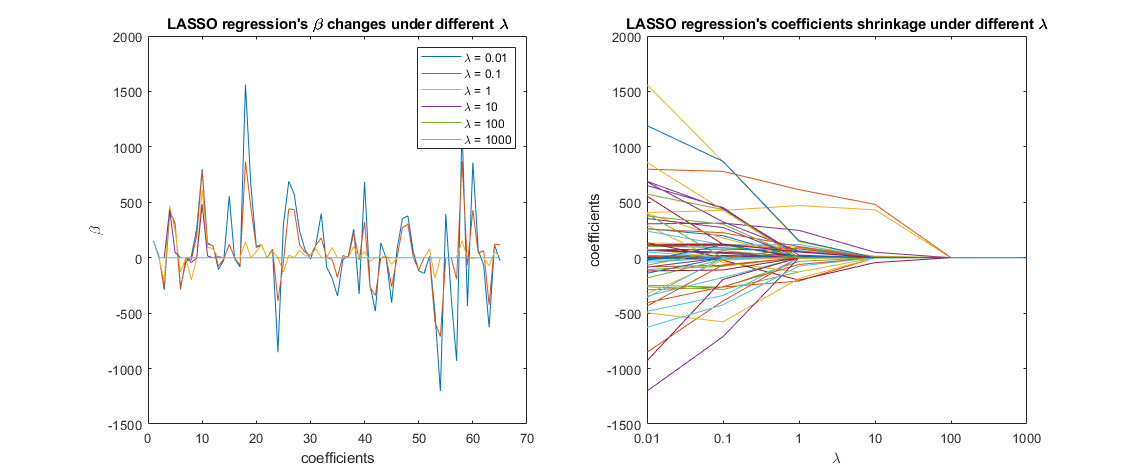
Problem 1

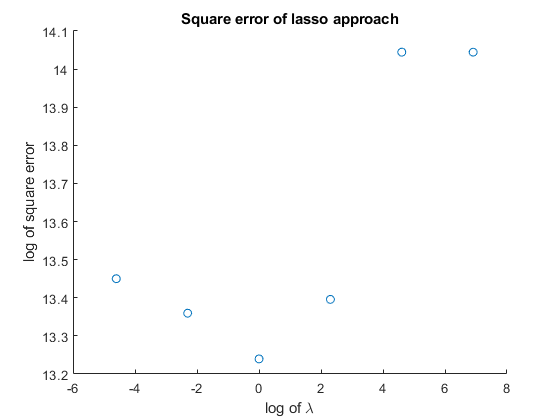
a)



Figure

從上圖左可以看出隨著lambda變大，震盪幅度愈來愈小，表示weighting縮成0的參數數量愈來愈多；從上圖右也可看出所有的參數在lambda變大時，其weighting都往0靠近。符合上課老師所講解的情況。

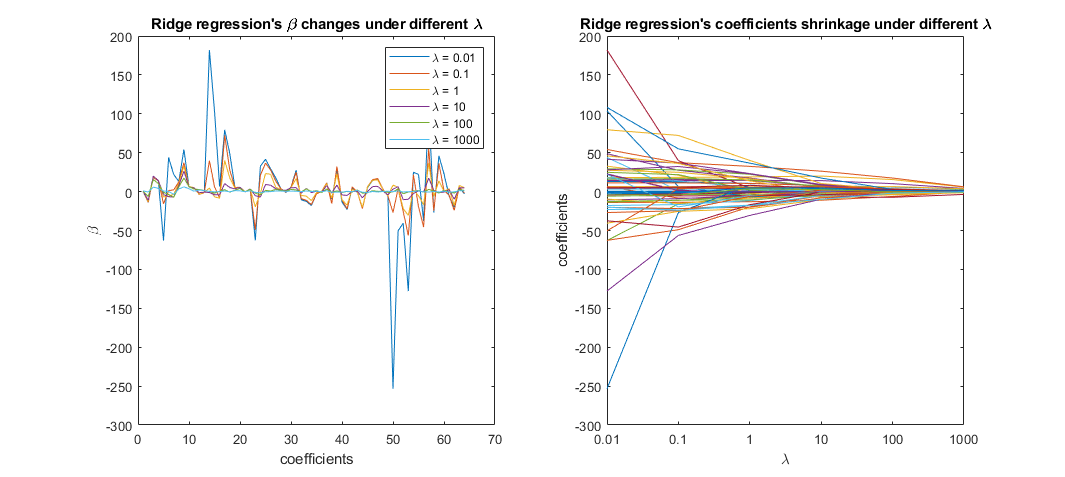
(b)



Figure

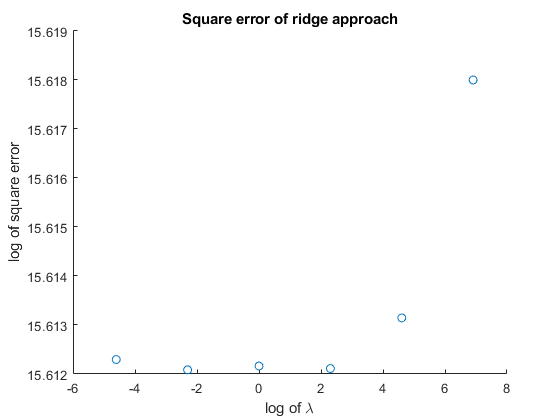
隨著lambda放大，誤差確有因多餘的參數被削去而達到降低的效果，但過度簡化模型反而使誤差在更大的lambda時暴增，所以上圖呈現的此種V型曲線是符合預想的結果。

(c)



Figure

和LASSO的結果比起來，Ridge regression的參數weighting縮得比較慢，甚至到lambda=1000時都還有很多參數被保留下來。

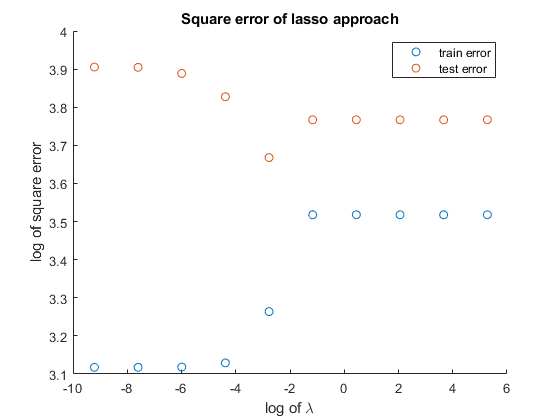


Figure

從rigde regression看不太出來像LASSO那樣的最佳lambda選擇(V型下降)，不過後面減少太多參數所導致的誤差暴增還是看得出來。

Problem 2

(a)



Figure

(b)

Figure 5由黃線區隔開的部分，左邊為underfitting、右邊為overfitting，因為test error在左邊仍有下降的空間，且training errorg尚小而不太變動；右邊則是training error已經過大，忽略的參數過多，表示overfitting。

(c)

所以最好的選擇就是中間的lambda = 0.0625。