	Train		Test	
Data-Set	Least MSE	Lambda (λ)	Least MSE	Lambda (λ)
100-10	0.2594184184 9708835	0	0.09088677117 243264	78
100-100	0.3387710480 377649	1	6.4817037162 75406	6
1000-100	2.1527504747 388497	0	7.58247129115 0516	101
50(1000)-100				
100(1000)-100				
150(1000)-100				
	Train		Test	
Data-Set	Least MSE	Lambda (λ)	Least MSE	Lambda (λ)
100-100	0.3387710480 377649	0	6.4817037162 75406	5
50(1000)-100				
100(1000)-100				

Solution: When lambda is set to zero, regularization is removed completely. So, chances of overfitting increases and is in fact the most because training in this case becomes solely focused on minimizing loss. So the value of MSE becomes huge.

Question- 2
Question- 2c)
Solution: CV, though an effective technic for reducing overfitting has several drawbacks.
CV does not eliminate overfitting completely but reduces its chances of occurring.
CV requires significantly more training than the conventional 3-split method.This problem becomes even more magnified when we consider multiple values of the lambda. Overfitting itself is less likely to occur as the size of the data increases, so CV becomes less effective and relatively unnecessary for a large data size.

Question- 2d)
Solution: The factors affecting the performance of CV are

- Size of data- The bigger the data the less is the need for CV, as the scope of overfitting reduces as data size increases. So it would be counter intuitive to apply the CV technic on data of a large size as the computation time would increase a lot and the benefits are minimal at best.
- Bias- The Validation set approach leads to an overestimate the test errors since in this approach the data gets divided into two parts and only one part of the data is used for building the statistical model. K-fold CV approach uses (k-1).n/k size data to build the model, which is fewer than the LOOCV model. So LOOCV model is preferred over the K-fold CV approach.
- Variance- In the LOOCV technic, the output gets average over n fitted models, which is trained on almost identical size of the data (n-1). So, the outputs for all the n folds are highly correlated with each other. In case of the K-fold technic, the model output is produced, which is less correlated with each other due to the less overlap between the data for each fold. Since, the mean of the highly correlated values has the higher variance than the mean of less correlated values. So, the test errors estimated from LOOCV tends to have the higher variance than the k-fold CV.