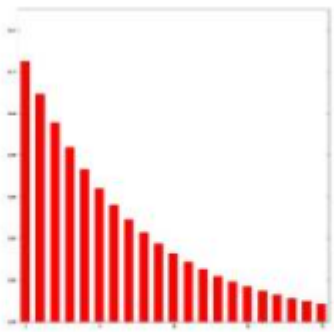

Figure: EMA weights $k = 20$

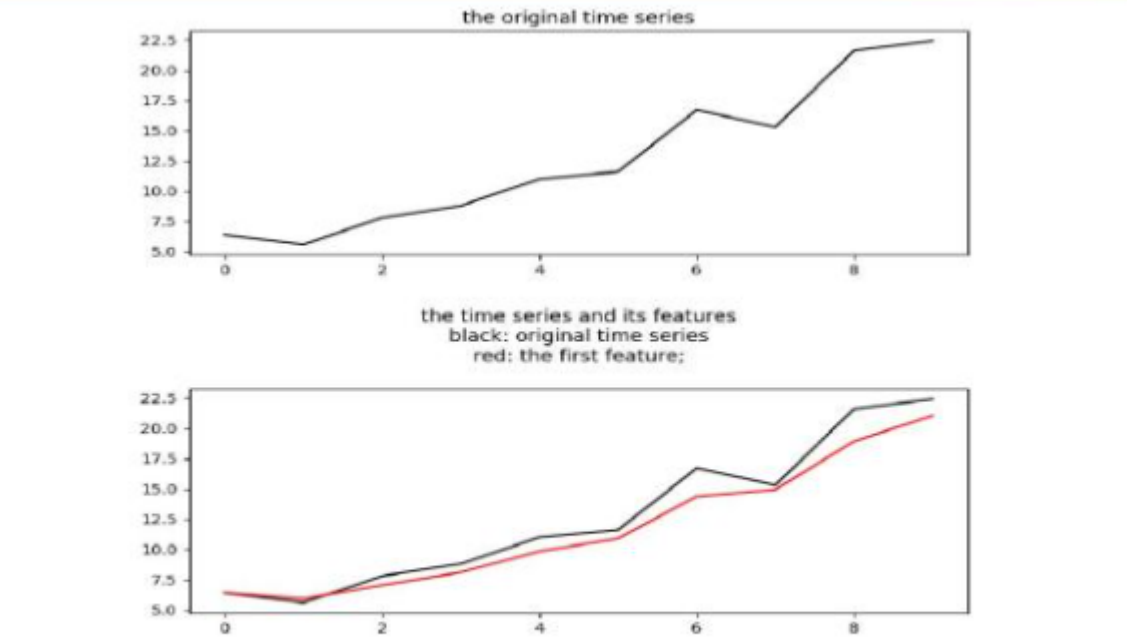

Figure: Exponential Weighted Moving Average Method for $\alpha = 0.6$

双指数平滑 Double Exponential Smoothing

Double Exponential Smoothing

Suppose $\{Y_t : t \geq 1\}$ is an observed data sequence, there are two equations associated with double exponential smoothing:

$$\begin{aligned} S_t &= \alpha Y_t + (1-\alpha)(S_{t-1} + b_{t-1}), \\ b_t &= \beta(S_t - S_{t-1}) + (1-\beta)b_{t-1}, \end{aligned}$$

where $\alpha \in [0,1]$ is the data smoothing factor and $\beta \in [0,1]$ is the trend smoothing factor.

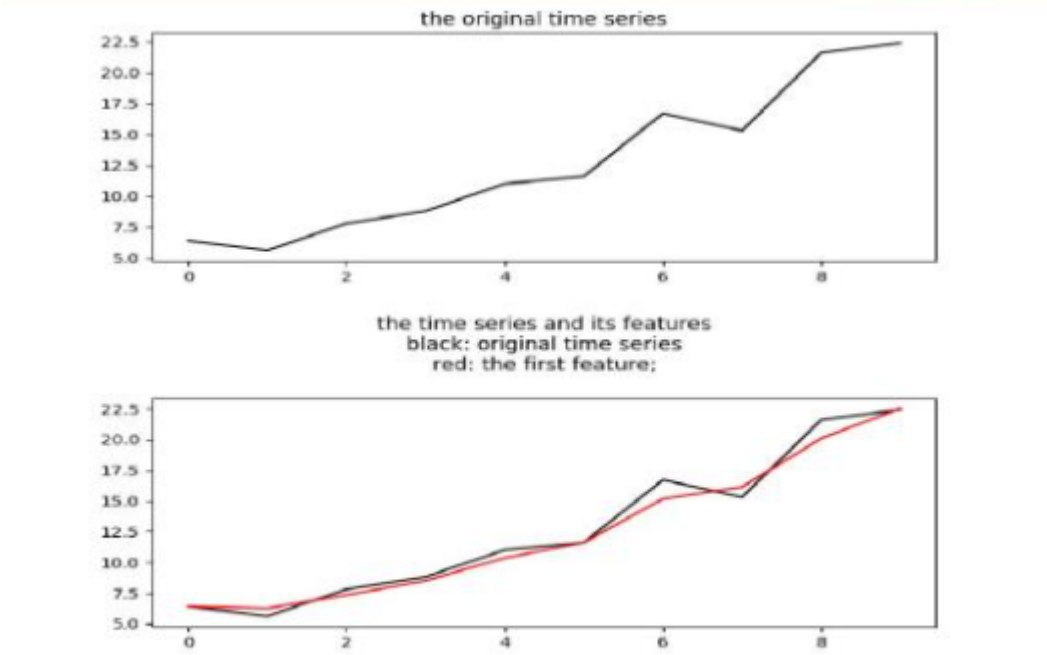

Figure: Double Exponential Smoothing for $\alpha = 0.6$ and $\beta = 0.4$

三指数平滑 Triple Exponential Smoothing

Triple Exponential Smoothing (Multiplicative Seasonality)

Suppose $\{Y_t : t \geq 1\}$ is an observed data sequence, then the triple exponential smoothing is

$$
\begin{aligned}
S_t &= \alpha \frac{Y_t}{c_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1}), \text{ Overall Smoothing} \\
b_t &= \beta(S_t - S_{t-1}) + (1-\beta)b_{t-1}, \text{ Trend Smoothing} \\
c_t &= \gamma \frac{Y_t}{S_t} + (1-\gamma)c_{t-L}, \text{ Seasonal Smoothing}
\end{aligned}
$$

where $\alpha \in [0,1]$ is the data smoothing factor, $\beta \in [0,1]$ is the trend smoothing factor, $\gamma \in [0,1]$ is the seasonal change smoothing factor.

2.控制图理论

控制图用来对图表的一些特征进行控制和分析，主要有三条线：

中心线：均值

上界和下界 Upper/Lower Control Limit

3 σ 控制图

$$UCL = \mu_w + L\sigma_w$$

$$CenterLine = \mu_w$$

$$LCL = \mu_w - L\sigma_w$$

L为控制线到中心线的距离，σ为w的标准差

累计和控制图 The Cumulative Sum Control Chart

CUSUM Control Chart

Let x_i be the i-th observation on the process $\{x_i : 1 \leq i \leq n\}$, $\{x_i : 1 \leq i \leq n\}$ has a normal distribution with mean μ and standard deviation σ. The cumulative sum control chart is calculated by, for all $1 \leq i \leq n$,

$$C_i = \sum_{j=1}^{i}(x_j - \mu_0) = C_{i-1} + (x_i - \mu_0),$$

where $C_0 = 0$ and μ_0 is the target for the process mean.

- If $|C_i|$ exceed the decision interval H, then the process is considered to be out of control.
- The decision interval H is 3σ or 5σ.

允许某些数据点有一定的偏移

表格和算法控制图要点：

Let x_i be the i-th observation on the process $\{x_i : 1 \leq i \leq n\}$, it has mean μ_0 and standard deviation σ. The statistics C^+ and C^- are computed as follows:

$$\begin{aligned} C_i^+ &= \max\left[0, x_i - (\mu_0 + K) + C_{i-1}^+\right] \\ C_i^- &= \max\left[0, (\mu_0 - K) - x_i + C_{i-1}^-\right] \end{aligned}$$

where $C_0^+ = C_0^- = 0$. K is the reference value, is calculated as

$$K = \frac{|\mu_1 - \mu_0|}{2}, \text{ where } \mu_1 = \mu_0 + \delta\sigma \text{ and } \delta = 1.$$

If either C_i^+ or C_i^- exceed the decision interval $H = 5\sigma$, the process is considered to be out of control. Here δ and H are parameters.

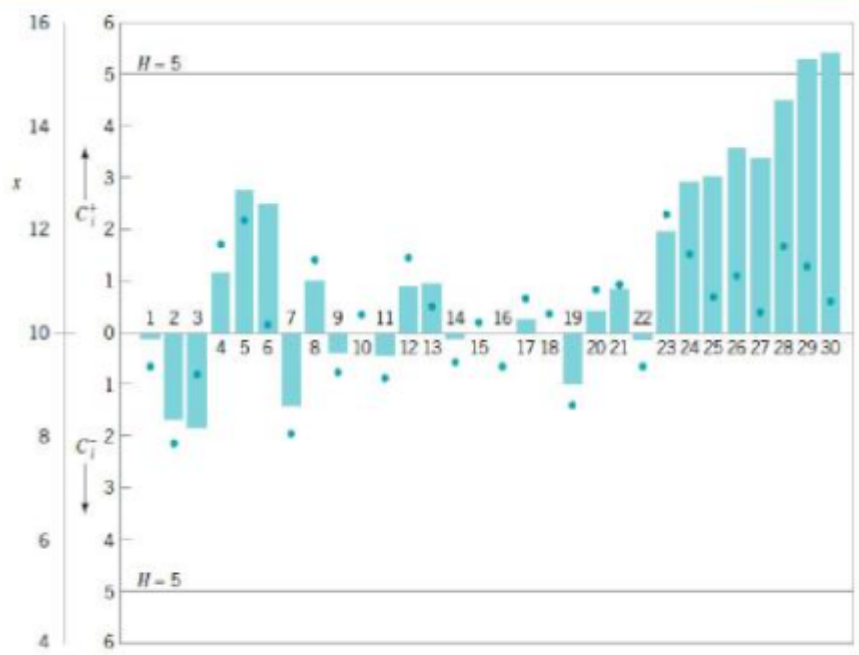

Figure: CUSUM status charts for the above example

指数滑动平均控制图 EWMA Control Chart

$$UCL = \mu_0 + L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}}$$

$$CenterLine = \mu_0$$

$$LCL = \mu_0 - L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}}$$

其中$0 \leq \lambda \leq 1$（省去推导过程）。

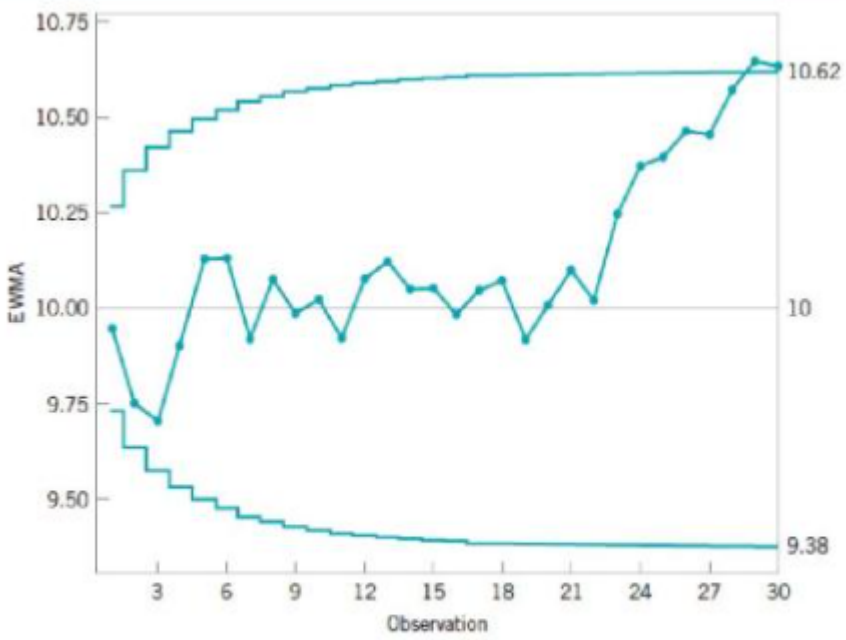

Figure: The EWMA control chart for the above example

滑动平均控制图 MA Control Chart

$$UCL = \mu_0 + \frac{3\sigma}{\sqrt{w}}$$

$$CenterLine = \mu_0$$

$$LCL = \mu_0 - \frac{3\sigma}{\sqrt{w}}$$

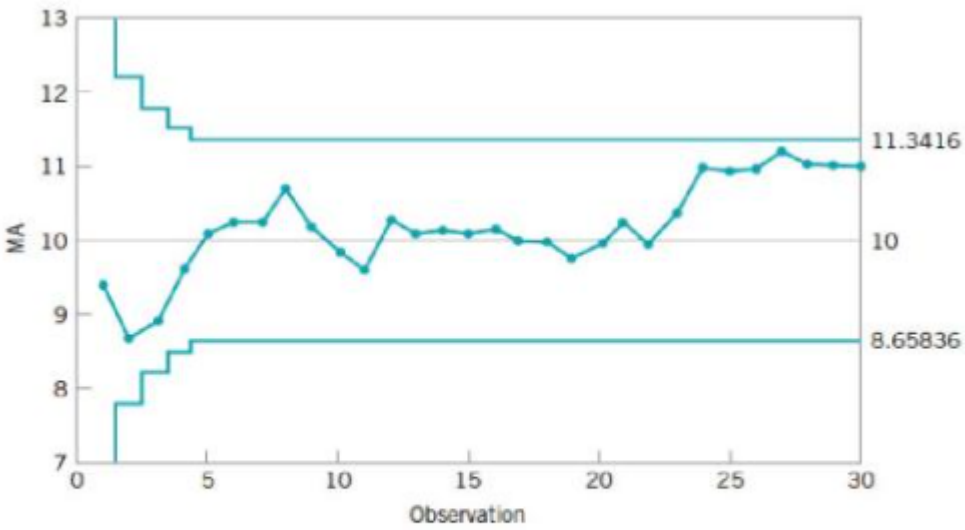

Figure: Moving average control chart with $w = 5$

多元数据控制图，包括：

Hotelling T^2 Control Chart

The Multivariate EWMA Control Chart

Regression Adjustment

Principal Components Method

Partial Least Squares

3.Opprentice 系统

通过用户标注异常数据来进行有监督的机器学习的异常检测系统

4.tsfresh python package

用来一次性提取时间序列中的千余种特征，并包括自动选择有效特征的功能，从而免去了为机器学习服务的特征工程。

相关使用样例在目录tsfresh_examples中。