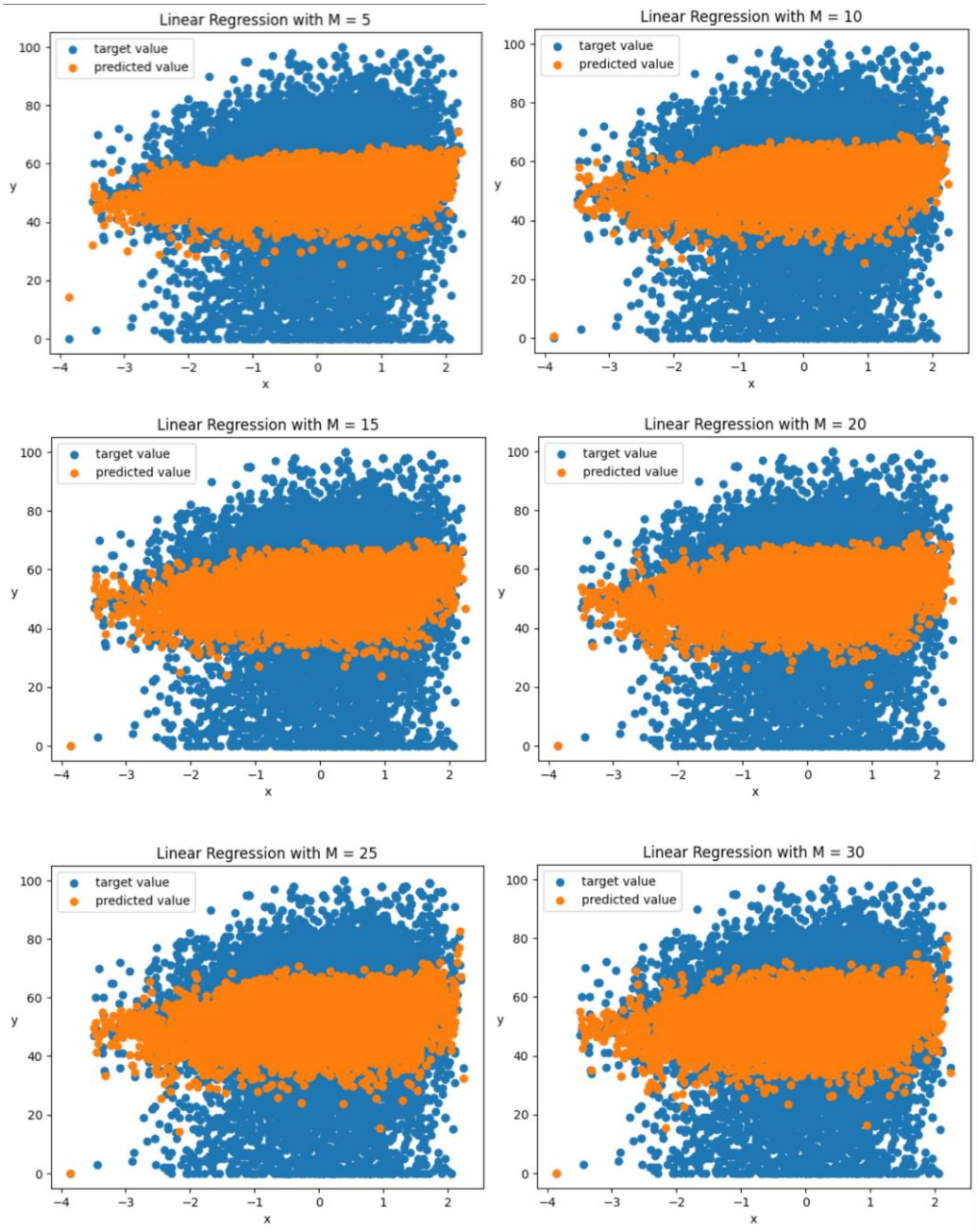
ML HW1 Report

110511277 蔡東宏

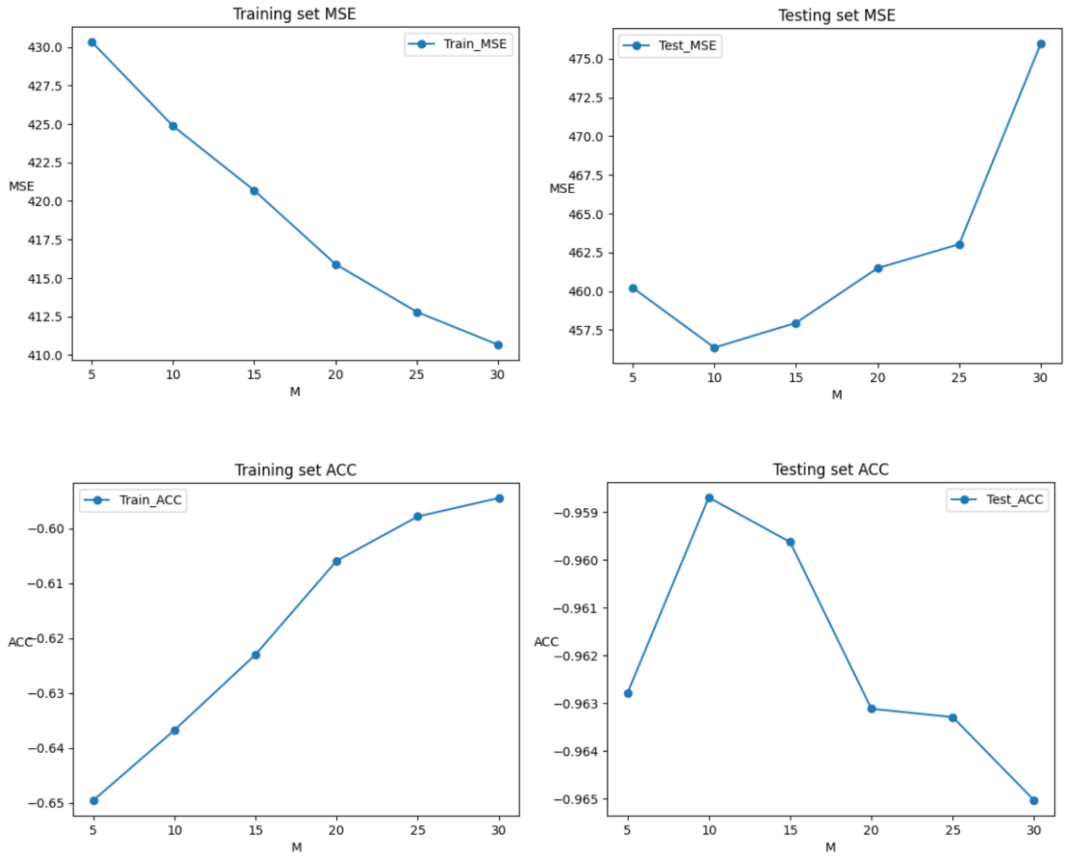
1. Please plot the fitting curve of the third input feature (x3: dace-ability) for M = 5, 10, 15, 20, 25, 30, respectively. (Change M for all input features, but you only need to plot the fitting curve of the third input feature.) 觀察 training set



觀察上面六張圖發現，當 M 越小時 training data 的 predicted value 的範圍越窄，而當 M 變大，training data 的 predicted value 的範圍越寬，因此我

們得知，當 M 越大時，training data 的 predicted value 能夠越接近 target value，因此大概可以推斷當 M 變大時 training set 的 MSE 越小，ACC 越接近 1。

2. Please plot the Mean Square Error evaluated on the training set and the testing set separately for M = 5, 10, 15, 20, 25, 30. Also, evaluate the accuracy of the training set and the testing set given as follows:

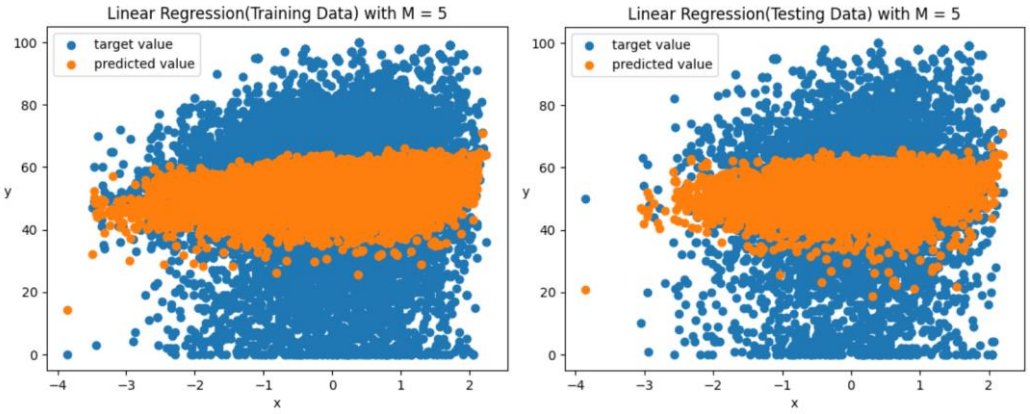


觀察當 M 提高時，training set 的 MSE 持續變小且 ACC 持續變大，代表 M 越大時，training set 的 predicted value 越接近 target value。然而觀察 testing set 的 MSE 以及 ACC 發現，隨著 M 變大，MSE 會先變小再逐漸變大，而 ACC 會先變大再逐漸變小，再 M = 10 時，MSE 會最小，ACC 會最大。因此我們可以得知，較大的 M 可以將 training data fit 得比較好，然而，較大的 M 可能會讓 testing data 出現 overfitting 的現象，失去 generalization。除此之外，我們可以發現無論 M 的大小，testing set 的 MSE 都比 training data 的 MSE 大，testing set 的 ACC 都比 training data 的 ACC 小，代表 training set 出來的資料比較精準。

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 of the third input feature (x3: daceability). You should briefly express how you select the best order M step-by-step.

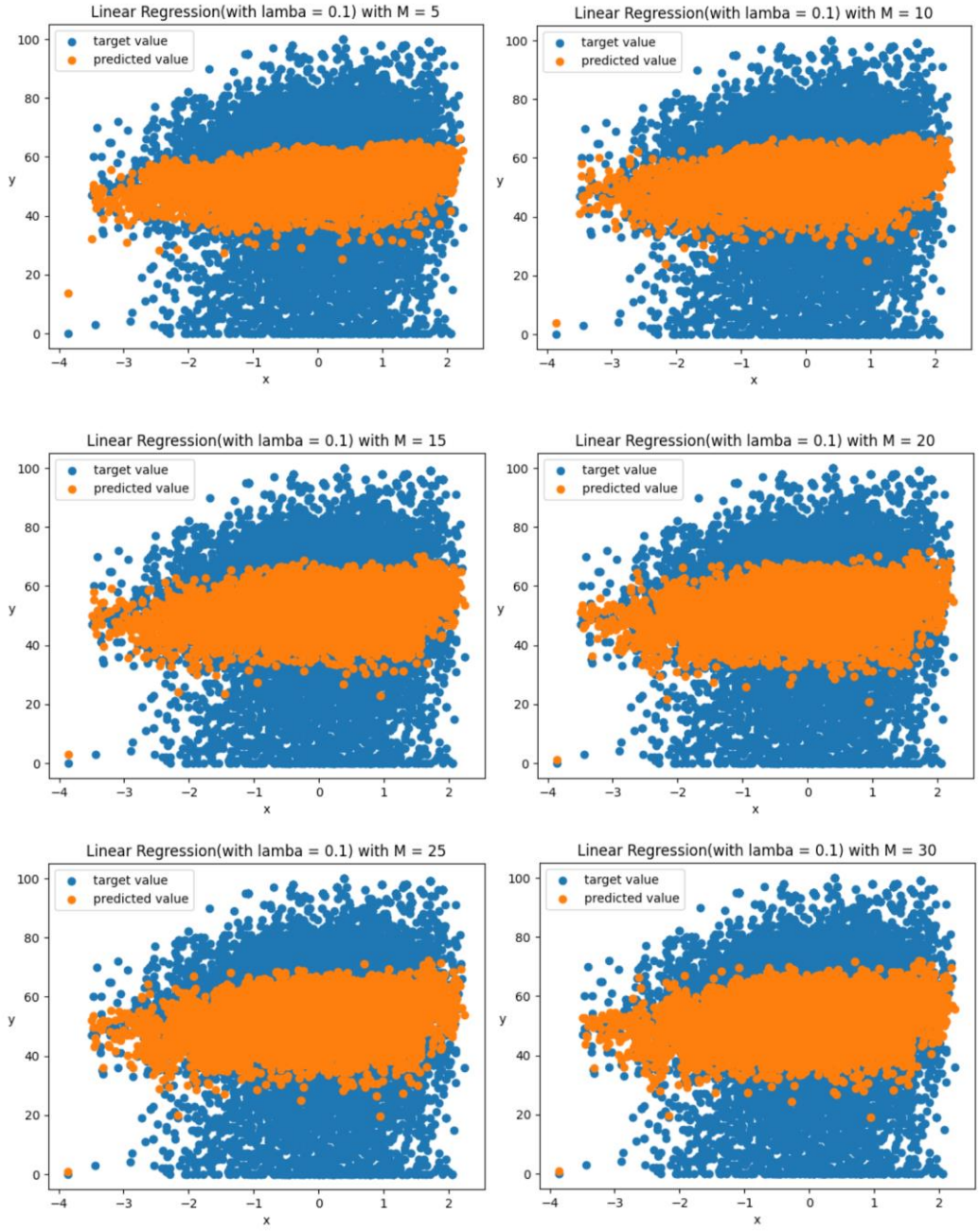
```
MSE for M = 5 : 456.9653834284075
MSE for M = 10 : 518.8828044927192
MSE for M = 15 : 481.3041051064837
MSE for M = 20 : 4842.462405506827
MSE for M = 25 : 565169.1168087681
MSE for M = 30 : 1459103.0577094248
```

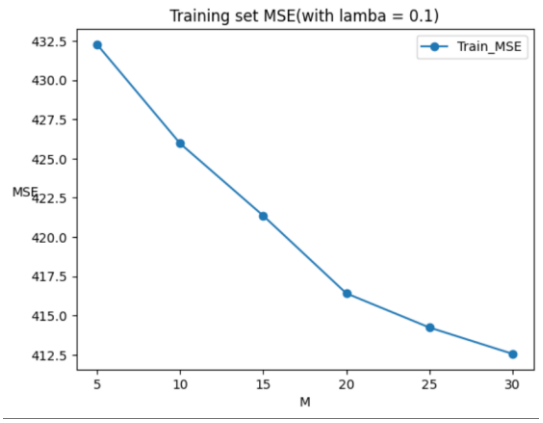
將 10000 筆 training data 拆成五組，每次取四組當作 training set，另外一組當作 validating set 算出 MSE，五種情況都考慮的情況下，將所有的 MSE 加總起來並取平均，並找出 MSE 最小的 M。我們會發現 MSE 最小時，M = 5，因此利用 M = 5 來分別對 Training Data 以及 Testing Data 進行繪圖。

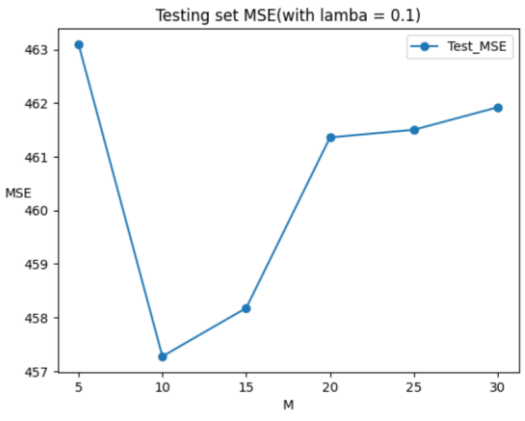


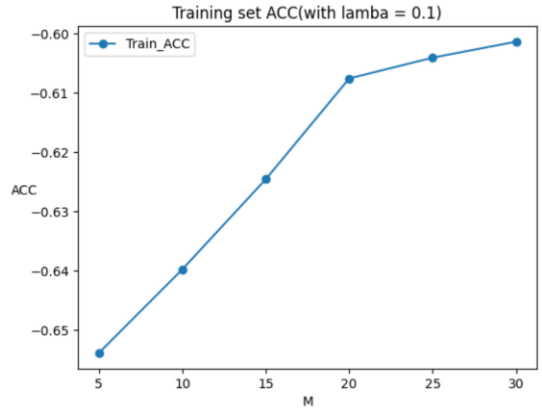
此題是為了找出最適合的 M，若 M 太小，模型不夠複雜而不能更精確地建出模型，然而若 M 太大，會導致模型太過複雜而造成 overfitting 的現象，也無法經確的建出模型，因此利用 5-fold cross validation 找出最小 MSE。而這個 M 會因為如何分組而有不同的結果，我是直接將 data 拆成 1~2000、2001~4000、4001~6000、6001~8000、8001~10000，如此得到的最小值 MSE 時，M = 5。

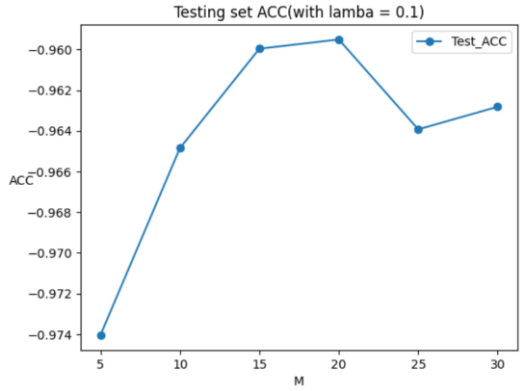
4. Considering regularization, please use the modified error function. Repeat Part I -1. and Part I-2. with λ =1/10 (You can also try to change the value of λ and discuss what happens under different λ values.)



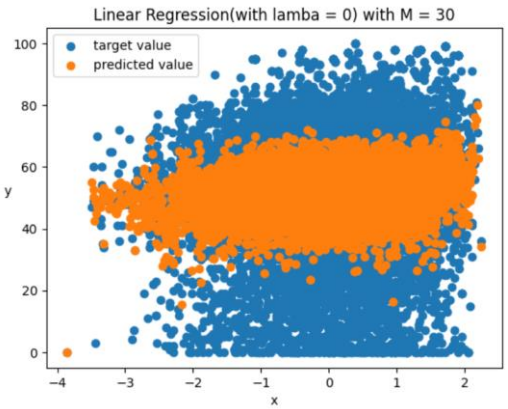


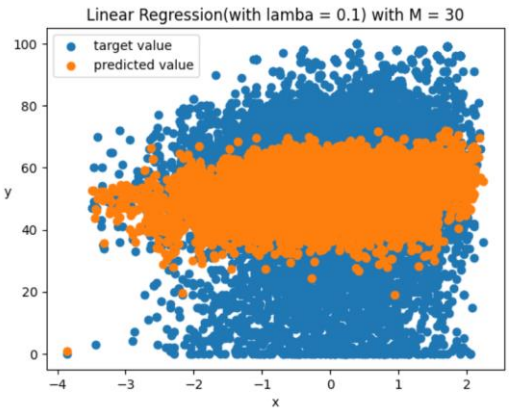


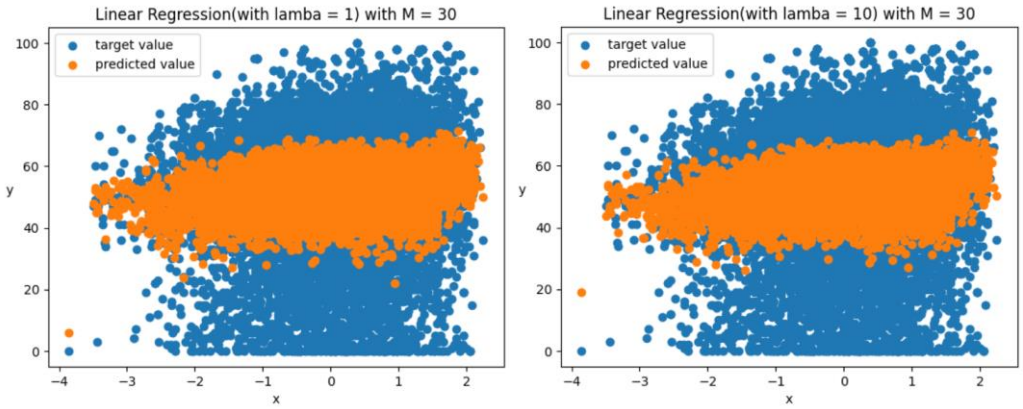


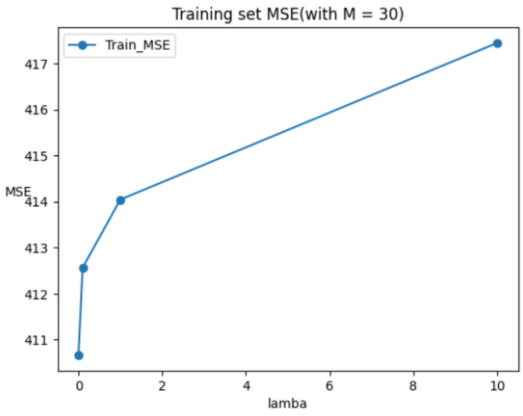


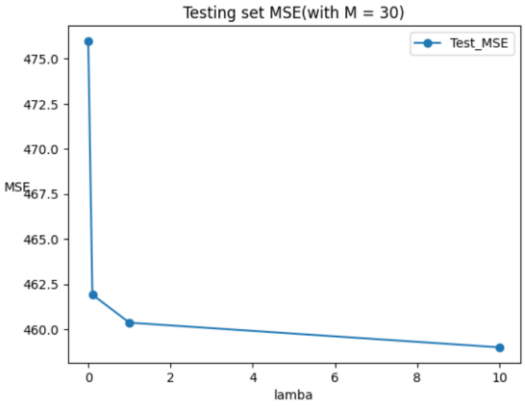
觀察 training set 的 MSE 以及 ACC，我們發現它的趨勢與 lamba = 0 時得一樣，MSE 皆隨著 M 變大而變小，而 ACC 皆隨著 M 變大而變大，而 testing set 的 MSE 趨勢也於 lamba = 0 時一樣，在 M = 10 時，MSE 最小。加了 lamba 之後發現，testing set 中較大 M 的 MSE 不會像 lamba = 0 時那麼大。加入 regularization term 是為了避免出現太大的 w 來造成 fitting curve 的 overfitting，特別是對於 M 很大時，加入 regularization 才能避免選擇太大的 w 造成 overfitting，如此一來，才能減少 M 很大時的 MSE。

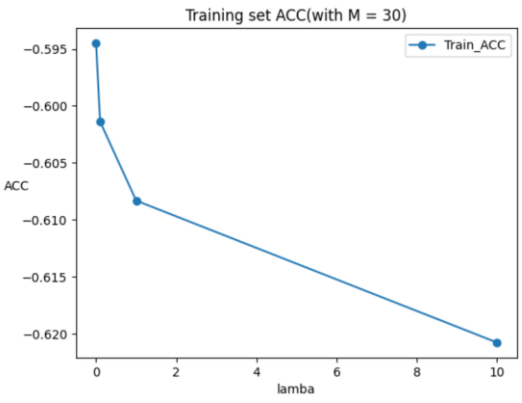


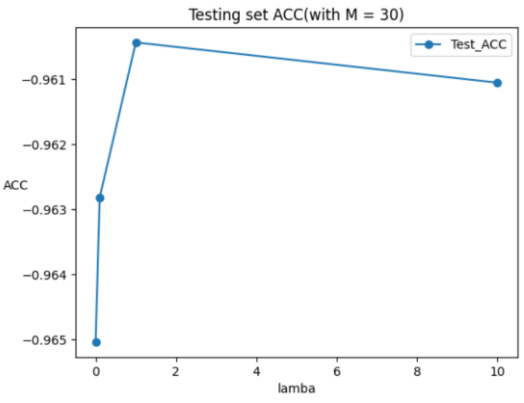












最後我固定 M = 30，調整 lamba 來觀察結果，我們可以發現對於 training set 的 MSE 會隨著 lamba 變大而變大，然而觀察 testing set 發現，MSE 反而隨著 lamba 變大而變小。Lamba 越大代表 w 的值會越小，fitting curve 會越平滑，比較不容易出現 overfitting 的情況，所以我將 lamba 調大時，training 的模型會比較不複雜，因此 training set 的 MSE 越大;然而 lamba 調大，導致 overfitting 越小，因此 testing set 的 MSE 就越小。

問題討論:

1. **If you change the number of basis functions, M, what effect will have on training and testing?**

   對於 training test 而言，隨著 M 增加，它的 MSE 以及 ACC 皆變好，其原因是因為當 M 變大時，模型的複雜度也會跟著變大，所以建出來的模型也會相對準確。

   對於 testing 而言，雖然 M 變大會增加模型的複雜度，但如果模型太過複雜會有 overfitting 的現象產生，因此可以得知增加 M 不一定能增加 generalization，而 M 太小會因為模型不夠複雜而不符合預期的 data，因此必須選擇適當的 M。

2. **Which features do you consider the most important? Why do you consider your selected features to be the most important?**

   Key 和 speechiness 是最重要的兩個 features，判斷最重要的 feature 是看它們的 weight，因此我將他們的 weight 分別印出來，並將同一個 feature 的權重取絕對值並相加，發現第六個 feature (Energy)以及第九個 feature(speechiness)的權重最高，而 speechiness 的權重又比 Key 的權重高，因此我認為他們兩個是最重要的 features。

```
[ 545.11072887 1910.37104784  695.84303581  718.60422242 2281.6692197
 4818.72927972 1545.46976224  904.55776466 9816.00889813  640.25510571
 1886.40106774]
```