

Question 1

**** How does λ affect the MSE in general?**

--> Lambda is reduces over fitting or the predicted model, hence as the lambda increase is reduces the MSE error in test data. but it adds bias to the model so it will increase the MSE error in training data. In general as lambda increase it will under fit the data. The higher the lambda value it will under fit data, the lower the lambda value it will over fit the data.

**** How does the choice of λ depend on the number of features vs. examples?**

--> As the number of features is increasing the choice for optimal value for lambda also increases, because with more features the model will try to over fit the data to reduce this we need to add higher bias to model, which will result into higher lambda value,

This is observed in plot 100_10 vs 100_100.

As the number of example increase the value of optimal lambda also increase, because, as the number of example increase the weight calculated using this examples try to predict as closely as possible, hence to reduce this over fitting we need to add more bias to the model which result into higher lambda value.

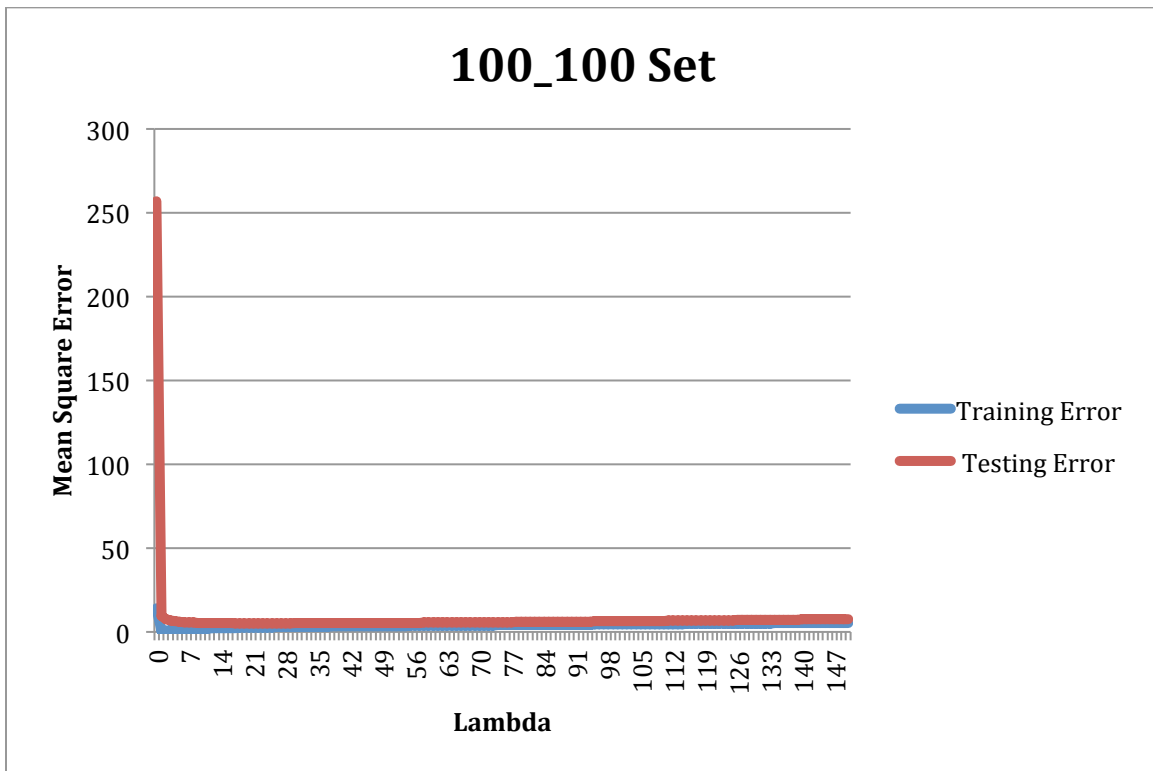
This is observed in plot 100_100 and 100_1000.

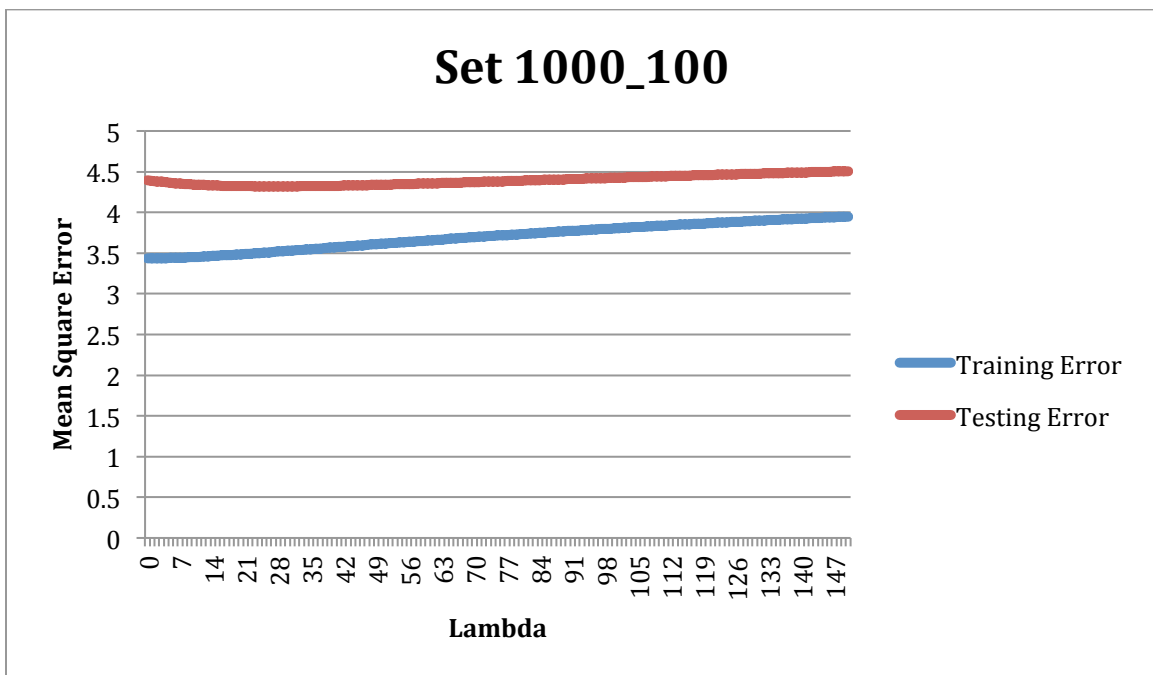
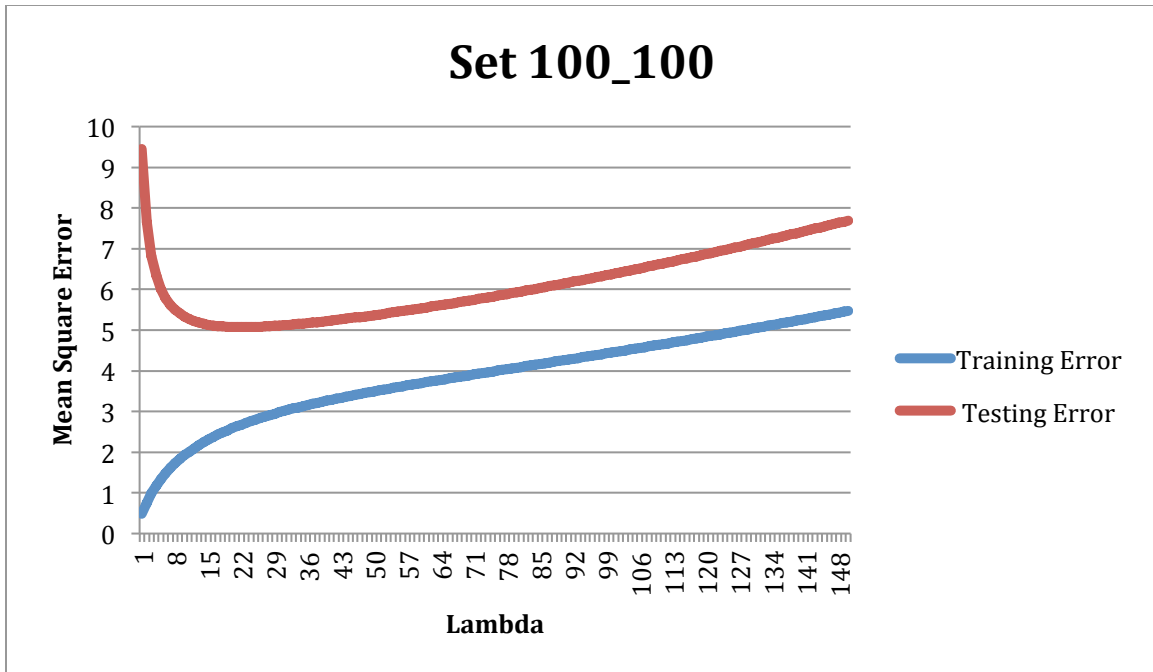
Hence, the general observation for lambda's value increases as the number of features increases or number of examples increase.

**** How does λ change with number of examples when the number of features is fixed?**

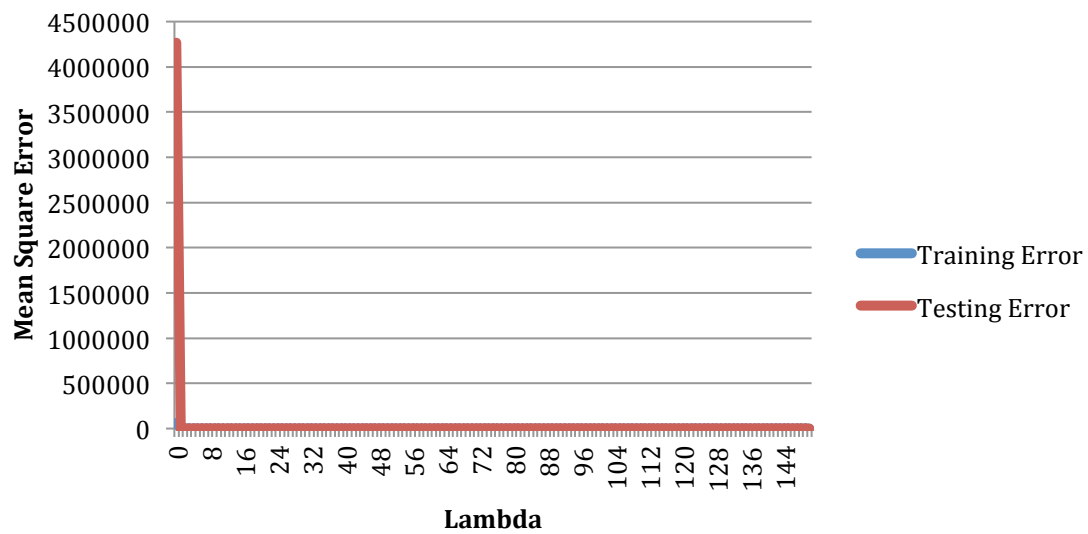
--> The value of lambda increase as the number of examples increase while, number of features remains fixed. Because, weight derived from the less examples will, have high variance to the model, hence as the number of example increase it try to fit as much data possible, this will over fit the data, so to reduce this over fitting, we need to introduce higher penalty to model, which reflect as higher lambda value.

This is observed in graph 50(1000)_100 vs 100(10000)_100 vs 150(1000)_100.

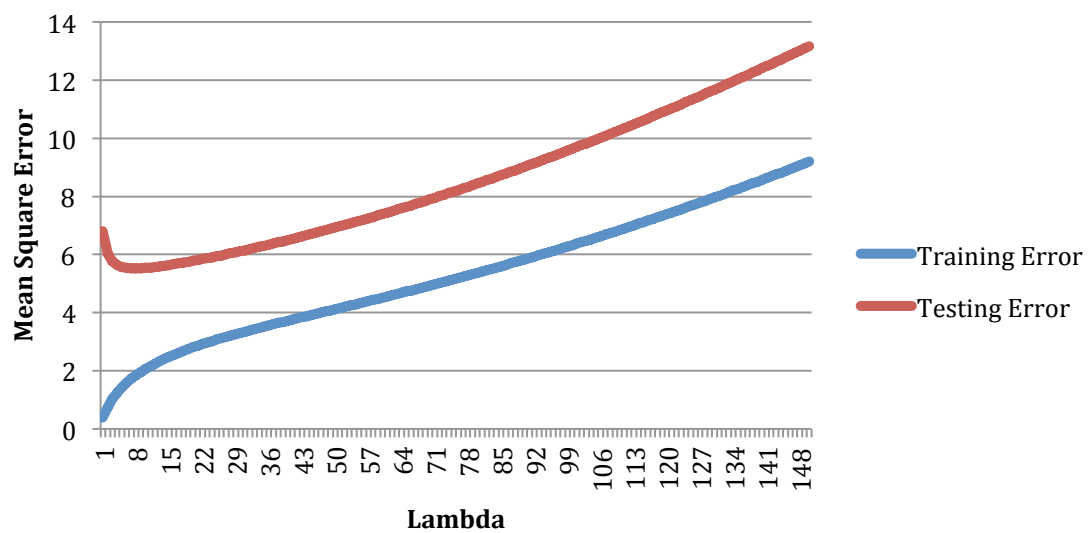




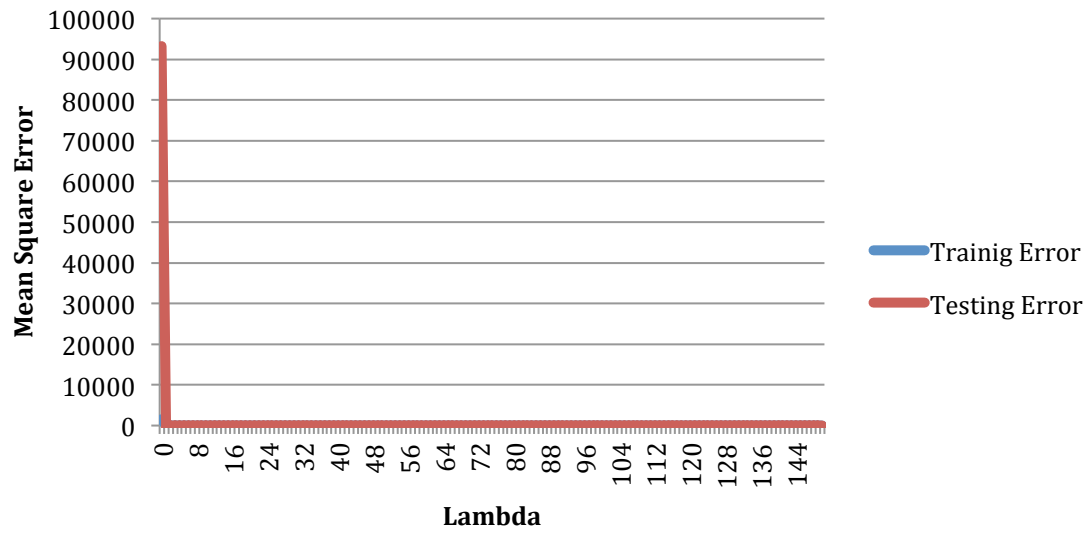
50(1000)_100 Set



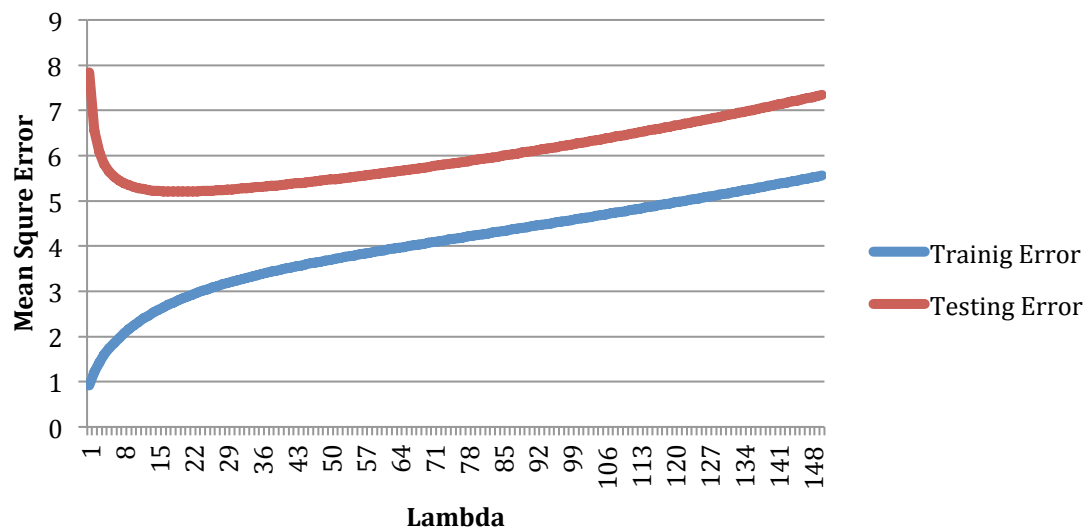
Set 50(1000)_100



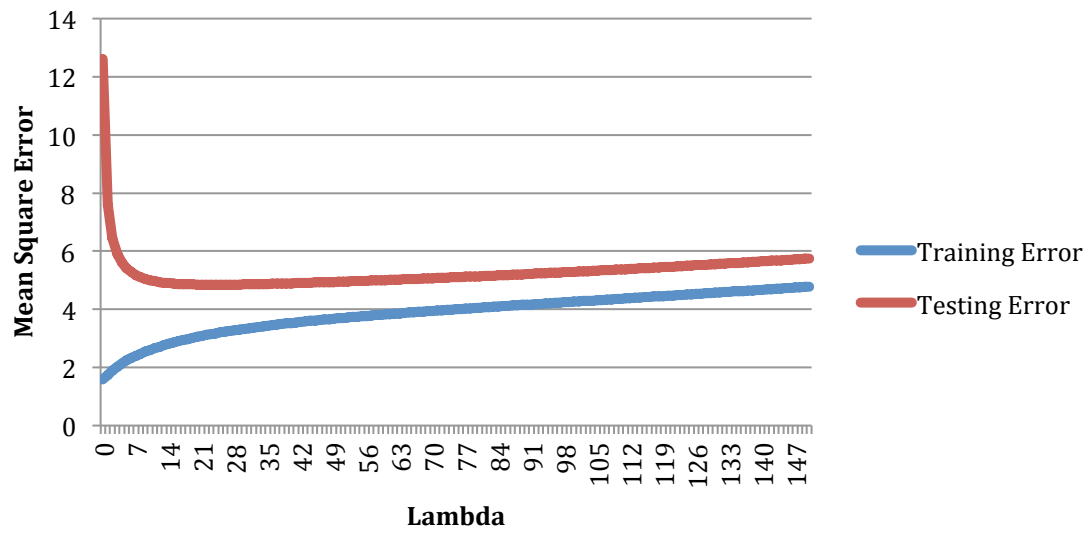
Set 100(1000)_100



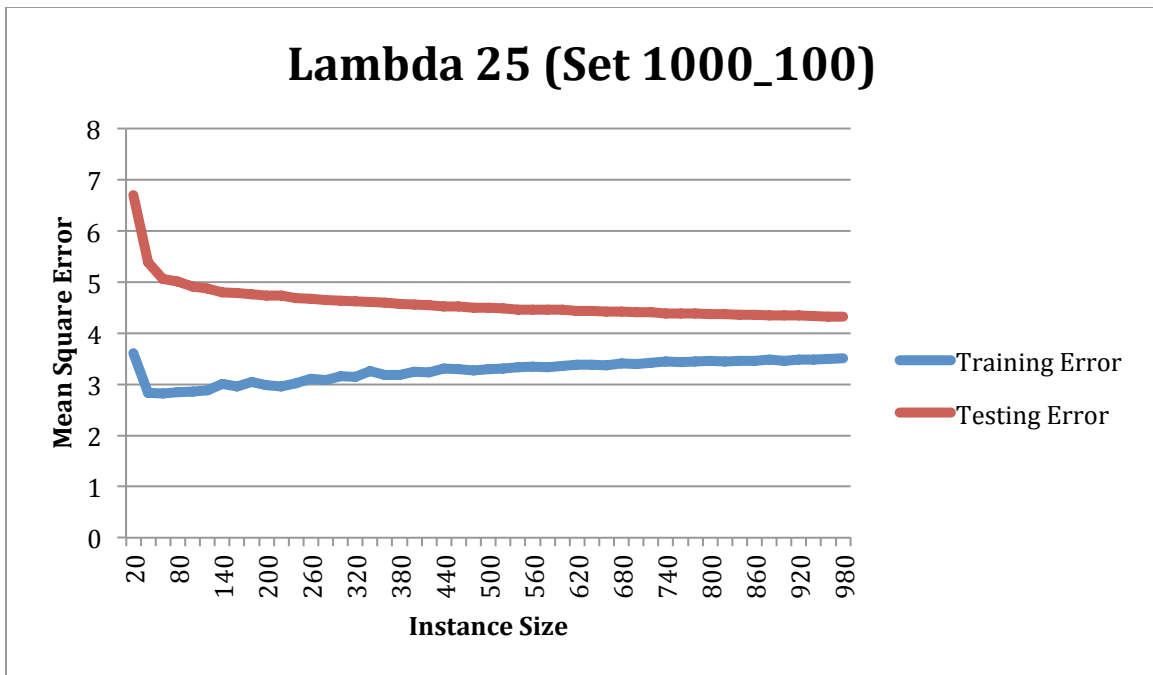
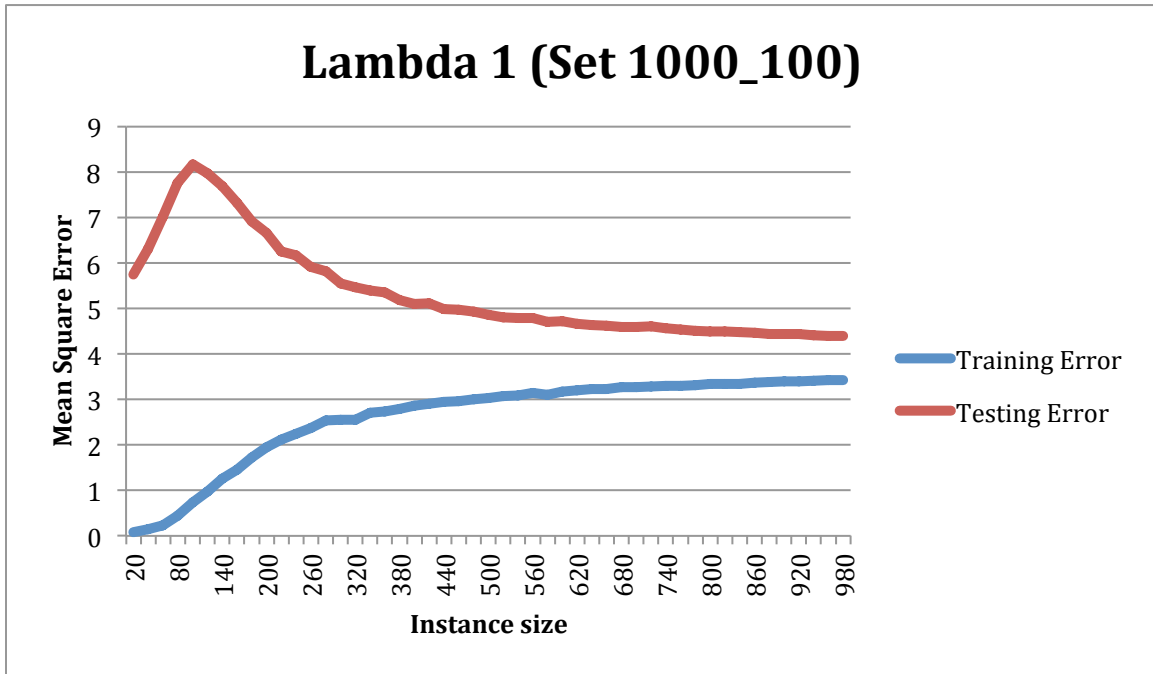
Set 100(1000)_100



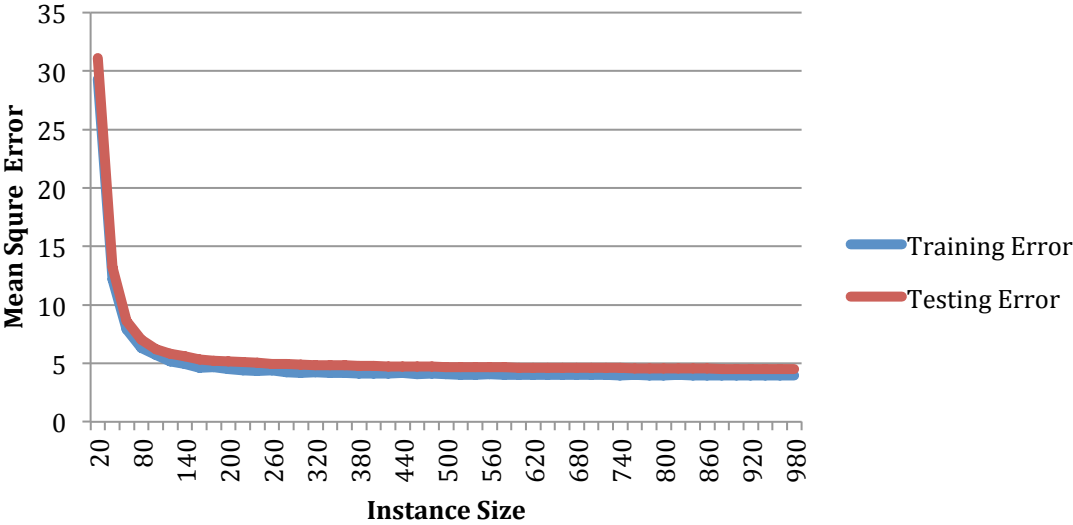
Set 150(1000)_100



Question 2



Lambda 150(Set 1000_100)



Question 3

**** How do the values for λ and MSE obtained from CV compare to the ones in question 1?**

-->

We perform k time training and testing on K folds, for each iteration we test on 1 fold and train on K – 1 fold. Then we find average value of testing error of this fold, now we do this step for each value of lambda and we record the minimum value of testing error and corresponding lambda.

After this using this lambda we train on entire set and determine the Training Error and Testing Error.

While in question 1 we choose the lambda value where the Training Error and Testing Error graph lines is having minimum distance.

Based on this observation the value obtained using question 1 and question 3 are nearly same.

	Question 1: Graph			Question 3 CV		
Set	Train Error	Test Error	Lambda	Train Error	Test Error	Lambda
10_100	3.85	4.19	17	3.81	4.17	13
100_100	2.91	5.09	29	2.58	5.07	20
1000_100	3.59	4.33	47	3.56	4.32	39
50(1000)_100	2.33	5.61	14	3.02	5.90	24
100(1000)_100	2.97	5.21	24	3.24	5.26	31
150(1000)_100	3.47	4.87	38	3.64	4.92	47

**** What are the drawbacks of CV?**

-->

- The main drawback of Cross Validation is we need to rerun K times from scratch.
- The other limitation of Cross Validation is it gives meaningful result, if the test data and training data are derived from the same population.
- Besides this if the system experience changes over time, which results into change of behavior, then predictive model is no longer valid for the given system, so we need to re run cross validation. So cross validation can not responds to changes.

**** What are the factors affecting the performance of CV?**

--> The mainly two factors affecting the performance of Cross Validation.

1. Total is number of folds we choose.
2. Total number of samples in testing vs total number of samples training set on each fold.