

R 語言與機器學習 (七)

丘祐瑋
David Chiu

時間序列分析

埃及人的時間序列分析

- 在古埃及，每當天狼星在黎明時從東方地平線升起時，正是一年一度尼羅河水泛濫的時候，尼羅河水的泛濫，灌溉了兩岸大片良田，於是埃及人又開始了他們的耕種
- 古代埃及人發現，天狼星兩次偕日升起的时间间隔不是埃及曆年的365天而是365.25天。古埃及把黎明前天狼星自東方升起的那一天確定為歲首。



感冒的搜尋趨勢

- 從感冒的搜尋趨勢 (Google Trends) 發現民眾每年搜尋感冒的時間點是有規律的



時間序列與迴歸分析

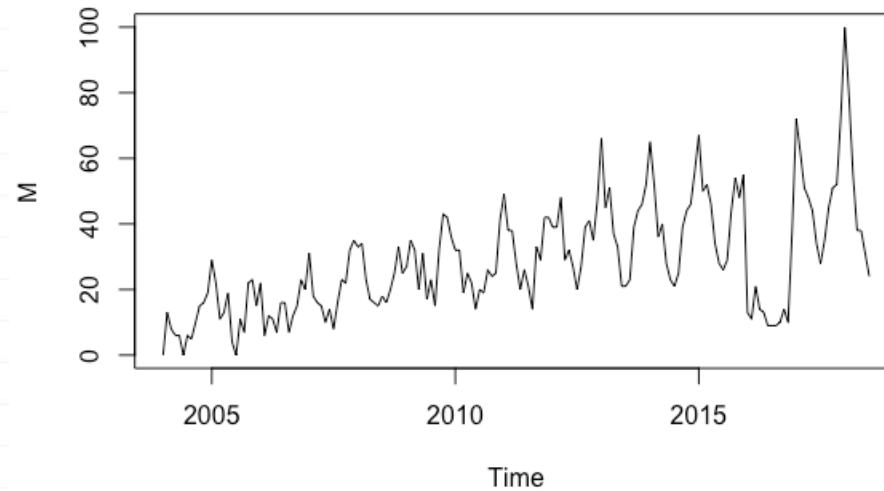
- 時間序列**與時間有關**。線性模型假設觀察值是獨立的情形不適用於時間序列
- 大多數時間序列具有**季節性趨勢**
 - 例如，冬天在Google 的感冒搜尋量會比較高

時間序列的平穩性

- 大部分時間序列模型是在假設它是穩定的前提下建立的
 - 如果一個時間序列隨著時間產生特定的行為，就有很高的可能性認為它在未來的行為是一樣的
- 如果一個時間序列的統計特徵如平均數(Mean)，變異數(Variance)隨著時間保持不變，我們就可以認為它是穩定的
- 平穩的定義
 - 恆定的平均數
 - 恆定的變異數
 - 共變異數Covariance不隨時間變化

讀取時間序列資料

```
cold <- read.csv('https://raw.githubusercontent.com/ywchiu/cdc_course/master/data/cold.csv')  
head(cold)  
M <- ts(cold$cold, frequency = 12, start = c(2004,1))  
plot.ts(M)
```



如何使一個時間序列趨於平穩

- 雖然許多時間序列模型都採用了平穩性假設，但幾乎沒有一個實際的時間序列是平穩的。實際上，幾乎不可能使一個序列完全平穩，但我們要盡可能地接近它
- 導致時間序列不穩定的兩個主要原因：
 - 趨勢-均值隨著時間變化而變化
 - 季節性-特定時間的變化

消除趨勢的方式

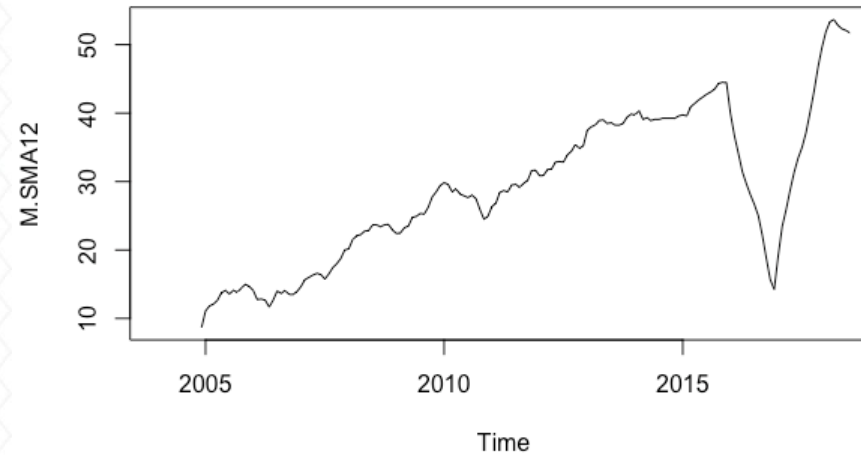
1. 聚合:取一段時間的平均值 (月/周平均值)
2. 平滑:取移動平均數
3. 多項式擬合:擬合一個回歸模型

找出非季節因素的趨勢 (使用均線)

```
library(TTR)
```

```
M.SMA12 = SMA(M, n = 12)
```

```
plot(M.SMA12)
```



指數平滑法

- 計算移動平均時，每出現一新的觀察值，就要從移動平均中減去最遠的觀測值，加入一個最新觀測值以計算移動平均，新的平均移動可做為下一期的預測值。
- 求出指數平滑：
 - 簡單指數平滑法(1 個參數)
 - Holt 雙參數線性指數平滑法(2 個參數)
 - Winters 線性與季節性指數平滑法(3 個參數)

簡單指數平滑法(1 個參數)

- 使用於可用加法模型描述的時間序列，用做恆定水平(數據有平穩性)且沒有季節性的數據做預測。

- 定義公式

- $F_{t+1} = ax_t + (1 - a)F_t$

- a 代表加權，參數介於0~1之間

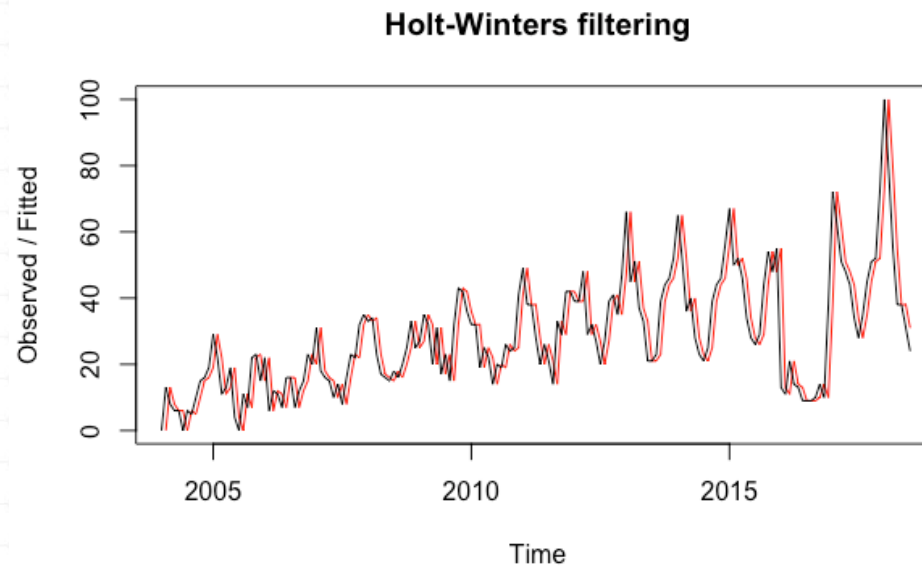
簡單指數平滑法(1 個參數)

```
library(forecast)
```

```
cold.pred<- HoltWinters(M,beta = FALSE, gamma = FALSE)
```

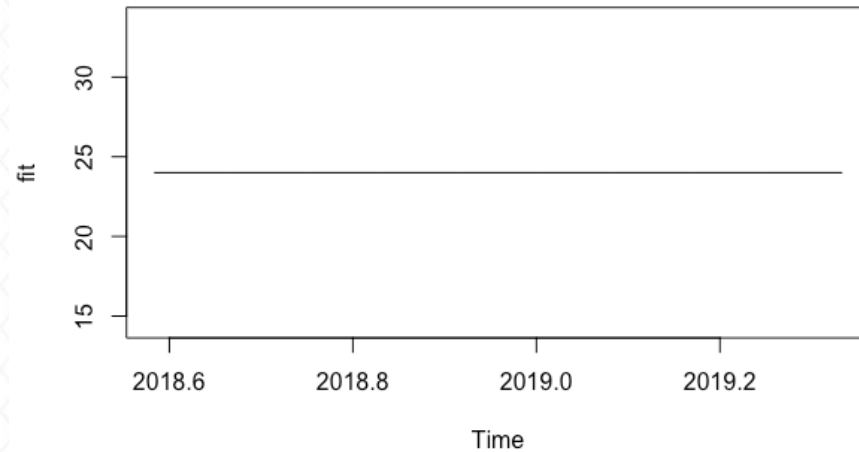
```
cold.pred
```

```
plot(cold.pred)
```



使用簡單指數平滑法產生預測結果

```
library(forecast)  
cold.pred2 <- predict(cold.pred, n.ahead = 10)  
plot(cold.pred2)
```



Holt 雙參數線性指數平滑法(2 個參數)

- 使用於時間序列有增長或降低趨勢，且可用加法模型描述的時間序列，可以估計當前時間點的水平與斜率
- 定義公式
 - $F_{t+m} = S_t + b_t m$
 - $S_t = ax_t + (1 - a)(S_{t-1} + b_{t-1})$
 - $b_t = B(S_t - S_{t-1}) + (1 - B)b_t$
 - b 代表趨勢值，
 - a 評估當前的水平， B 評估當前的斜率

Holt 雙參數線性指數平滑法(2 個參數)

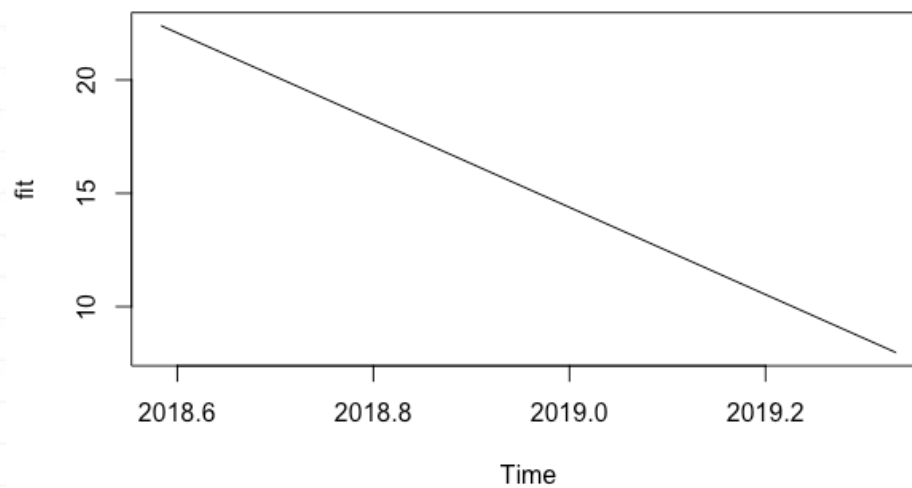
```
library(forecast)
```

```
cold.pred<- HoltWinters(M,gamma = FALSE)
```

```
cold.pred2 <- predict(cold.pred, n.ahead = 10)
```

```
cold.pred2
```

```
plot(cold.pred2)
```



消除趨勢和季節性

- 簡單的趨勢減少技術不能在所有情況下使用，特別是在有季節性的影響因素
- 消除趨勢與季節性的方式
 - 差分(Difference)—取一個特定時間間隔的差值
 - 分解(Decomposition)—建立有關趨勢和季節性的模型然後從模型中刪除它們

時間序列的分解

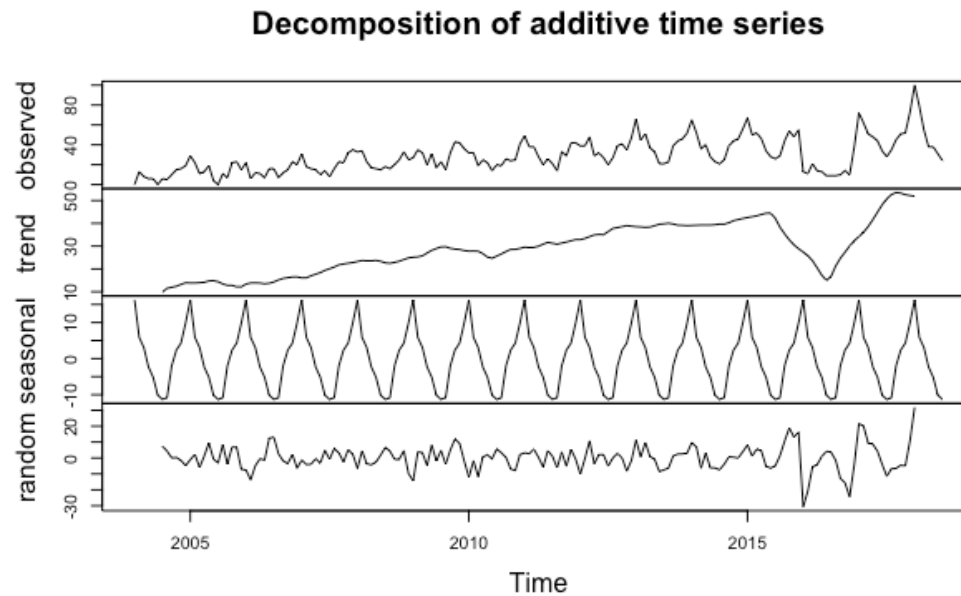
- 長期趨勢 **T**: 時間序列隨時間逐漸增加或減少的變化趨勢
 - 季節變動 **S**: 時間在固定時間內出現固定規則(週期性)的變動
 - 循環變動 **C**: 隨著趨勢性出現鐘擺的變動
 - 不規則變動 **e**: 時間序列受到隨機因素而產生的變動
-
- 加法模型
 - $Y = T + S + C + e$
 - 乘法模型
 - $Y = T * S * C * e$

找出季節因素的趨勢

```
components <- decompose(M)
```

```
components$seasonal
```

```
plot(components)
```



Winters 線性與季節性指數平滑法(3 個參數)

■ 使用於時間序列有增長或降低趨勢，存在季節性，且可用加法模型描述的時間序列

■ 定義公式

$$\square F_{t+m} = (S_t + b_t m) I_{t-L+m}$$

$$\square S_t = a \frac{x_t}{I_{t-L}} + (1 - a) (S_{t-1} + b_{t-1})$$

$$\square b_t = B(S_t - S_{t-1}) + (1 - B)b_{t-1}$$

$$\square I_t = r \frac{x_t}{S_t} + (1 - r)I_{t-L}$$

■ L 代表季節長度，I 代表季節修正係數

Winters 線性與季節性指數平滑法(3 個參數)

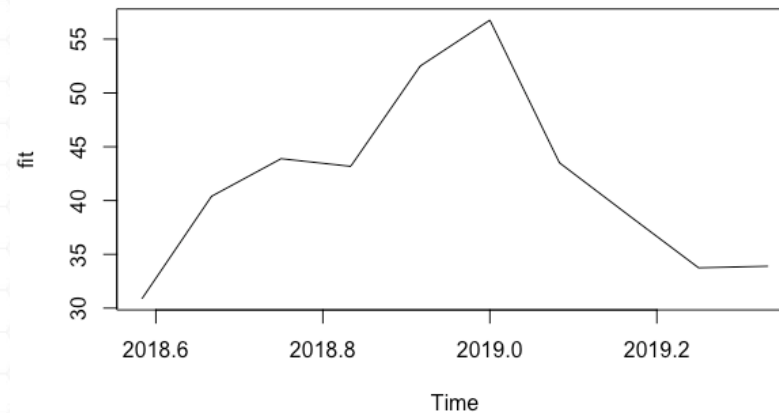
```
library(forecast)
```

```
cold.pred<- HoltWinters(M)
```

```
cold.pred2 <- predict(cold.pred, n.ahead = 10)
```

```
cold.pred2
```

```
plot(cold.pred2)
```



ARIMA

- ARIMA代表差分自回歸移動平均。平穩時間序列的ARIMA預測的只是個線性方程(如線性回歸)。ARIMA使用三個模型參數(p, d, q)產生預測：
 - 自回歸項的數目(p)：AR項僅僅是因變數的時間區間。例如，如果 p 等於5，那麼預測 $x(t)$ 將是 $x(t-1) \dots x(t-5)$
 - 移動平均的數目(q)：MA項是預測誤差的時間區間。例如，如果 q 等於5，那麼預測 $x(t)$ 將是 $e(t-1) \dots e(t-5)$ ， $e(i)$ 是第 i 個時刻的移動平均和真實值的差值。
 - 差分次數(d)：這裡指的是非季節性的差分次數，即這種情況下我們採用一階差分。因為無論我們傳遞差分後的變數且 $d=0$ ，還是傳遞原始變數且 $d=1$ ，得到的結果是一樣的

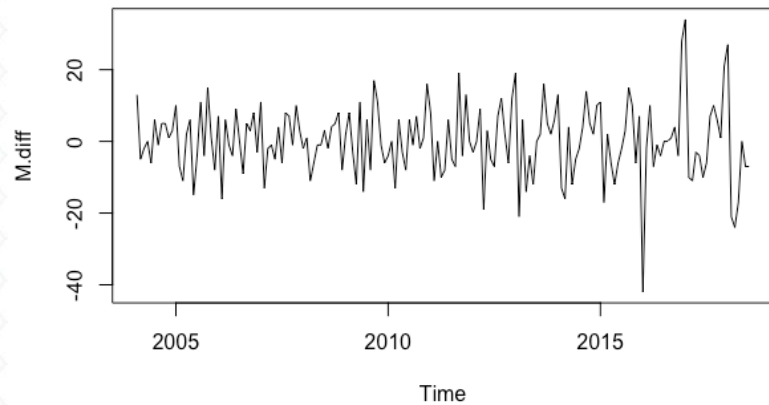
ARIMA 建模過程

- 對時間序列做平穩化處理
- 根據ACF 與 PACF 選擇參數
- 進行參數估計
- 利用模型做預測分析

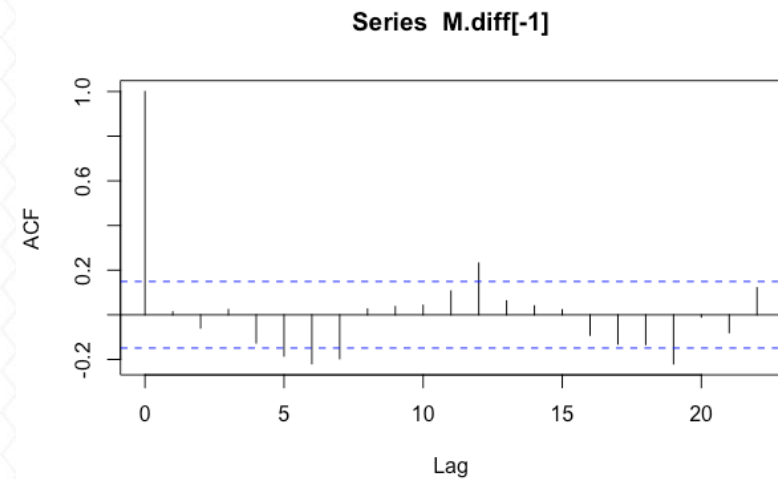
檢視 ACF 圖

```
M.diff <- diff(M, lag = 1, differences = 1)
```

```
plot(M.diff)
```

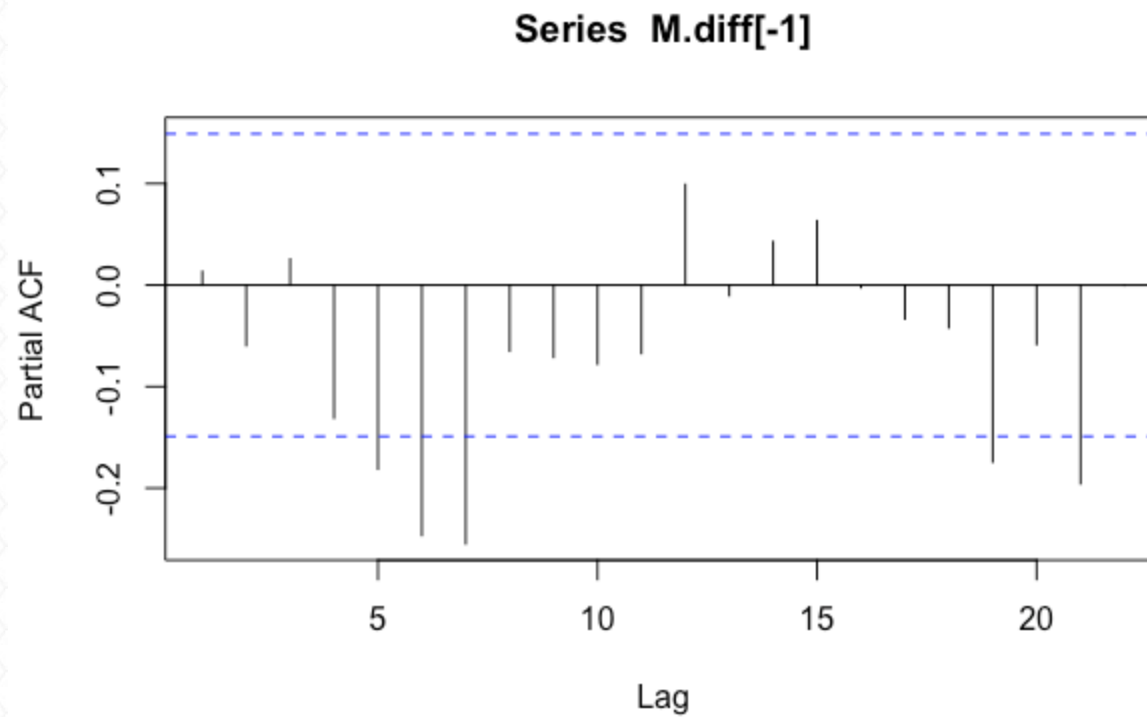


```
acf(M.diff[-1])
```



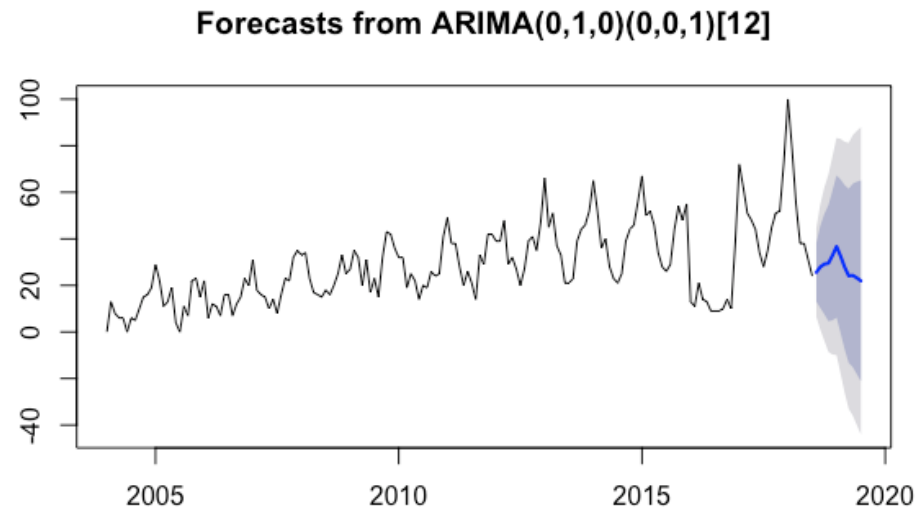
檢視 PACF 圖

`pacf(M.diff[-1])`



使用auto.arima 產生預測

```
M.arima <- auto.arima(M, ic= 'aic')  
M.forecast <- forecast(M.arima,h = 12)  
plot(M.forecast)
```



The background features a light gray hexagonal grid pattern. Overlaid on this are several concentric, semi-transparent circles in shades of light blue and white. A solid dark blue horizontal line runs across the top of the image, and a dark teal horizontal band is at the bottom. The text "THANK YOU" is centered in a bold, dark blue font.

THANK YOU