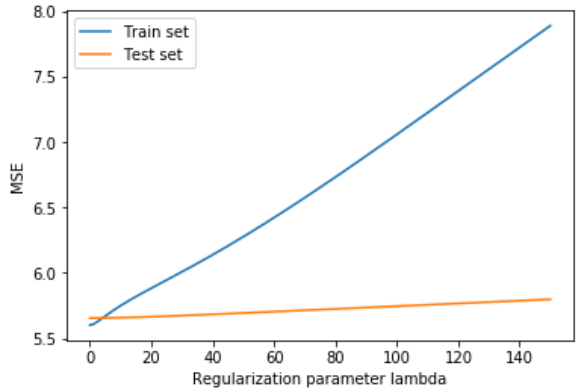
B555: Machine Learning

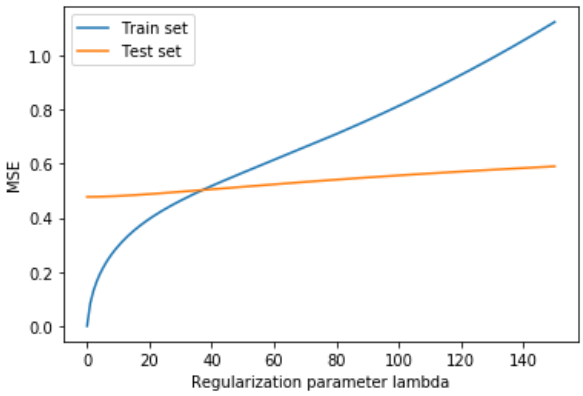
Programming Project 2

Task 1:

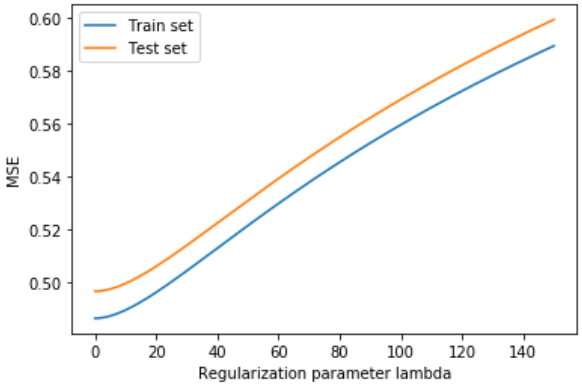
100-10:



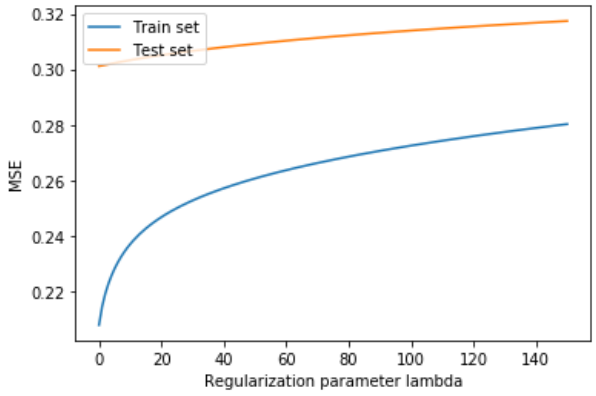
100-100:



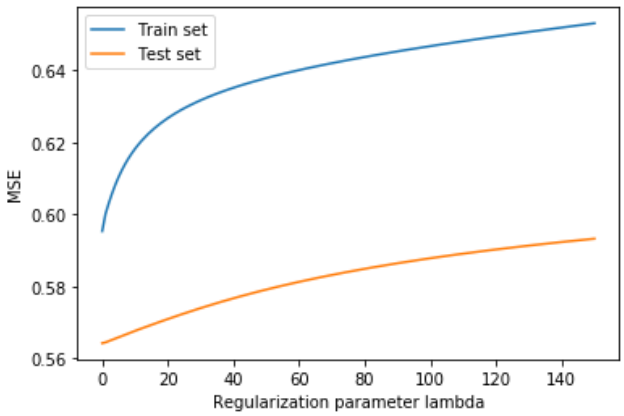
1000-10:



Crime:



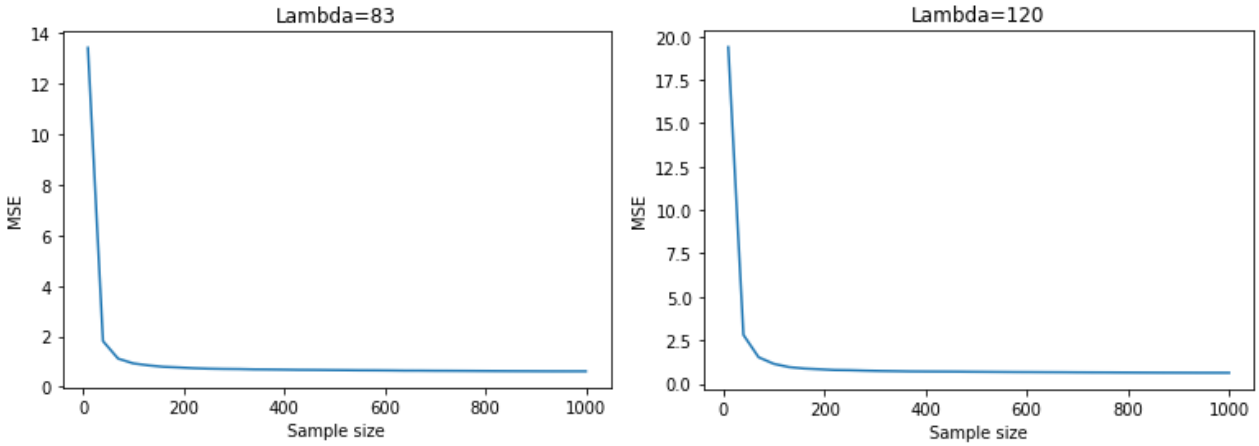
Wine:

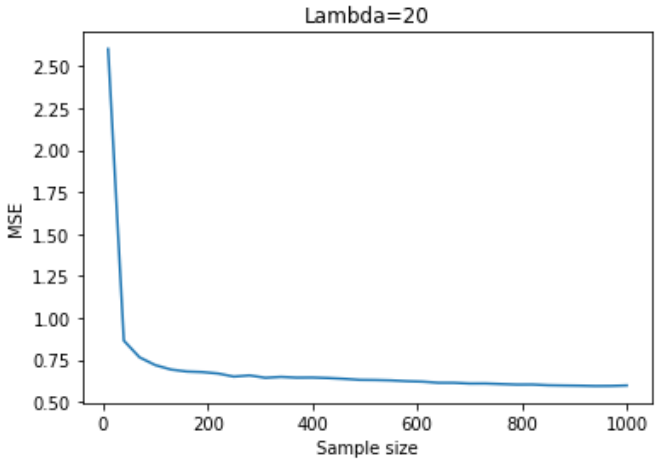


The training set MSEs are very different from the test set MSE. So, they cannot be used to select λ.

For the optimal value of lambda, the MSE is the minimum. For value of lambda less than or more than that value increases the value of MSE. The effect of lambda is different for different datasets based on the number of examples and the number of features.

Task 2:





The error decreases as the sample size increases. The error is minimum for an optimal value of lambda. For values other than the optimal value, the error is higher.

If the training size is less, there is not enough sample to train the algorithm, so the error is more. As the sample size increases, the accuracy increases and thus the error decreases.

Task 3.1:

Dataset	Lambda	MSE	Run Time (in seconds)
Wine	2.0	0.6253	1.75
Crime	149.0	0.39227	10.64
100-10	18.0	6.26639	1.015
100-100	34.0	0.73702	8.906
1000-100	149.0	0.64513	19.48

The results are a bit higher in terms of the test set MSE as compared to task 1. This is because in first task we plotted the lambda and corresponding MSE from the test dataset itself instead of training it first using train set. So, the results are different.

Task 3.2:

Dataset	Alpha	Beta	Lambda	MSE	Run Time (in seconds)
Wine	6.1639	1.6098	3.829	0.6267	0.328
Crime	425.64	3.2504	130.95	0.3911	0.6875
100-10	0.88206	0.16516	5.34068	6.0879	0.0468
100-100	5.1546	3.1543	1.6342	1.0635	0.625
1000-100	10.28579	1.8603	5.52907	0.6083	0.359

The results are a bit higher in terms of the test set MSE as compared to task 1. This is because in first task we plotted the lambda and corresponding MSE from the test dataset itself instead of training it first using train set. So, the results are different. The values of lambda are also very different.

Task 3.3:

The MSE values are almost similar in both the model selection methods although the values of lambda are drastically different. The run time for evidence maximization method is less compared to cross validation method. It also gives less error value in most of the cases.