Introduction to Machine Learning HW1 report

109062202 陳禹辰

In basic part, I use temperature and number of cases to be my regression equation's input. Also, because it seems that three city's case are independent to each other, I use three different regression equation to each of the cities respectively, which look like below.

City A : $Y = W_0 + W_1 * X^1 + W_2 * X^2 + \Phi_1 * Y_{t-1} + \Phi_2 * Y_{t-2}$

City B : $Y = W_0 + W_1 * X^1 + W_2 * X^2 + W_3 * X^3 + \Phi_1 * Y_{t-1} + \Phi_2 * Y_{t-2}$

City C : $Y = W_0 + W_1 * X^1 + W_2 * X^2 + W_3 * X^3 + \Phi_1 * Y_{t-1} + \Phi_2 * Y_{t-2}$

(X is temperature input and Y is lagging output)

In advanced part, I use the temporal index according to the time by converting epiweek number to increasing order index and the number of lagging case as input. It looks like below.

City A : $Y = W_0 + W_1 * I^1 + W_2 * I^2 + W_3 * I^3 + \Phi_1 * Y_{t-1} + \Phi_2 * Y_{t-2}$

City B : $Y = W_0 + W_1 * I^1 + W_2 * I^2 + \Phi_1 * Y_{t-1} + \Phi_2 * Y_{t-2}$

City C :

$Y = W_0 + W_1 * I^1 + W_2 * I^2 + W_3 * I^3 + W_4 * I^4 + \Phi_1 * Y_{t-1} + \Phi_2 * Y_{t-2}$

(I is increasing order index according time and Y is lagging output)

遇到的困難是最開始不知從何下手，以及做出 model 後，拿現有數據測出來的 MAPE 都非常高，也找不到溫度跟病例數的關係，以及在選好 model 之後如何調整要用的 input。

後來是把數據丟到 EXCEL 內，透過裡面的工具做回歸曲線，以年份週數去當作 input 發現他跟時間變化是較有關係的，所以在 basic part 就採用了 autoregression 的方法下去做，發現確實 MAPE 要變得比較好一點，然後再慢慢調整 input order 要去到多少，至於 advanced part 就是直接拿時間順序的 index 當作 input 了。