|  |
| --- |
| R05631027\_楊皓文\_機器學習\_作業#5 |
|  |
| Problem 1 (LASSO and Ridge regression) |
| (a) Plot of weights 𝜷 for each λ |
|  |
| Discussion:  上圖為64種不同特徵，由LASSO regression改變λ得到的𝜷值，可以看出隨著λ增加，各個特徵的𝜷值”分別”歸0”。 |
| (b) Plot of test error for different λ |
|  |
| Discussion:  test error隨著λ增大而縮小，但是又隨之增大。由(a)的圖可以了解，在λ=1之前，𝜷值隨著λ有些已經歸0，因此可以篩選出較有用的𝜷值; 但隨著λ=10之後，幾乎所有𝜷值都歸0了，造成model不準確，所以test error又開始升高。 |
| (c) Comparing the solution 𝜷 obtained by the ridge regression to that by LASSO |
|  |
| Discussion:  上圖為64種不同特徵，由ridge regression改變λ得到的𝜷值，可以看出隨著λ增加，各個特徵的𝜷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𝜷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𝜷歸零。(如下圖左) |
|  |
| Problem 2 (LASSO regression) |
| (a) Plot the training and test errors versus different 𝜆 |
|  |
| (b) Briefly explain the fitting behavior of the models with different 𝜆 |
| 由上圖我們可以看到，隨著𝜆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𝜆增加到一定數目時，𝜷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 𝜆 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𝜆，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𝜆 = 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。 |