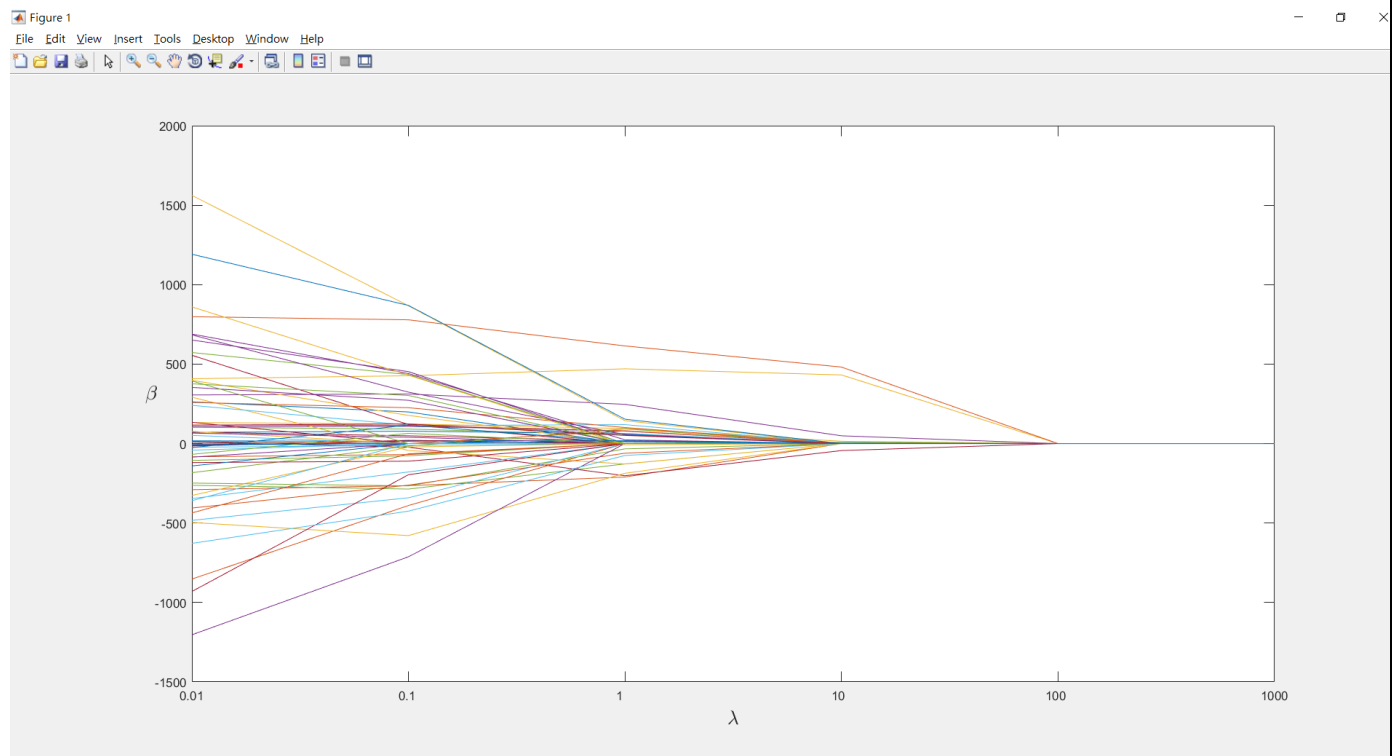
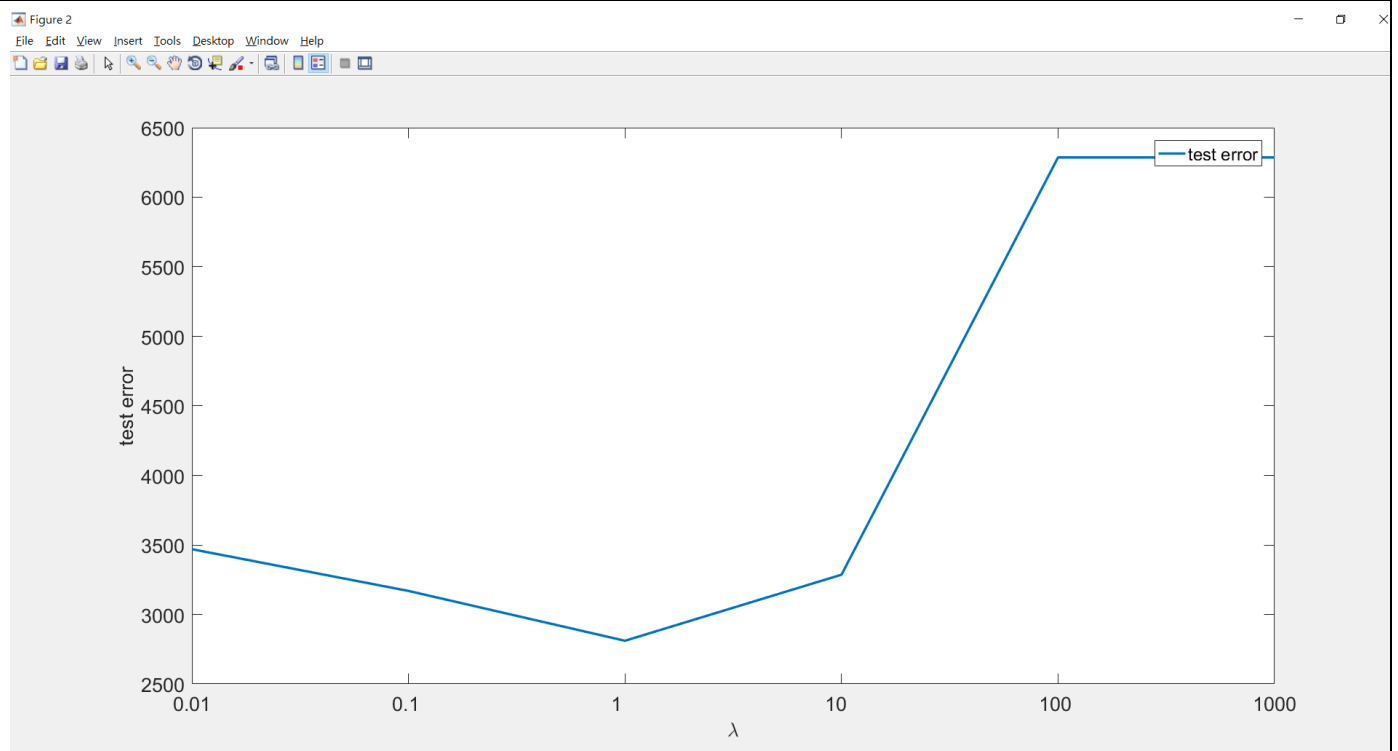


## Problem 1 (LASSO and Ridge regression)

(a) Plot of weights  $\beta$  for each  $\lambda$ 

## Discussion:

上圖為 64 種不同特徵，由 LASSO regression 改變  $\lambda$  得到的  $\beta$  值，可以看出隨著  $\lambda$  增加，各個特徵的  $\beta$  值“分別”歸 0。

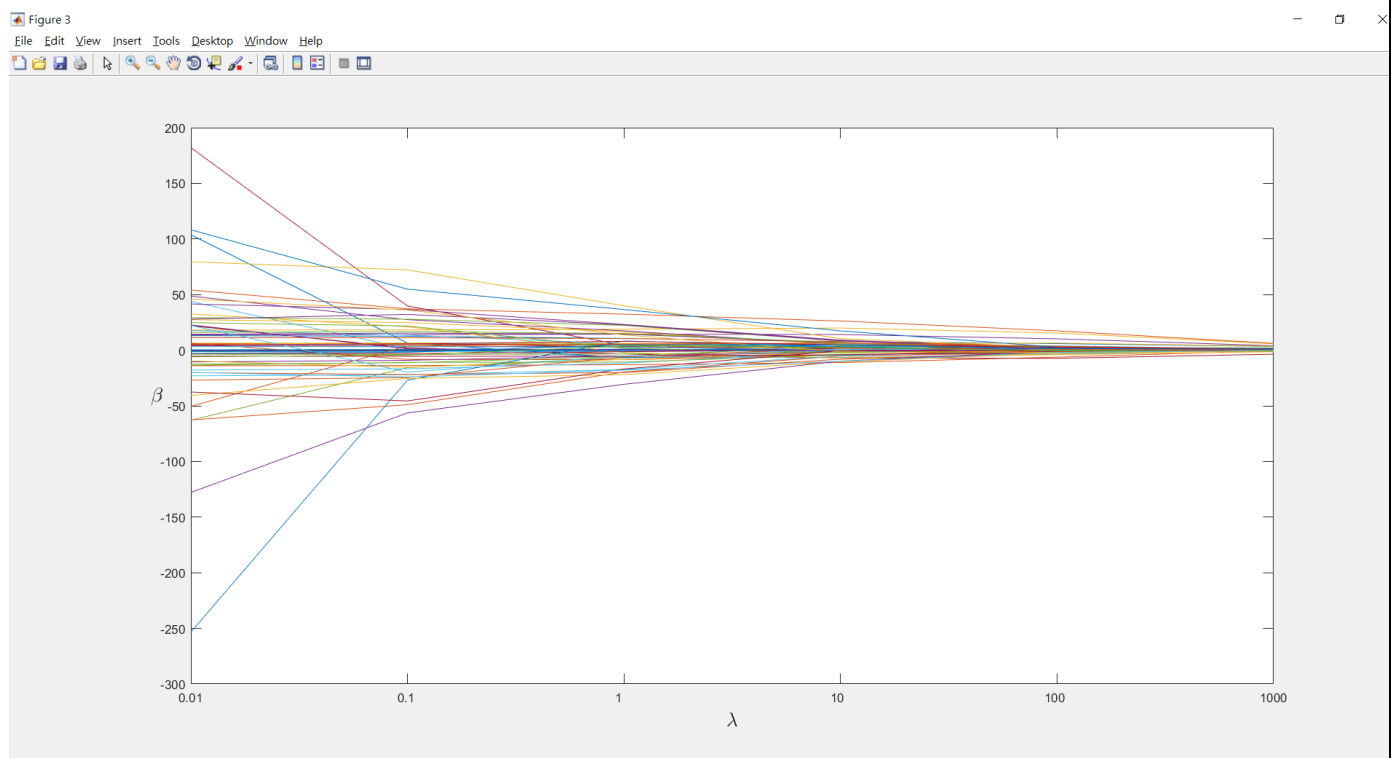
(b) Plot of test error for different  $\lambda$ 

## Discussion:

test error 隨著  $\lambda$  增大而縮小，但是又隨之增大。由(a)的圖可以了解，在  $\lambda=1$  之前， $\beta$  值隨著  $\lambda$  有些

已經歸 0，因此可以篩選出較有用的 $\beta$ 值；但隨著 $\lambda=10$  之後，幾乎所有 $\beta$ 值都歸 0 了，造成 model 不準確，所以 test error 又開始升高。

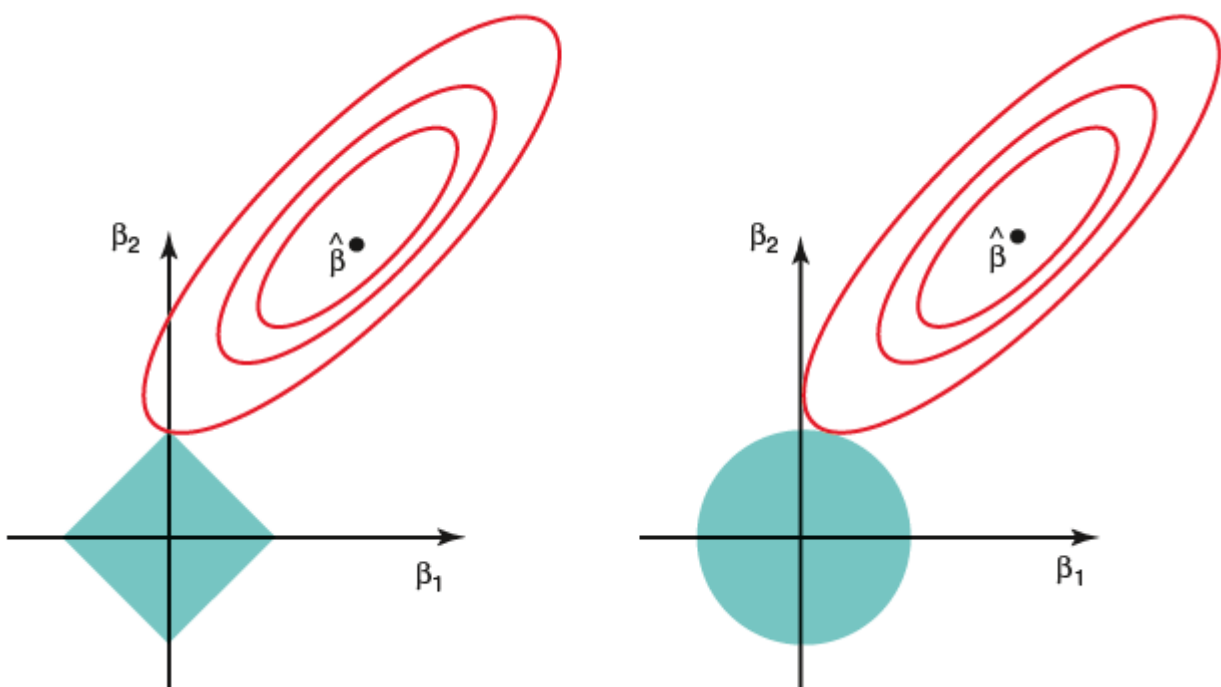
(c) Comparing the solution  $\beta$  obtained by the ridge regression to that by LASSO



Discussion:

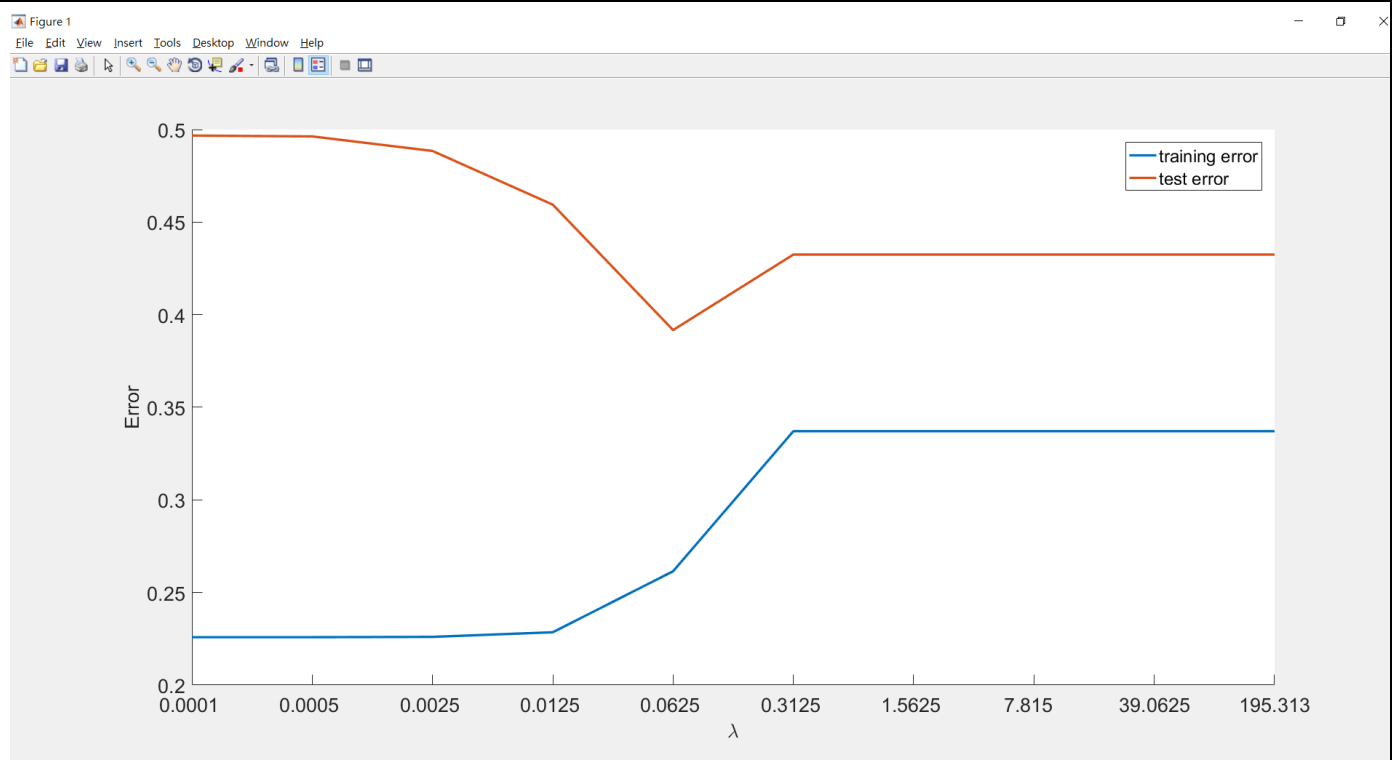
上圖為 64 種不同特徵，由 ridge regression 改變  $\lambda$  得到的 $\beta$ 值，可以看出隨著  $\lambda$  增加，各個特徵的 $\beta$ 值”同時”趨近於 0”。

ridge regression 的 constraint 為 circular constraint，此 constraint 並沒有尖銳的點，因此和橢圓交界處並不會在軸上，所以 $\beta$ 只會趨近於非零的值（如下圖右）。而 LASSO regression 的 constraint 為 diamond constraint，此 constraint 有尖銳的點，因此和橢圓交界處就會在軸上，此時，就會有部分 $\beta$ 歸零。（如下圖左）



## Problem 2 (LASSO regression)

### (a) Plot the training and test errors versus different $\lambda$



### (b) Briefly explain the fitting behavior of the models with different $\lambda$

由上圖我們可以看到，隨著 $\lambda$ 增加，training error 逐漸上升，test error 先降後升。一開始 training error 小，但 test error 高，可以觀察到可能有 over-fitting 的現象；之後由於當 $\lambda$ 增加到一定數目時， $\beta$ 就會歸 0，因此使得整個特徵的權重都是 0，就會有 under-fitting 的現象，使得 training error 和 test error 同時上升。

### (c) Which $\lambda$ would you choose to train your final model? Why?

一般來說可以用 cross-validation error，並選擇 one standard error 不再增加的 $\lambda$ ，做為最好的 model。不過這邊單就上圖來看，我們可以選擇 test error 最小的點，也就是 $\lambda = 0.0625$ 的地方，做為最後的 model，因為此時雖然 training error 略為上升，但 test error 大幅下降，代表這個地方擁有不錯的 model。