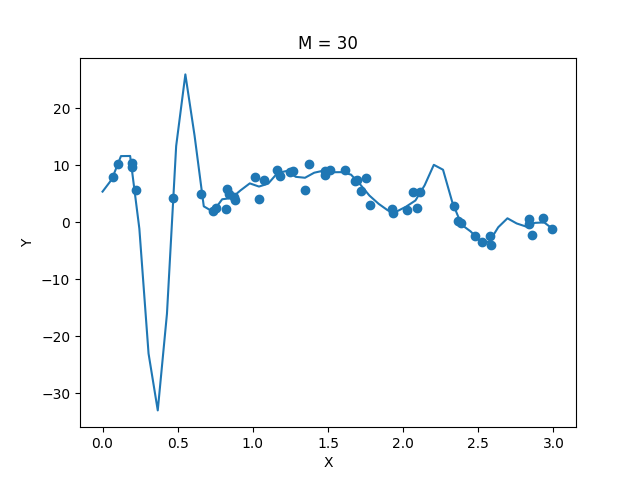
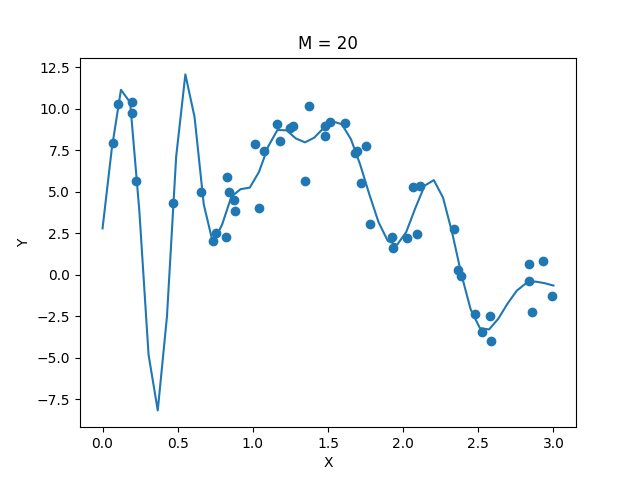
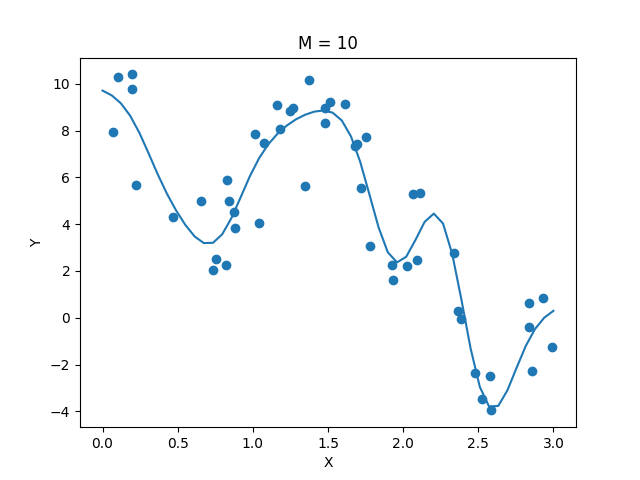
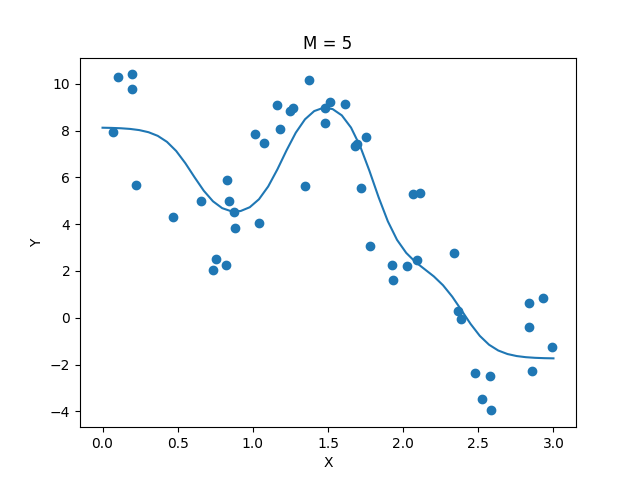
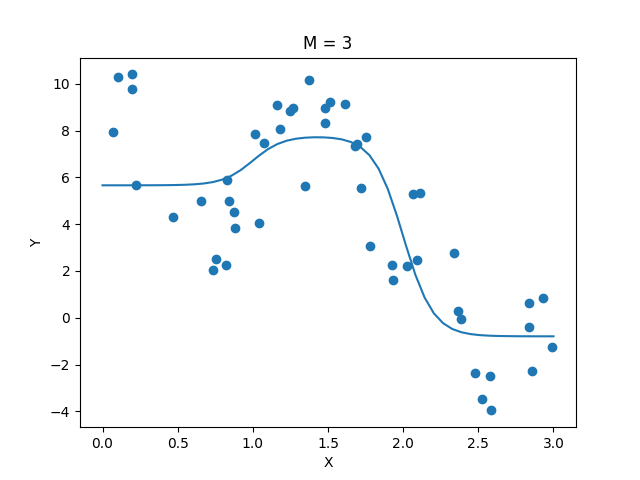
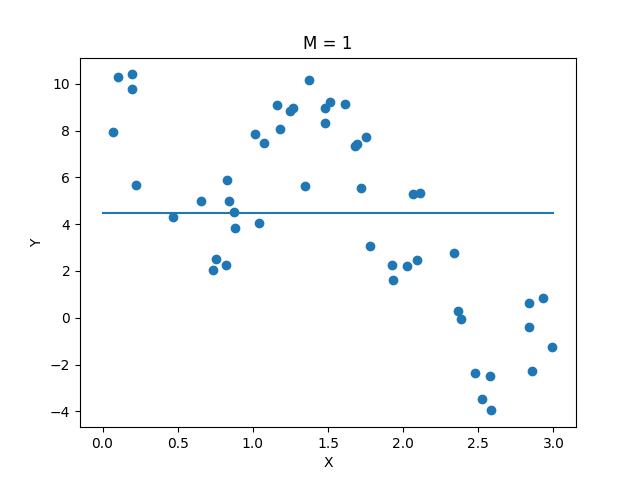
HW1: Linear Regression

109511219 林錦樑

**Part I. Linear Regression**

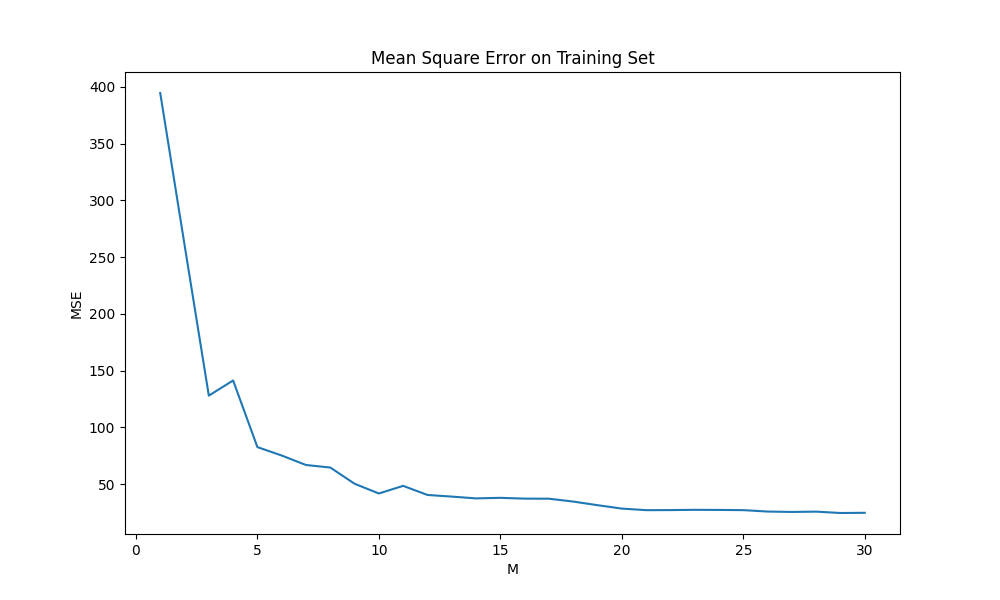
1. Fitting curve for M=1,3,5,10,20 and 30

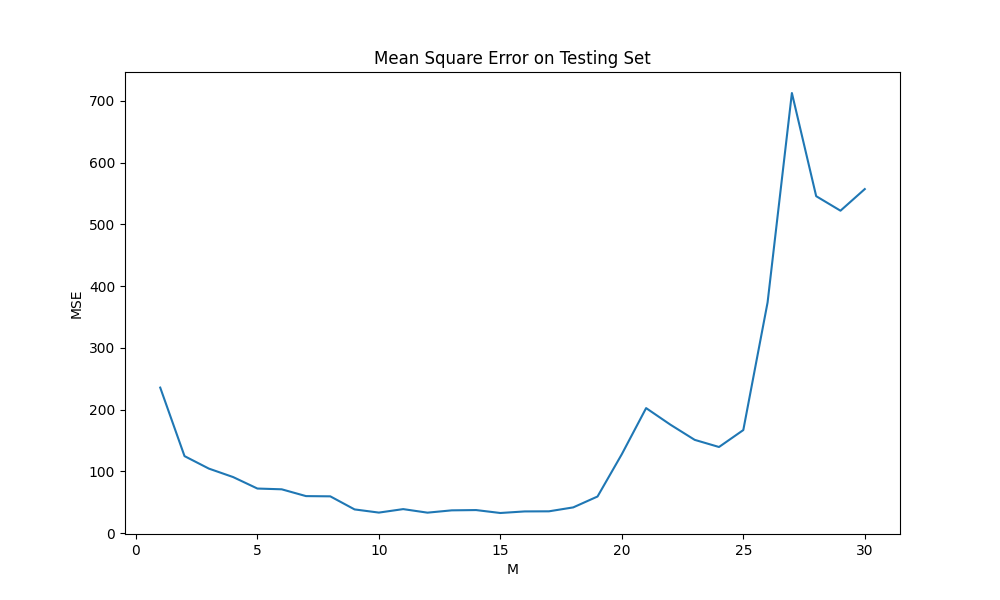
下面六張圖分別為不同M時的fitting curve與datapoint。可以看出當M=1時，fitting curve為水平線，Y與X無關。隨著M值增大，fitting curve越來越彎曲，且更加貼合datapoint。從M=20開始時，可以發現fitting curve在datapoint多的地方貼合得很好，但在沒有datapoint的地方時，會出現相當大的起伏。推測是fitting curve為了盡可能地貼合datapoint，使得W值變大，出現overfitting的現象。



1. 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。



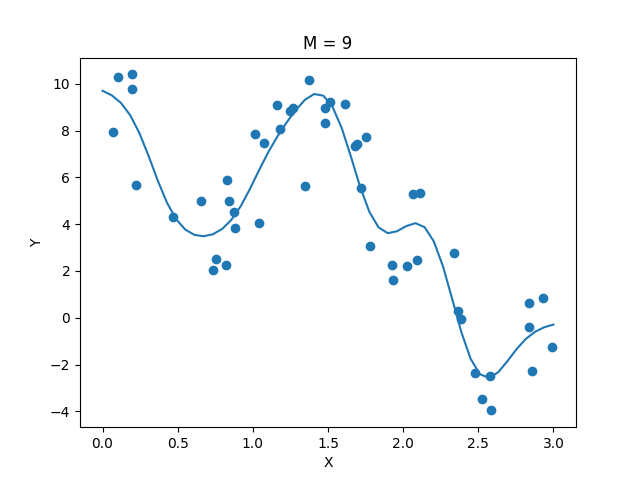


1. 5-fold cross-validation

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。

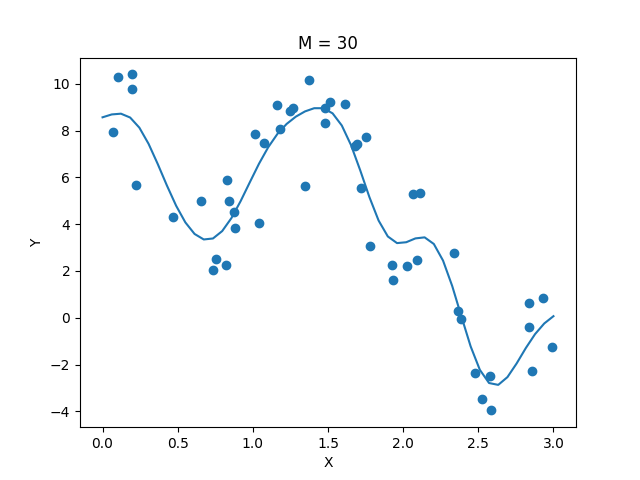
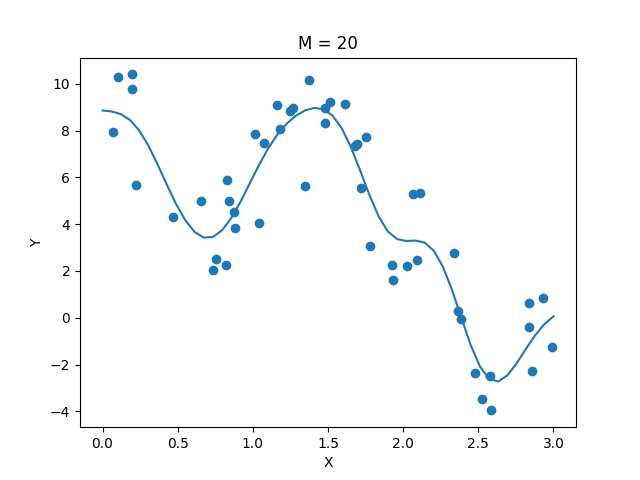
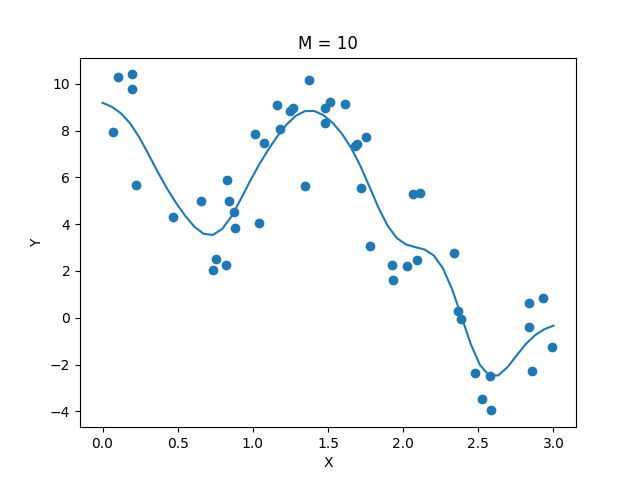
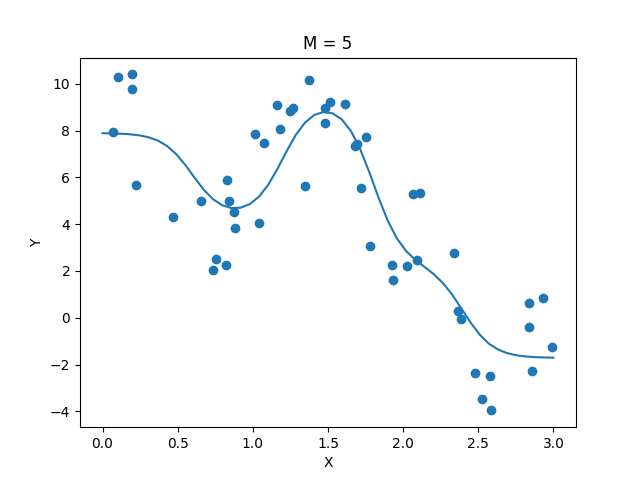
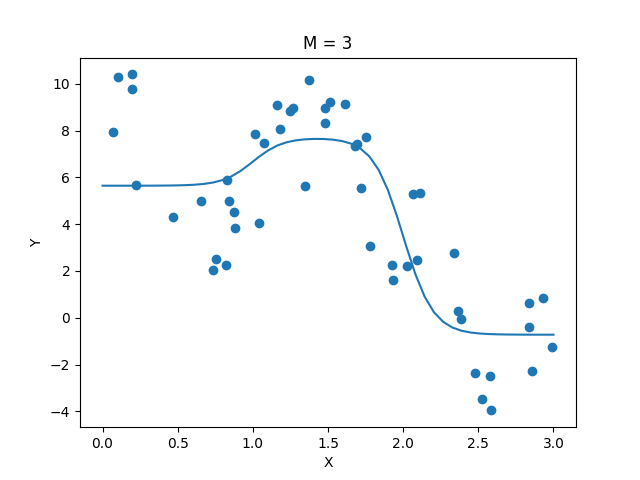
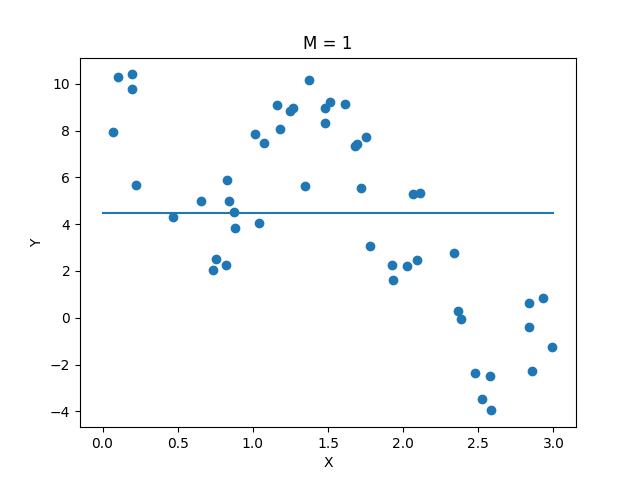
同個M值總共會訓練五次，每次訓練完後將model拿到驗證集去計算MSE，最後再將五次的MSE取平均，找出平均MSE最小值者，即可獲得最佳的M值，我透過5-fold cross-validation所得出的最佳M值為9。

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。

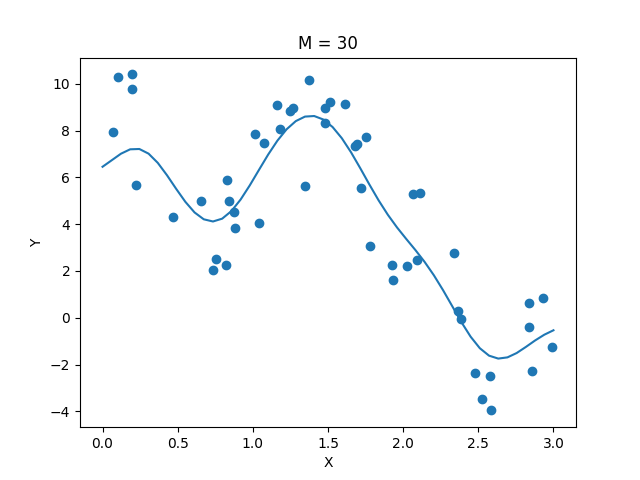
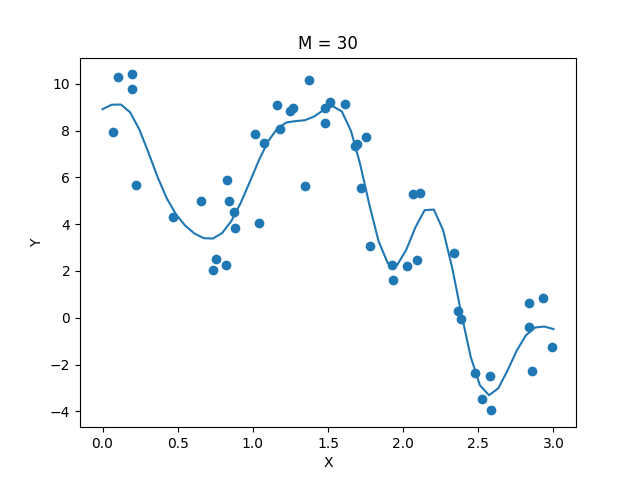


1. Regularization
   1. Fitting curve for M=1,3,5,10,20 and 30

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。



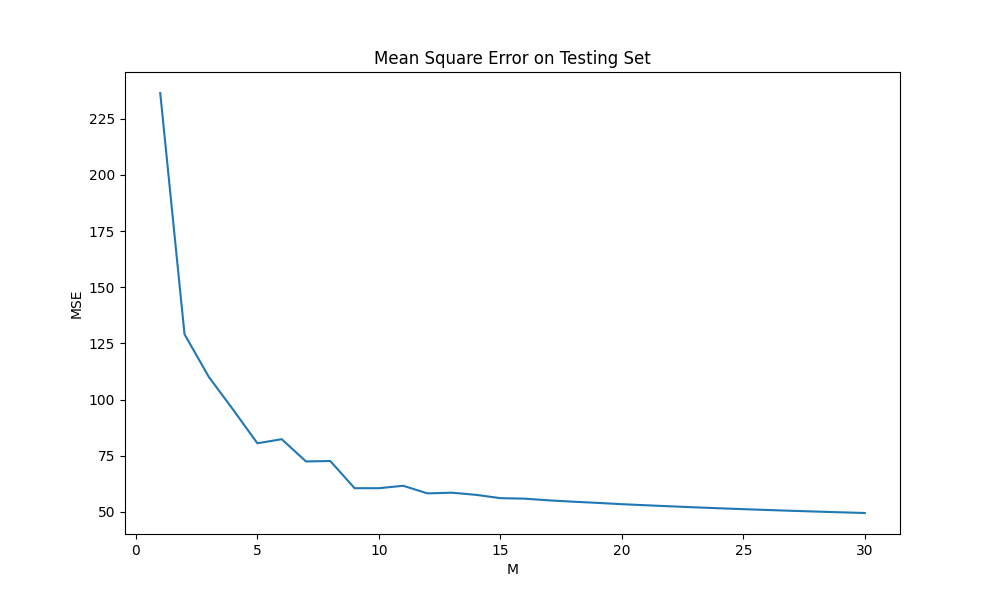
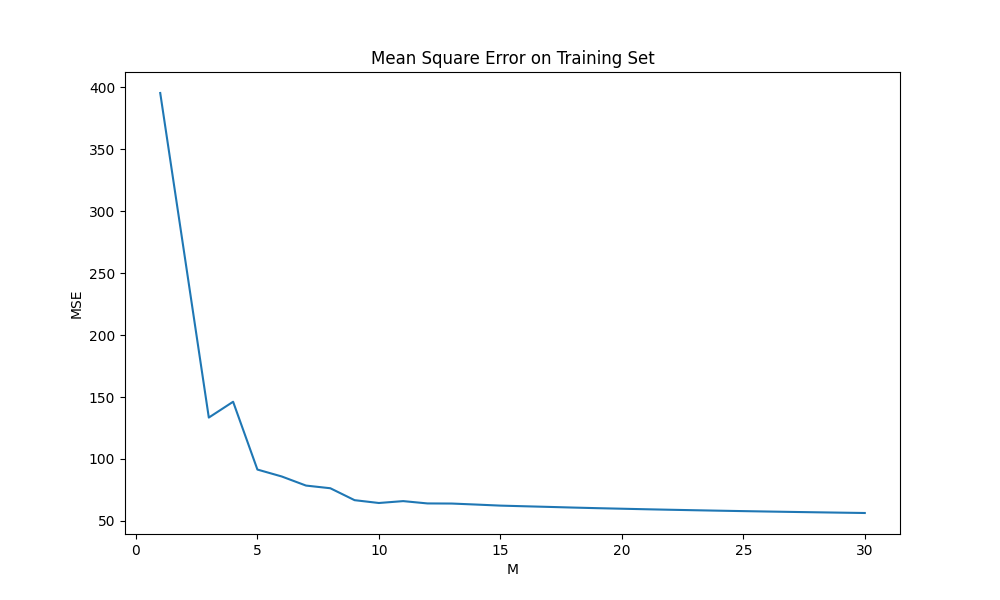
如果將λ調小到0.01，則fitting curve更彎曲，如下列左圖。λ調大到1，則fitting curve更平滑，如下列右圖。



* 1. 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。

