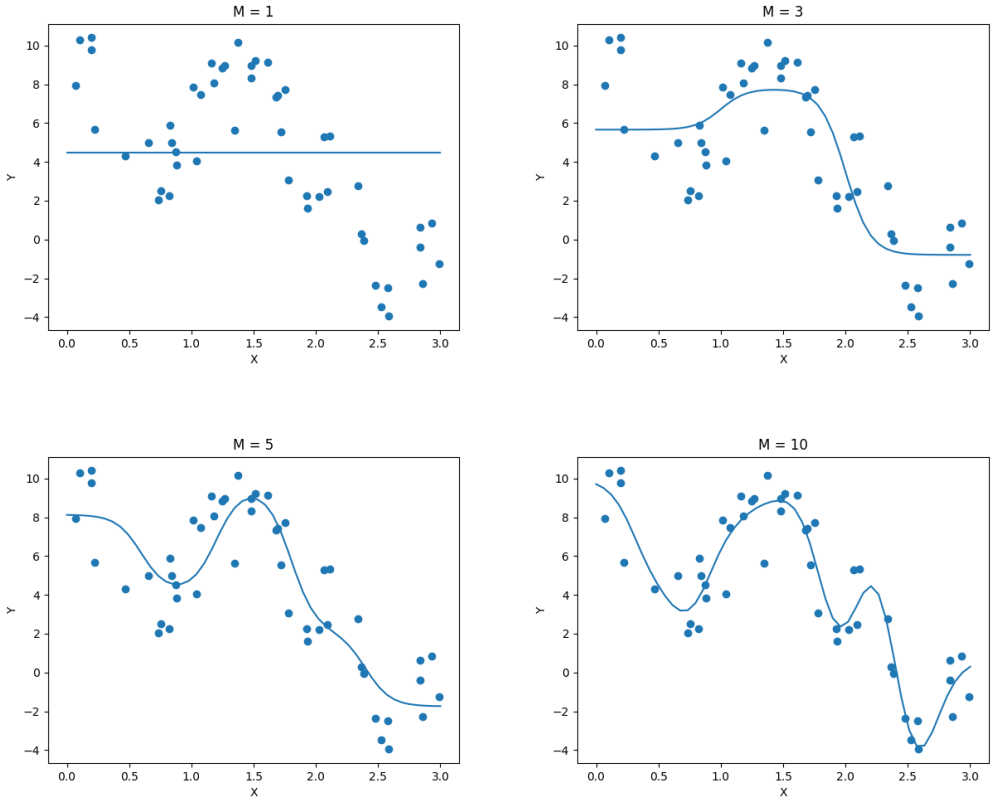
HW1: Linear Regression

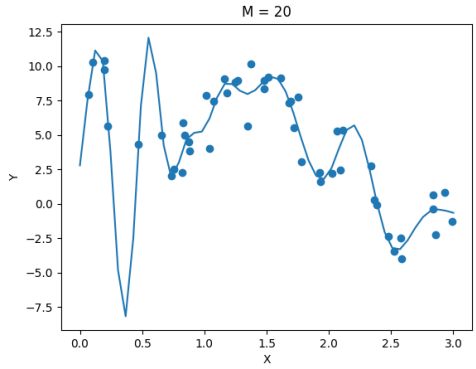
109511219 林錦樑

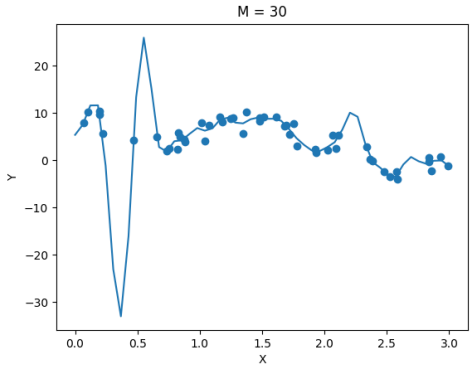
Part I. Linear Regression

1. Fitting curve for M=1,3,5,10,20 and 30

下面六張圖分別為不同 M 時的 fitting curve 與 datapoint。可以看出當 M=1 時，fitting curve 為水平線，Y 與 X 無關。隨著 M 值增大，fitting curve 越來越彎曲，且更加貼合 datapoint。從 M=20 開始時，可以發現 fitting curve 在 datapoint 多的地方貼合得很好，但在沒有 datapoint 的地方時，會出現相當大的起伏。推測是 fitting curve 為了盡可能地貼合 datapoint，使得 W 值變大，出現 overfitting 的現象。

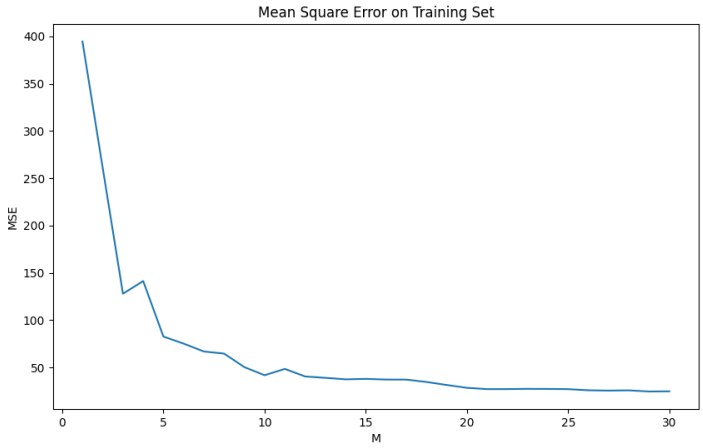


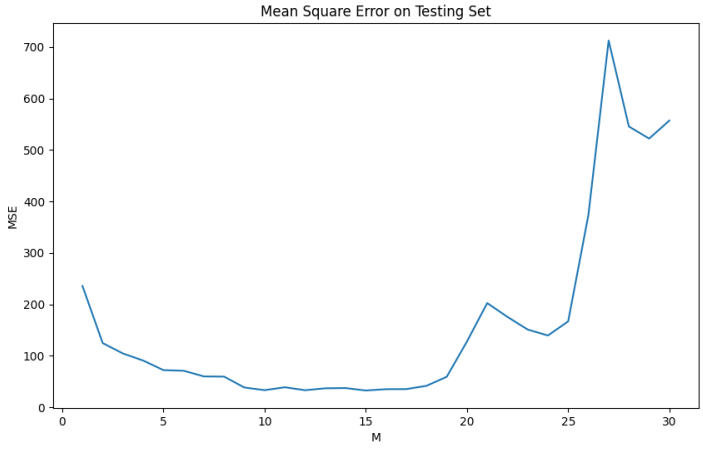




2. Mean Square Error evaluated on the Training/Testing Set

下圖為在訓練集與測試集的 MSE。可以看出在訓練集上，整體趨勢是 MSE 隨著 M 增加而下降，代表 M 越大時，越貼合在 datapoint 上。在測試集上，在 M=15 以前，MSE 是呈現下降趨勢，M=20 以後 MSE 快速上升，甚至超越最初的 M=1 時的 MSE，代表在 M=20 以後，fitting curve 過度依賴在 datapoint 上，出現 overfitting 的現象，與第一點的觀察結論相同。



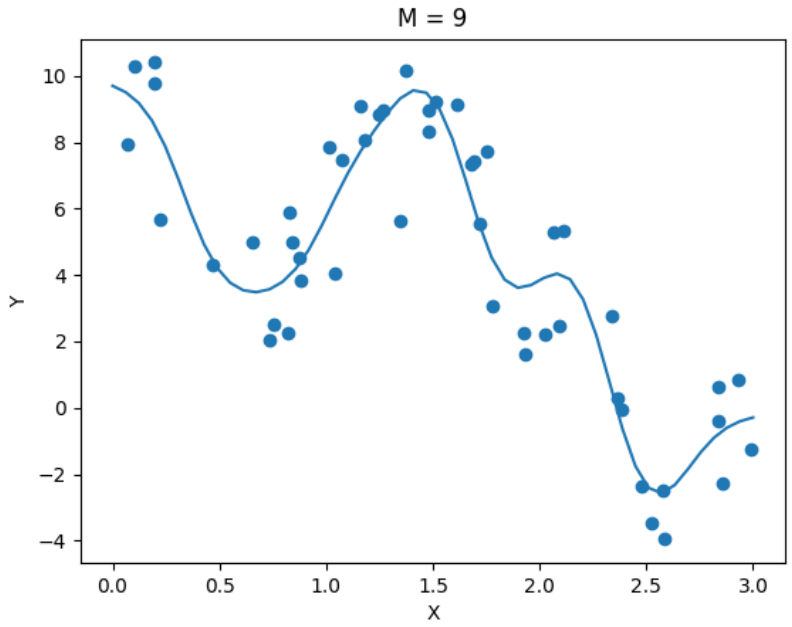


3. 5-fold cross-validation

首先，將 training data 分成五份，每份 10 筆 data，我是按照 data 順序分組的。每次訓練時取其中四份訓練，剩下一份作驗證，每份 data 都會被用來當作驗證集過。

同個 M 值總共會訓練五次，每次訓練完後將 model 拿到驗證集去計算 MSE，最後再將五次的 MSE 取平均，找出平均 MSE 最小值者，即可獲得最佳的 M 值，我透過 5-fold cross-validation 所得出的最佳 M 值為 9。

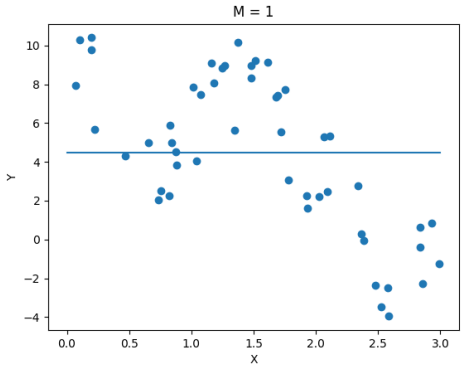
再來將 M=9 的 model 拿到測試集做測試，可以得到 MSE=38.22。我同時也使用了挑選出來的 M 值的正負一做測試，M=8 時，MSE=59.41，M=10 時，MSE=33.08。可以看出用 5-fold cross-validation 挑選出來的 model 在測試時不一定會表現最好，但整體表現已相當不錯，MSE 只稍微高了一點。

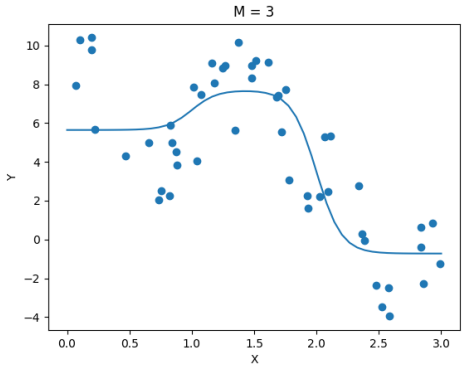


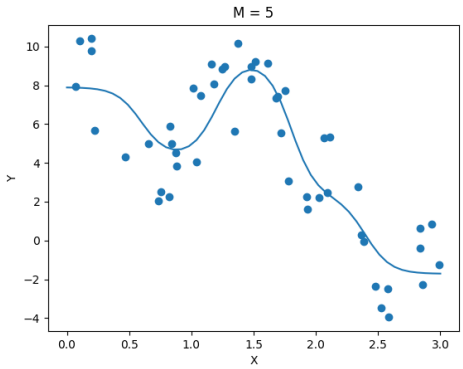
4. Regularization

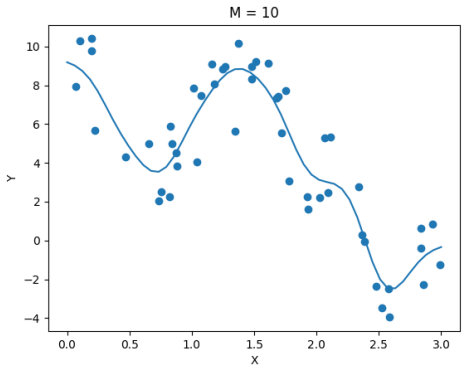
i. Fitting curve for M=1,3,5,10,20 and 30

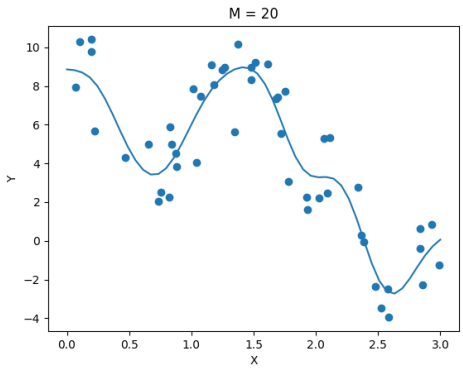
可以觀察出與第一點的趨勢很像，M 增加時，線段更彎曲，更 fitting。比較不同的地方是從 M=20 開始，由於 regularization term 的緣故，線段不會再出現很大的起伏。同時也注意到，從 M=10 之後的 fitting curve 都蠻相似的，沒有太大的變化。

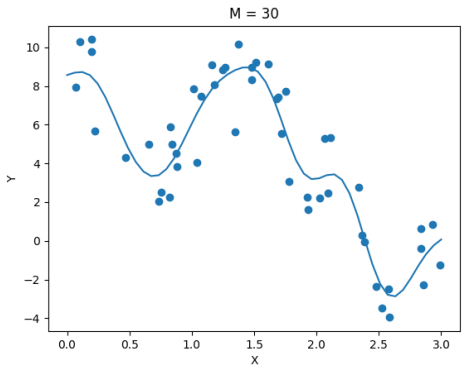




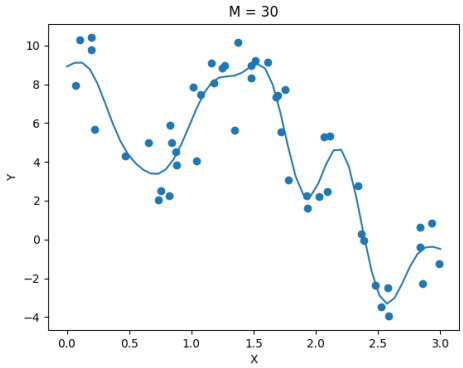


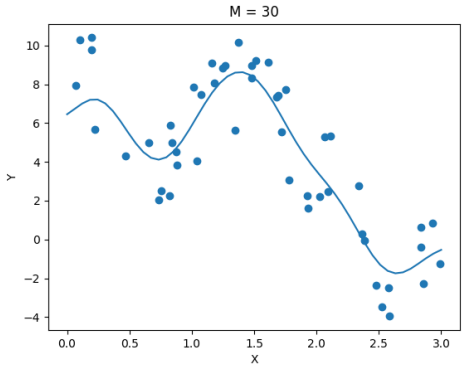






如果將 λ 調小到 0.01，則 fitting curve 更彎曲，如下列左圖。λ 調大到 1，則 fitting curve 更平滑，如下列右圖。

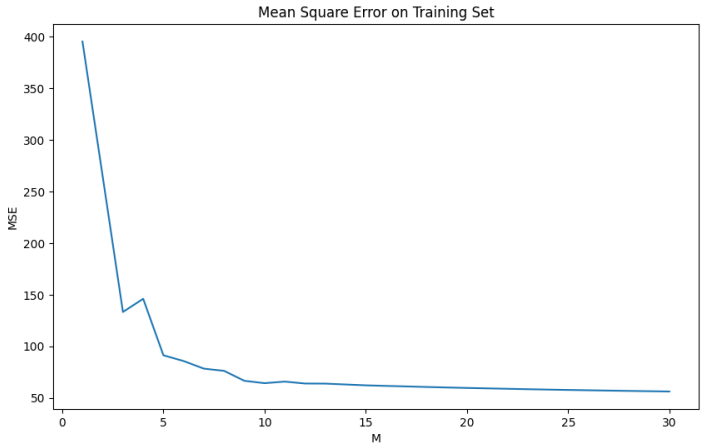


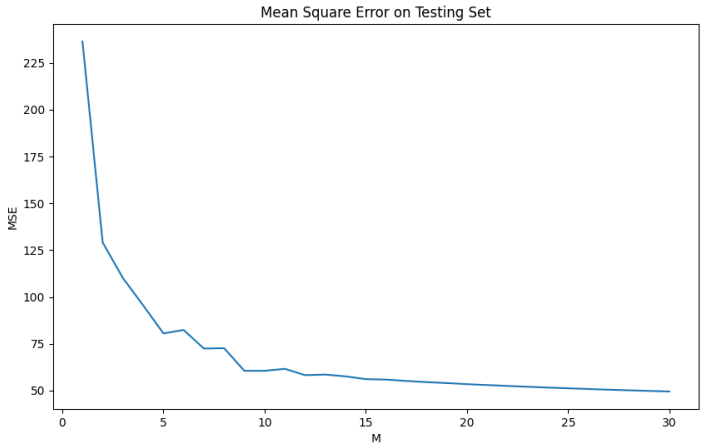


ii. Mean Square Error evaluated on the Training/Testing Set

在訓練集上，整體趨勢與第二點相同，都是往下，但 MSE 的值由於 regularization term 的緣故而增加。在測試集上的趨勢也與訓練集類似，整體向下，不像第二點測試集的 MSE 會隨 M 值增加而變大，代表透過 loss function 增加 regularization term 可以解決 overfitting 的問題。

如果將 λ 調小到 0.01，regularization term 的值會下降，使 MSE 減少。相反地，如果 λ 調大到 1，regularization term 的值會增加，造成 MSE 上升。



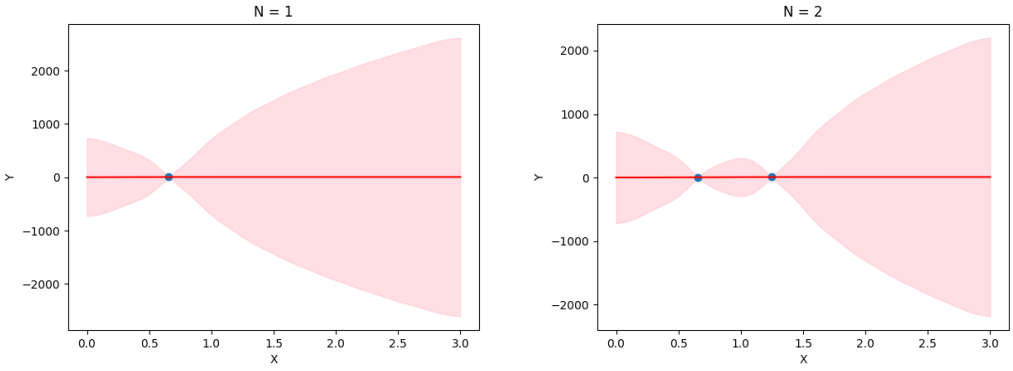


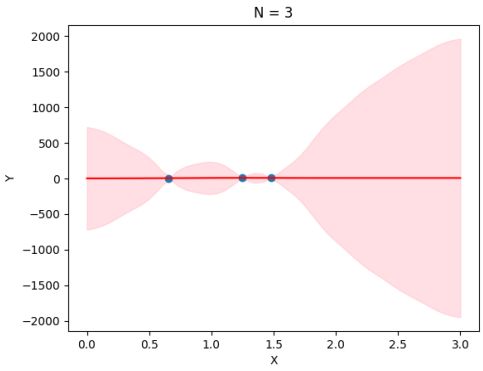
Part II. Bayesian Linear Regression

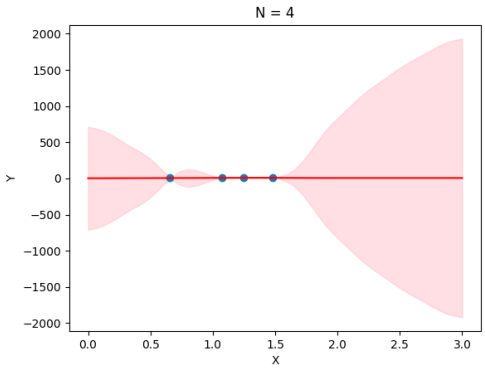
從觀察到的結果來看，在資料量少時，平均值的曲線較為平滑，隨著資料量增加而越曲折。

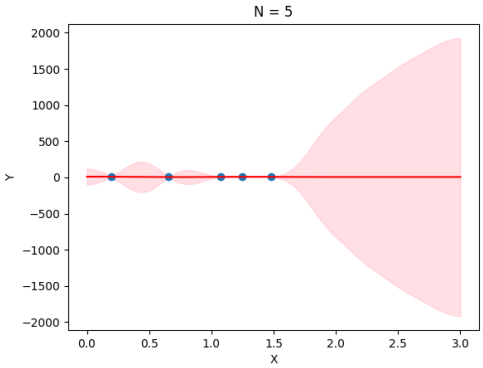
在資料量少時，標準差非常大，代表不確定性高，平均值的曲線對於預測較無把握，即便對於已知的資料，其標準差仍不小，只是相對於其他點把握較高而已。當資料慢慢變多時，整體的標準差逐漸減少，曲線對於預測的把握越來越高，最後只有 data point 較少的部分標準差較大。

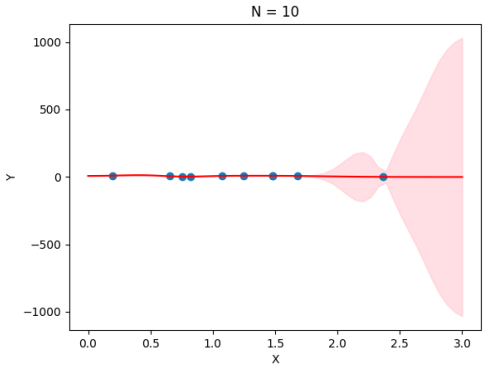
整體而言，當資料量少時，曲線非常依賴已知的 data point，對其他地方較無把握。當資料量增加後，曲線便不會過度依賴在已知的 data point 上（通過該點），而是有把握地找出整體資料分布的趨勢，較易於預測未知的部分。

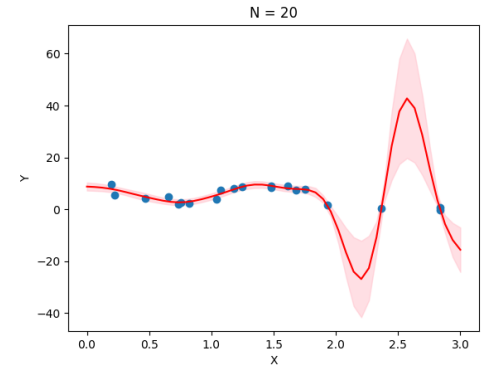


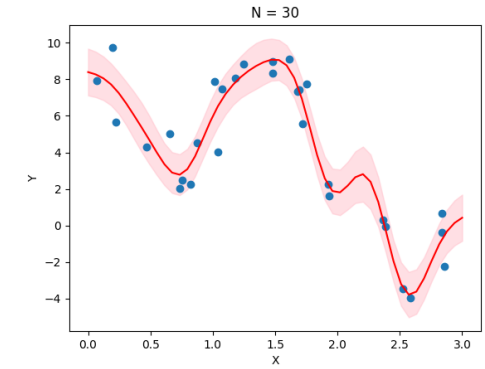


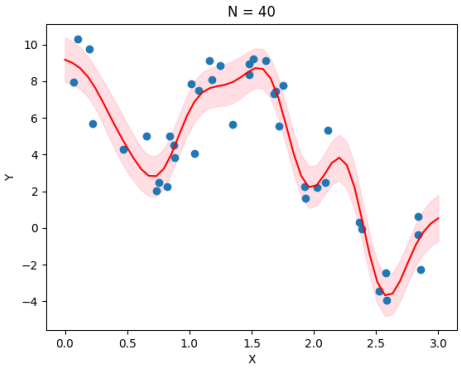


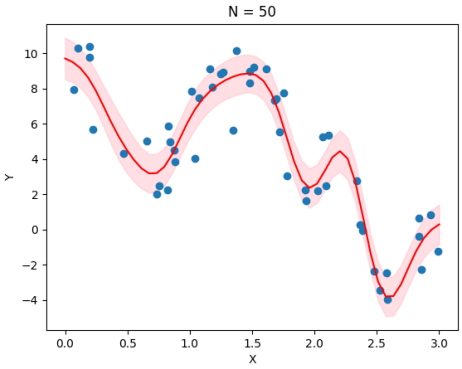









N = 40
Y
X
10
8
6
4
2
0
−2
−4
0.0
0.5
1.0
1.5
2.0
2.5
3.0


N = 50
Y
X
10
8
6
4
2
0
−2
−4
0.0
0.5
1.0
1.5
2.0
2.5
3.0