ML HW1 Report

姓名：陳品媛

學號：A063021

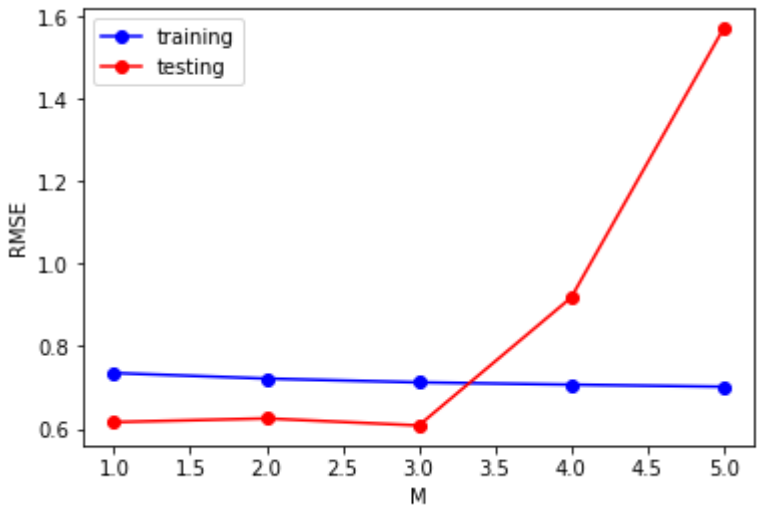
Problem 3 Polynomial Regression

首先我對資料有做shuffle，避免資料有按照某種順序做排序，另外我對資料做標準化的動作。

Problem 3-1

1. Apply the polynomials of order M = 1 to M = 3 over the 3-dimensional input data
2. Evaluate the corresponding Root-Mean-Square error (ERMS = root(2E(w)/N)) on the Training Set and Test Set
3. Plot their RMS error versus order M .
4. Describe in details about what you see in the plot.

My discussion




```
M=1, training error: 0.735338, testing error: 0.616191
M=2, training error: 0.721775, testing error: 0.625101
M=3, training error: 0.712586, testing error: 0.608078
M=4, training error: 0.707082, testing error: 0.918149
M=5, training error: 0.702049, testing error: 1.570390
```

這個dataset很剛好的在一開始對testing data fit的很好，因為理論上應當training data的RMSE要比testing data來的低，但是在M=4之後明顯誤差值往上飆高，原因是對training data更加的fit以至於對於testing data誤差會更大。而training data的部份呈現不斷遞減。

另外，也有遇到助教公告的M越大 training error 會變大的反矩陣問題，原先使用inv會有計算誤差，導致問題的產生。因此，用*Gauss-Jordan method*計算反矩陣，解決這個問題。

Problem 3-2

1. Select the most contributive attribute or dimension which has the lowest RMS error on the Training Set

```
Dimension: 2, Attribute_indexes:  [0, 1]
RMSE:  0.9327096193022578

Dimension: 2, Attribute_indexes:  [0, 2]
RMSE:  0.7270082413880575

Dimension: 2, Attribute_indexes:  [1, 2]
RMSE:  0.7255938183712136
```

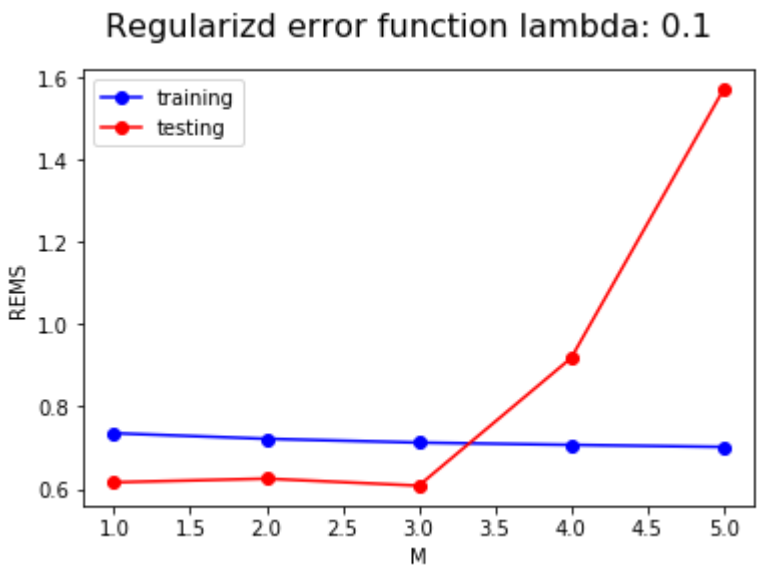
My discussion

從三種attribute任選兩個所組合出來的結果，可以看出用total_rooms, population所組合出來的結果誤差值是最高的。 因此可以得知median_income會是最有貢獻的attribute。

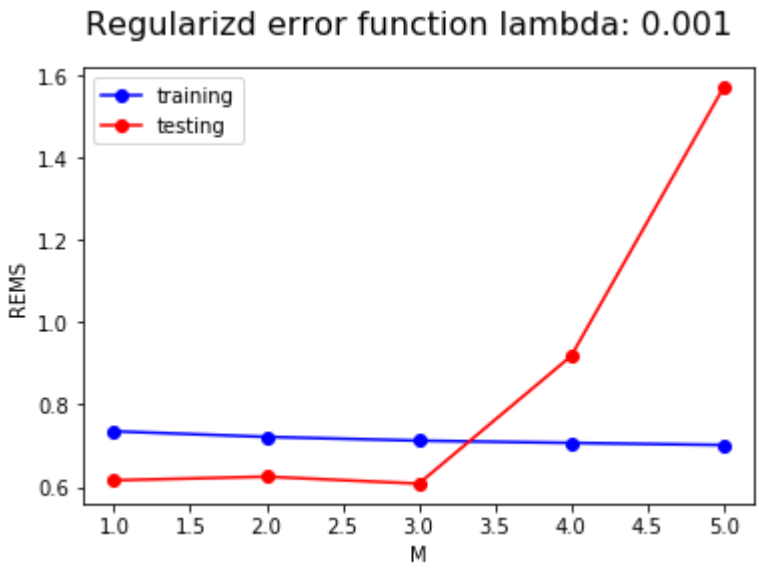
Problem 3-3

1. Set two values for regularization parameter as λ = 0.1 and λ = 0.001 and repeat part 1
2. RMSE = root(2E(w)/N) is calculated using E(w) not E(w) hat
3. Plot the regularized regression result on Training Set and Testing Set for various order M from 1 to 3. Compare the result with different λ
4. Describe the difference between part 1 and part 3.

My discussion



```
M=1, training error: 0.735338, testing error: 0.616192
M=2, training error: 0.721775, testing error: 0.625097
M=3, training error: 0.712586, testing error: 0.608074
M=4, training error: 0.707082, testing error: 0.917825
M=5, training error: 0.702049, testing error: 1.569162
```




```
M=1, training error: 0.735338, testing error: 0.616191
M=2, training error: 0.721775, testing error: 0.625101
M=3, training error: 0.712586, testing error: 0.608078
M=4, training error: 0.707082, testing error: 0.918146
M=5, training error: 0.702049, testing error: 1.570378
```

Regularized error function的用意是在於為了避免overfit的狀況產生，所以加上懲罰項。 但以M:1~3做訓練，由於誤差值本身很高，所以不存在overfit的問題，因此加上regularize term並不會有顯著的效果。