

# HW1 Regression Report

學號:108062373 姓名:蘇勇誠

## Basic

1. Model (Linear Regression) :

$$y(i) = w_0 + w_1 * x(i)$$

$y(i)$ 為第 $i$ 天TSMC之股價,  $x(i)$ 為第 $i$ 天mtk之股價)

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \times \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$$

使用Ordinary Least Square, 在training data中找出第 $i$ 天MTK的股價與第 $i$ 天TSMC股價的線性關係。並利用training data找出的線性關係預測TSMC的股價。

2. Model選擇之想法: 我嘗試了三種model, 分別為 $y(i) = w_0 + w_1 * x(i)$ 和 $y(i) = w_0 + w_1 * x(i) + w_2 * x(i)^2$ ,  $y(i) = w_0 + w_1 * x(i) + w_2 * x(i)^2 + w_3 * x(i)^3$ 。我分別train 3個model, 分別嘗試了很多種training data set跟validation data set。我發現後面兩個model在validation data中的MAPE都較 $y(i) = w_0 + w_1 * x(i)$ 高, 於是我就選擇 $y(i) = w_0 + w_1 * x(i)$ 作為basic part的model。
3. Input & Output format: 以附件中input.csv作為input data, 以附件中108062373\_basic\_prediction作為output data

## 2. Dataset

Training Data:取2021/1/4~2021/8/31作為Training Data

Validation Data:取2021/9/1~2021/10/14作為Validation Data

, 可以計算出MAPE=2.5%

Testing Data:利用2021/10/15~2021/11/11MTK的股價預測2021/10/15~2021/11/11之TSMC的股價

## Bonus

1. Model (Auto Regression) :

$$y(t) = w_0 + w_1 * y(t-1) + w_2 * y(t-2) + w_3 * y(t-3) + w_4 * y(t-4) + w_5 * y(t-5)$$

$y(t)$ 為第 $t$ 天TSMC的股價， $y(t-1)$ 為第 $t$ 天之前一天TSMC的股價， $y(t-2)$ 為第 $t$ 天之前兩天TSMC的股價，以此類推。

$$\begin{bmatrix} y_5 \\ y_6 \\ \vdots \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & y_4 & y_3 & y_2 & y_1 & y_0 \\ 1 & y_5 & y_4 & y_3 & y_2 & y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & y_{t-1} & y_{t-2} & y_{t-3} & y_{t-2} & y_{t-5} \end{bmatrix} \times \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \end{bmatrix}$$

使用Ordinary Least Square，以欲預測股價的前一天至前五天的股價作為model的 feature

2. Input & Output format：以附件中input\_bonus.csv作為input data，input\_bonus.csv的第一個row為日期，第二個row為TSMC的股價。以附件中108062373\_bonus\_prediction作為output data，108062373\_bonus\_prediction的第一個row為日期，第二個row為預測的TSMC的股價。
3. Model選擇之想法：由於我只採用TSMC 2019/1/3~2021/10/14的股價預測2021/10/15~2021/11/11的股價，然而，我認為股價的漲跌會呈現某種規律而且會受到前幾天股價的影響，於是我採用Auto Regression作為bonus part的Traing Model

4. Dataset

Training Data:取2019/1/3~2021/9/3作為Training Data

Validation Data:取2021/9/6~2021/10/14作為Validation Data

，可以計算出MAPE=3.26%

Testing Data:預測2021/10/15~2021/11/11之TSMC的股價