系級：電機系 2E
姓名：張峻瑋
學號：110511194

機器學習導論 作業 1：線性回歸

Part I. Linear Regression

1. Please plot the data points (only the Training Set) and the fitting curve for M=1,3,5,10,20 and 30, respectively.

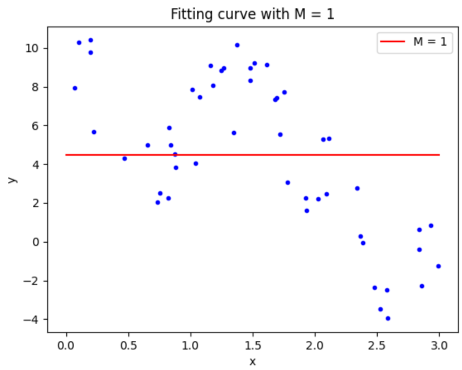

圖 1：M=1 之最適曲線

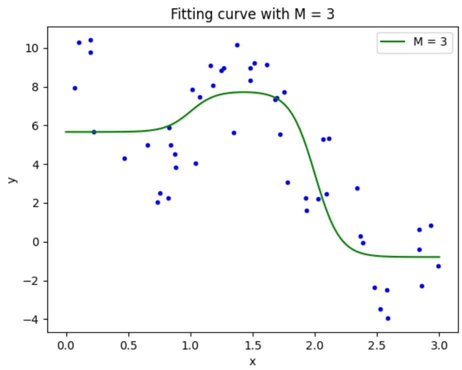

圖 2：M=3 之最適曲線

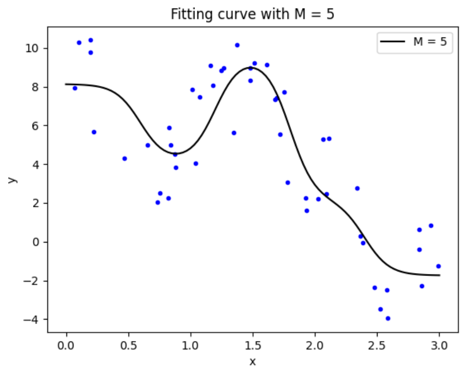

圖 3：M=5 之最適曲線

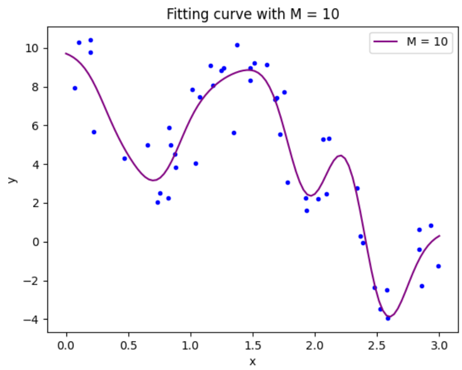

圖 4：M=10 之最適曲線

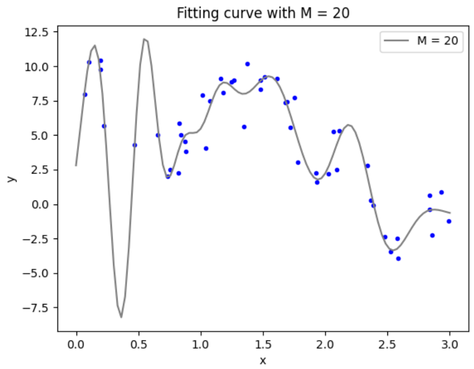

圖 5：M=20 之最適曲線

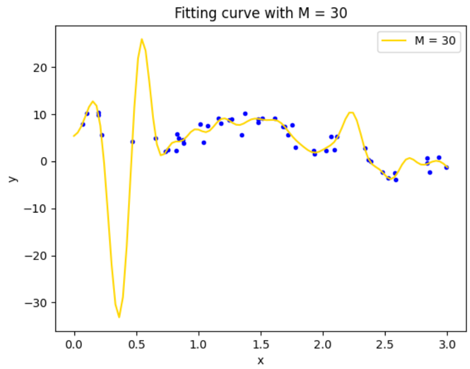

圖 6：M=30 之最適曲線

結果分析：

第一題的做法，即先透過 generate_Phi 函式產生設計矩陣 Φ，再藉由 Φ 產生權重向量 **w**，得出 **w** 向量後再與題目所提供的 ϕ 做內積形成的 y 值畫出最適曲線。其中 M 代表該最適曲線的次方數。從圖 1 到圖 6 我們發現，當 M 愈大，曲線會與資料點愈相符，圖 3 到圖 6 都有這個結果，但圖 5、圖 6 中，為了使曲線更符合資料點而使得曲線 x = 0.5 前後變動巨大，似乎有過適（overfitting）的現象。

2. Please plot the Mean Square Error evaluated on the Training Set and the Testing Set separately for M from 1 to 30.

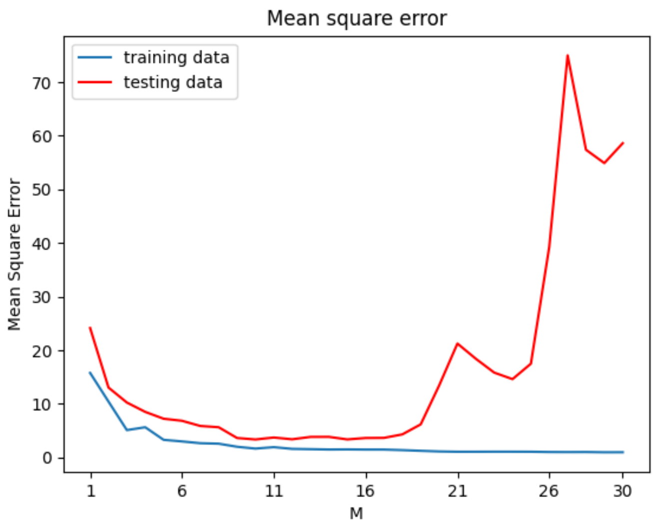

圖 7：在不同的 M 下訓練資料與測試資料的均方誤差

結果分析：

第二題使用題目的均方誤差公式，將訓練資料與測試資料分別以 M = 1 到 M = 30 去跑，發現藍線（訓練資料）隨著 M 愈大，均方誤差愈小；然而紅線

（測試資料）在 18 過後大幅提升，顯見 M>18 存在過適現象。

3. Please apply the 5-fold cross-validation in your training stage to select the best order M and then evaluate the mean square error on the Testing Set. Plot the fitting curve and data points (only the Training Set). You should briefly express how you select the best order M step-by-step.

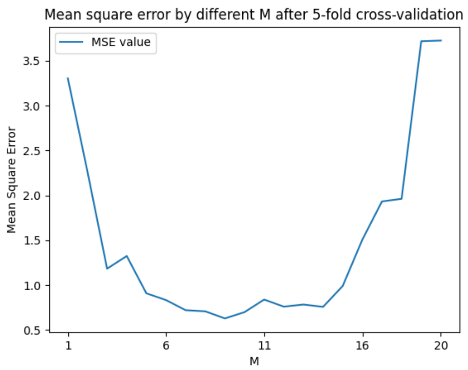

圖 8：各種 M 下 5 折交叉驗證結果
（資料未打散，M > 20 省略）

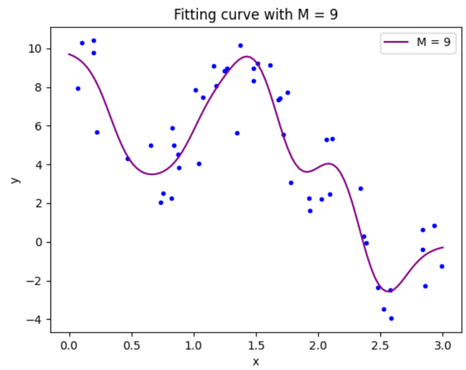

圖 9：M = 9 之最適曲線

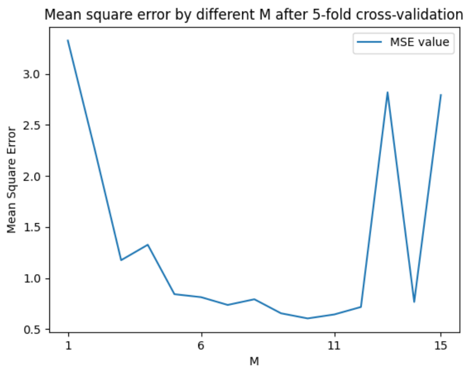

圖 10：各種 M 下 5 折交叉驗證結果之一
（資料隨機打散，M > 15 省略）

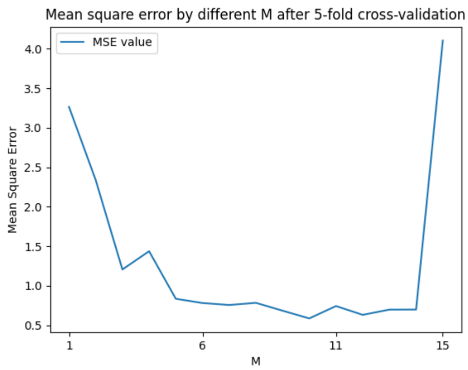

圖 11：各種 M 下 5 折交叉驗證結果之二
（資料隨機打散，M > 15 省略）

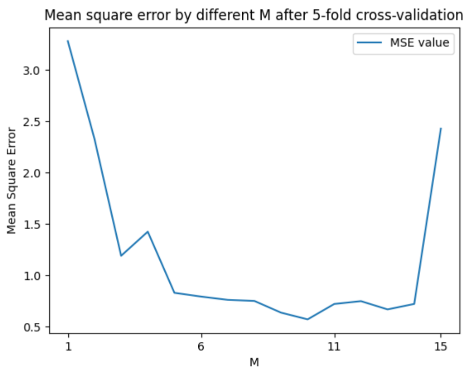

圖 12：各種 M 下 5 折交叉驗證結果之三

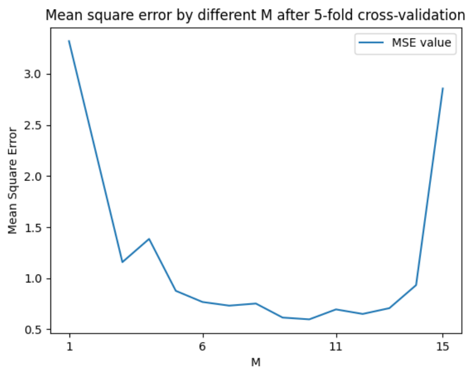

圖 13：各種 M 下 5 折交叉驗證結果之四

（資料隨機打散，M > 15 省略）　　　　（資料隨機打散，M > 15 省略）

結果分析：

利用 5 折交叉驗證，將 50 筆訓練資料依序拆分成 5 個集合，每個集合共 10 筆。將 5 次實驗之方均誤差取平均，即可得圖 8 之結果。從圖 8 看出，M = 9 時方均誤差最小，其曲線如圖 9。

然而將資料拆分成 5 個等分的方式不同，會影響圖 8 之結果。圖 8 之結果只是一種拆分方法下得到之結果。故在程式中又用另一種方式去跑，也就是將資料打散，隨機分成 5 組去算均方誤差，並將此過程做 10 次取平均。圖 10 到圖 13 都是依上述方式跑出來的其中一種結果，可見每次跑出來的結果都不一樣。不見得每次跑出來的結果均方誤差值最低者皆為 M = 10，但剛好這 4 張圖最低點都在 M = 10 處，故取最佳的 M 為 10，其曲線同圖 4。output.xlsx 的最佳 M 也是採用 M = 10。

4. Considering regularization, please use the modified error function

$$\tilde{E}(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{N}\{y(x_i\,,\boldsymbol{w}) - t_i\}^2 + \frac{\lambda}{2}\|\boldsymbol{w}\|^2$$

where $\|\boldsymbol{w}\|^2 \equiv w_1^2 + w_2^2 + \cdots + w_M^2$. Repeat Part I -1. and Part I-2. with $\lambda = \frac{1}{10}$. (You can also try to change the value of λ and discuss what happens under different λ values.)

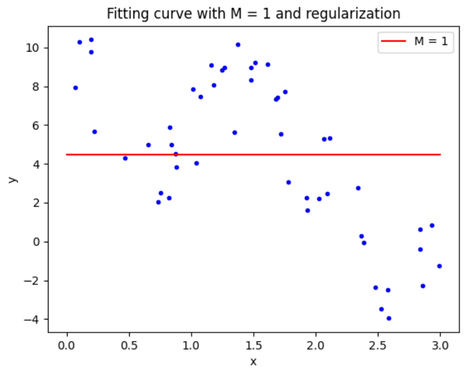

圖 14：正則化後 M=1 之最適曲線

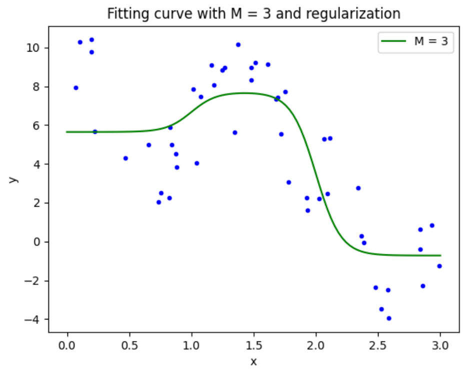

圖 15：正則化後 M=3 之最適曲線

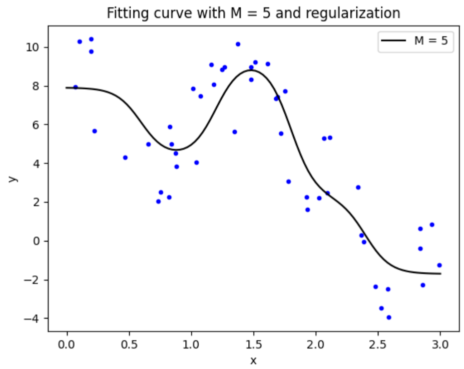

圖 16：正則化後 M=5 之最適曲線

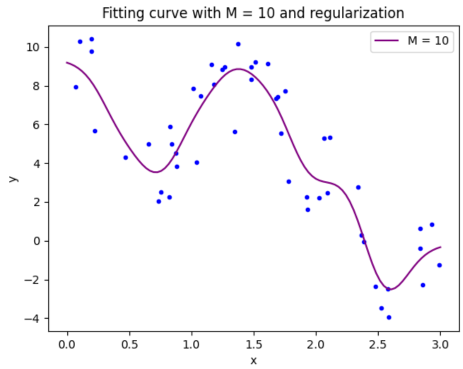

圖 17：正則化後 M=10 之最適曲線

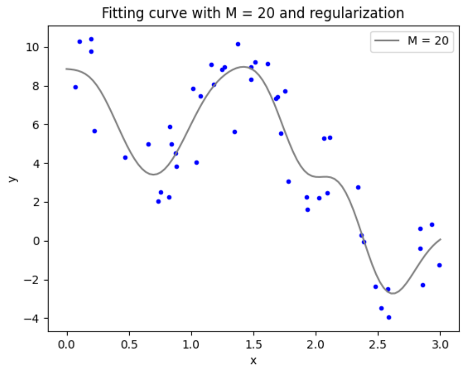

圖 18：正則化後 M=20 之最適曲線

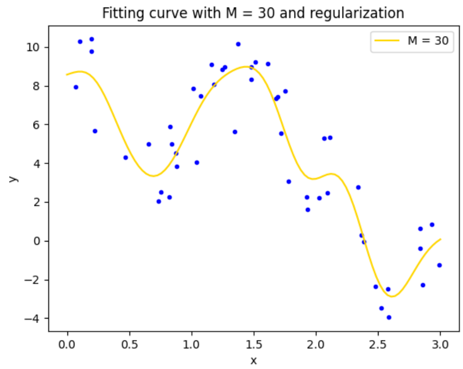

圖 19：正則化後 M=30 之最適曲線

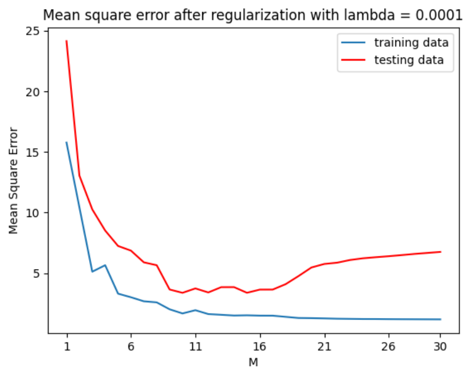

圖 20：λ = 0.0001 時各 M 之均方誤差

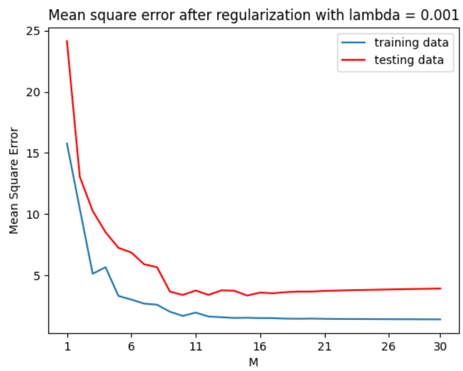

圖 21：λ = 0.001 時各 M 之均方誤差

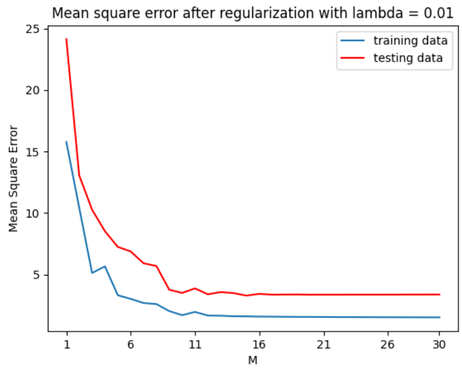

圖 22：λ = 0.01 時各 M 之均方誤差

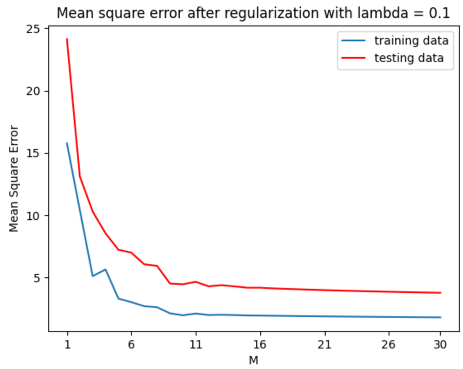

圖 23：λ = 0.1 時各 M 之均方誤差

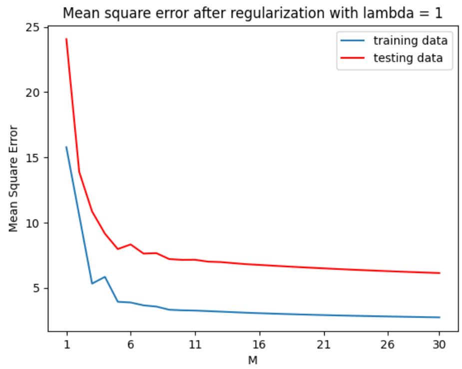

圖 24：λ = 1 時各 M 之均方誤差

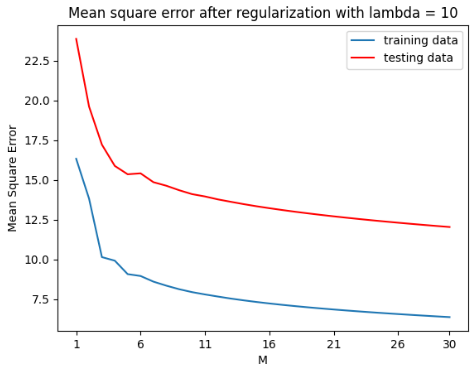

圖 25：λ = 10 時各 M 之均方誤差

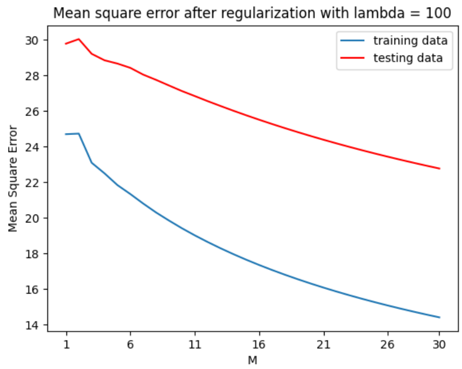

圖 26：λ = 100 時各 M 之均方誤差

結果分析：

利用第四題的式子及帶入 λ = 1/10 後，得到圖 14 到圖 19 之結果。由圖 18、圖 19 可看出，正則化後的結果不會使曲線為了使均方誤差愈小而產生過適現象，可與圖 5、圖 6 做比對。

調整 λ 值亦會改變各種 M 所對應的圖形。從圖 20 到圖 23 的過程可看出，當 λ 愈來愈大，會使得 w 被限制得愈來愈多，使得在 M > 18 處過適現象逐漸簡少。從圖 23 到圖 26 來看，會發現當 λ 愈來愈大，w 被過度壓縮，使其趨近於

0，反而會使均方誤差愈來愈大。

Part II. Bayesian Linear Regression

In this part, use the Training Set in Part I, apply the sigmoidal basis functions in Part I with *M*=10, and implement Bayesian linear regression. In order to discuss how the amount of training data affects the regression process, please implement a sequential estimation:

Please ***sequentially*** compute the mean and the covariance matrix for the posterior distribution $p(\mathbf{w}|\mathbf{t}) = \mathcal{N}(\boldsymbol{w}|\boldsymbol{m_N}, \boldsymbol{S_N})$ with the given prior $p(w) = \mathcal{N}(w|m_0, S_0)$ with $m_0 = 0, S_0^{-1} = 10^{-6}\mathbf{I}$. The predictive distribution $p(t|X, w, \beta)$ is chosen to be $\beta = 1$. Similar to the following figures, please plot the curve of the posterior mean versus x and the region spanning one standard deviation on either side of the mean curve for N = 1, 2, 3, 4, 5, 10, 20, 30, 40, 50.

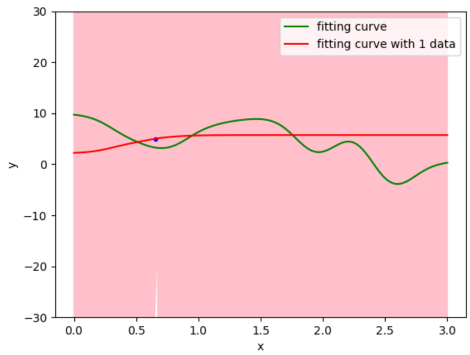

圖 27：1 筆資料下的曲線與標準差

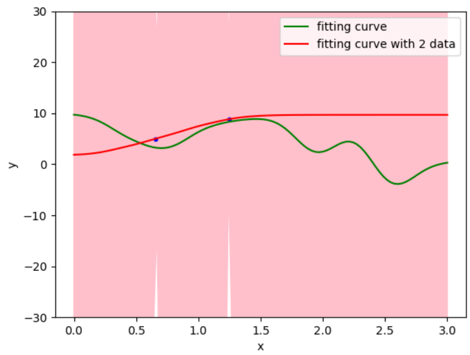

圖 28：2 筆資料下的曲線與標準差

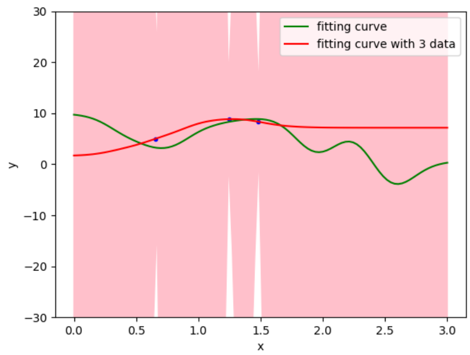

圖 29：3 筆資料下的曲線與標準差

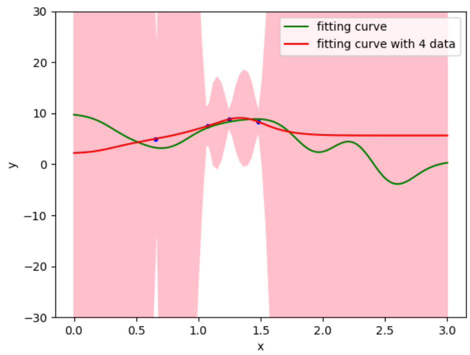

圖 30：4 筆資料下的曲線與標準差

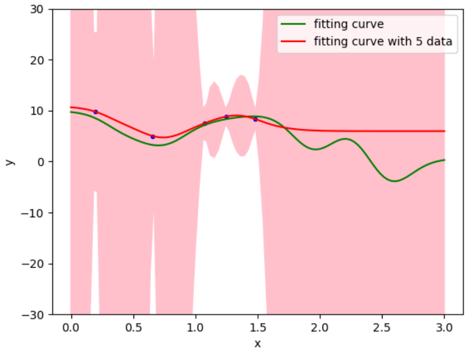

圖 31：5 筆資料下的曲線與標準差

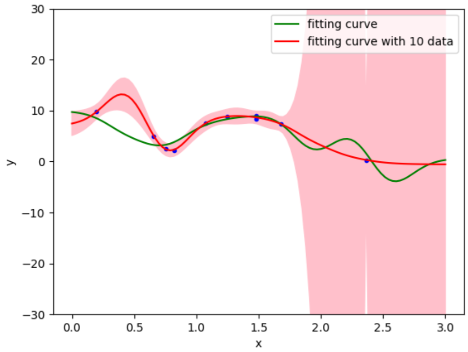

圖 32：10 筆資料下的曲線與標準差

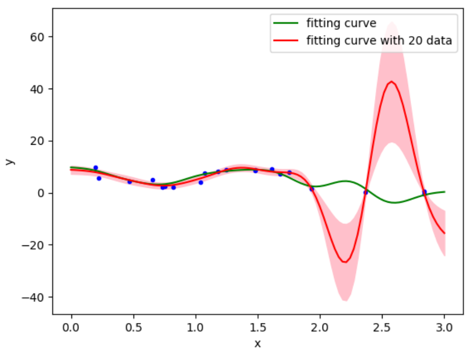

圖 33：20 筆資料下的曲線與標準差

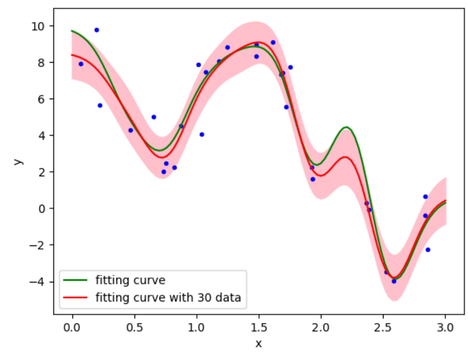

圖 34：30 筆資料下的曲線與標準差

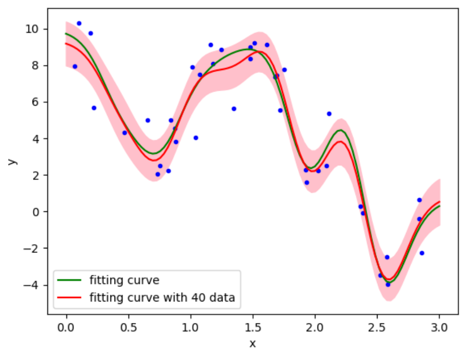

圖 35：40 筆資料下的曲線與標準差

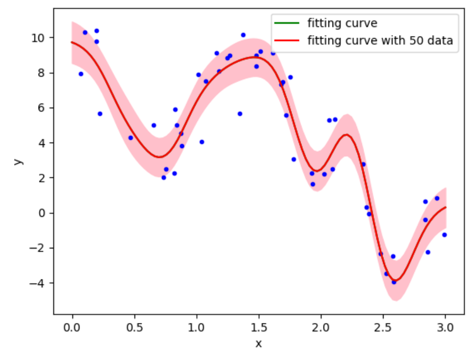

圖 36：50 筆資料下的曲線與標準差

結果分析：

利用 Part I 所建構的模型，設定 M = 10，將資料一筆一筆送進來。從圖 27 到圖 36 可發現，在資料只進來一筆時，除了該資料點所對應之 x 值附近，其他的標準差都非常大，最適曲線也與 50 筆資料的曲線有極大差異。由於先進來的資料點集中在 x < 2 處，到圖 22 已可發現 10 筆資料已讓 x < 2 處的標準差控制在一定範圍內，曲線也與 50 筆資料的曲線逐漸吻合。圖 34 與圖 35，30 筆資料進來後，兩條線的差距已不大，標準差的縮小也有限。至圖 36，綠線與紅線吻

合。