Part1 :

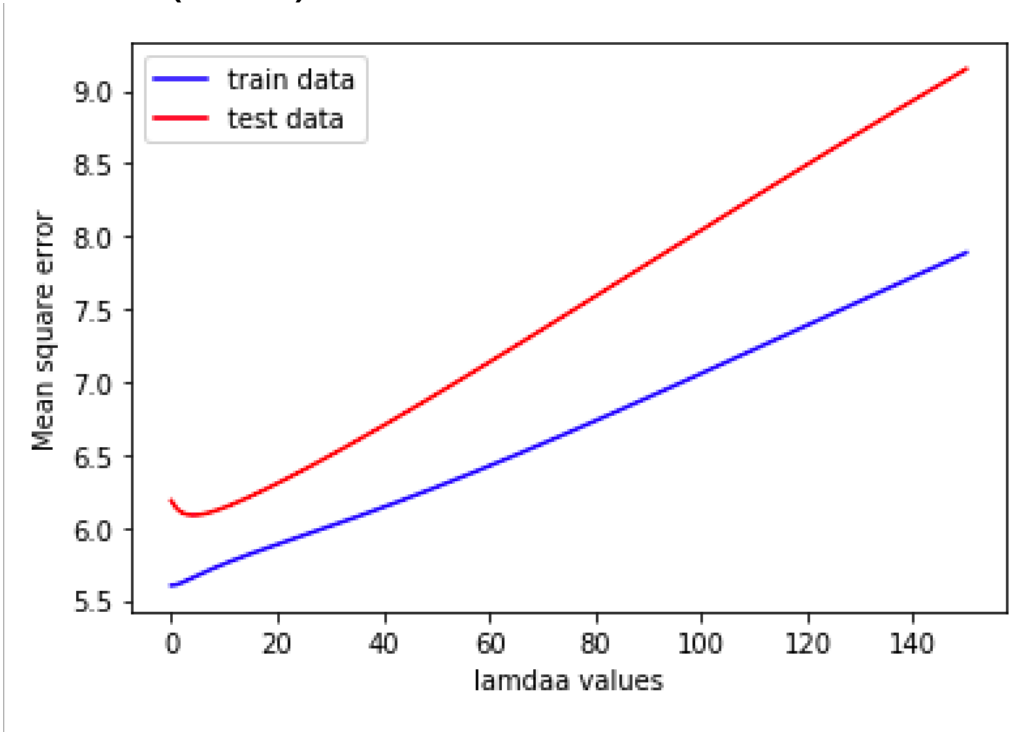
(Input format is : (part1.py) (train-data) (train-data-label) (test-data) (test-data-label))

Procedure :

->Here as mentioned in the question , I have taken the training dataset and test dataset for all the five datasets. It is then trained for lambda values ranging from o to 150 and calculate w for all values of lambda . We calculate train and test use using the w we have calculated and plot the graph

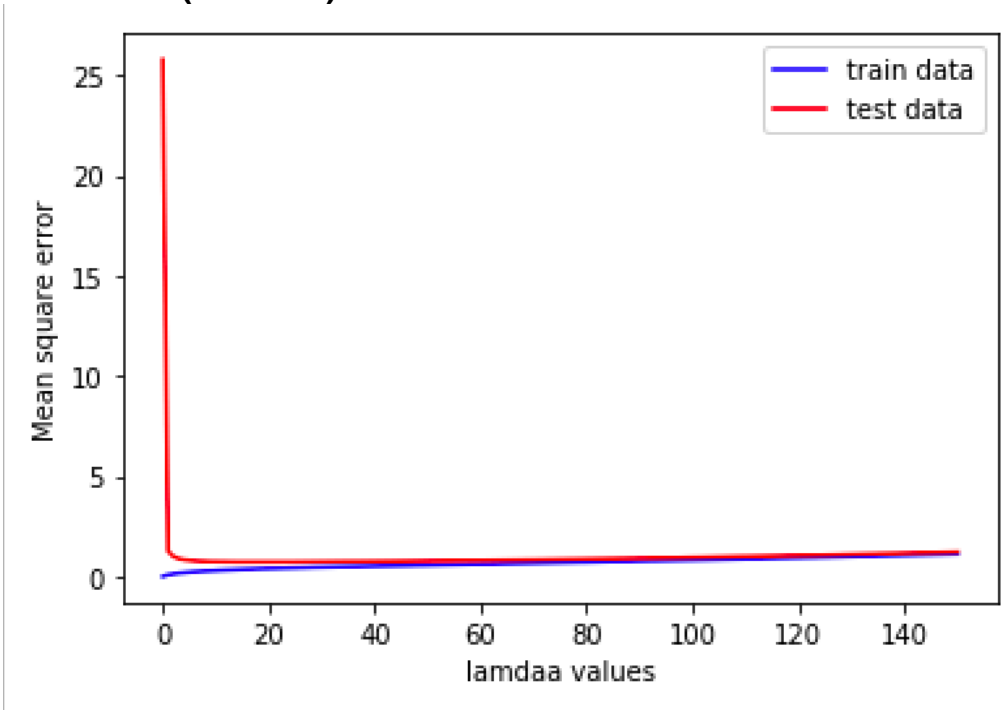
Results :

For data: (100-10)



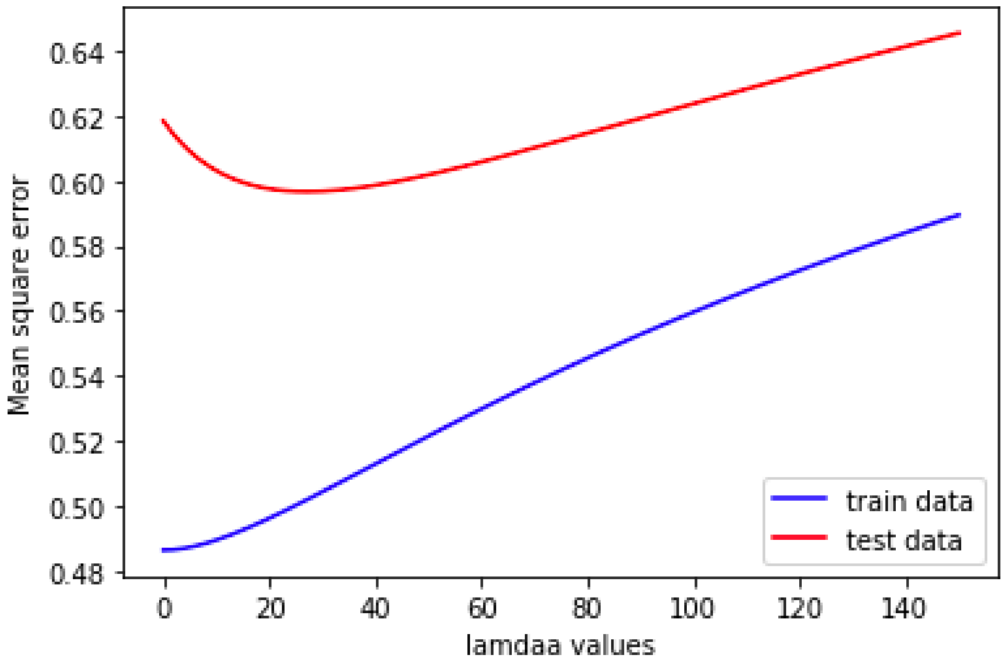
Here as seen in the graph, as the value of lambda increases, the training and testing MSE also increases. However for the range of lambda from 0-20 the the difference between test and train MSE decreases.
For 100-10 MSE is 5.714. Here that MSE is achieved when lambda is around 12(looking from the graph).

For data : (100-100)



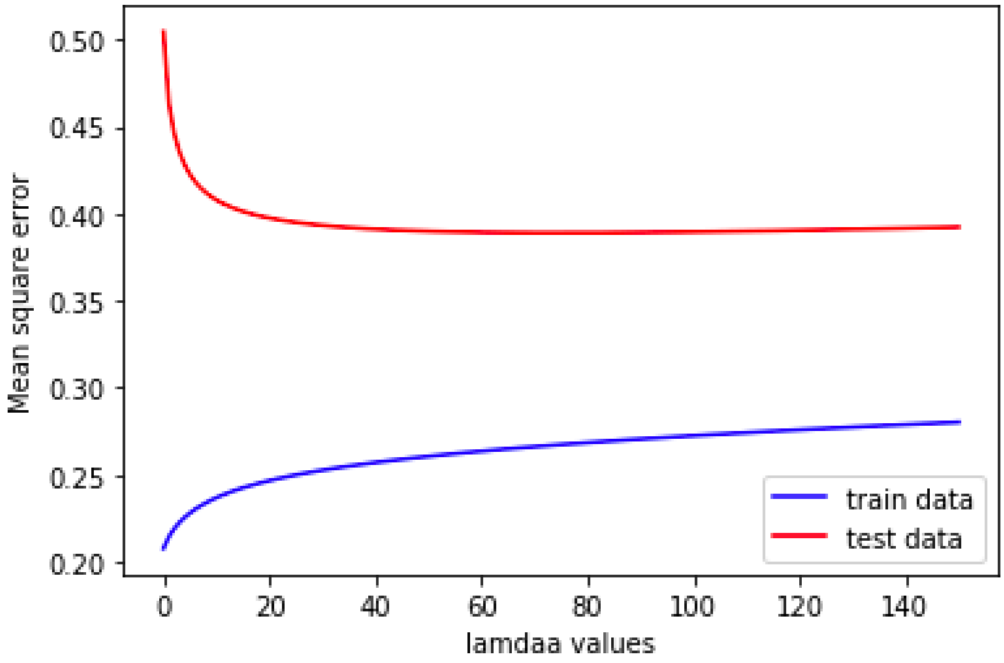
Here, as seen in the graph, for lambda = 0, the test MSE is very bad(extremely high). After that there is a sharp decrease and then test MSE decreases a bit and then again increases. Here for 100-100 dataset MSE given is 0.533 which will be achieved for lambda ranging from 10-15.

For data : (1000-100)



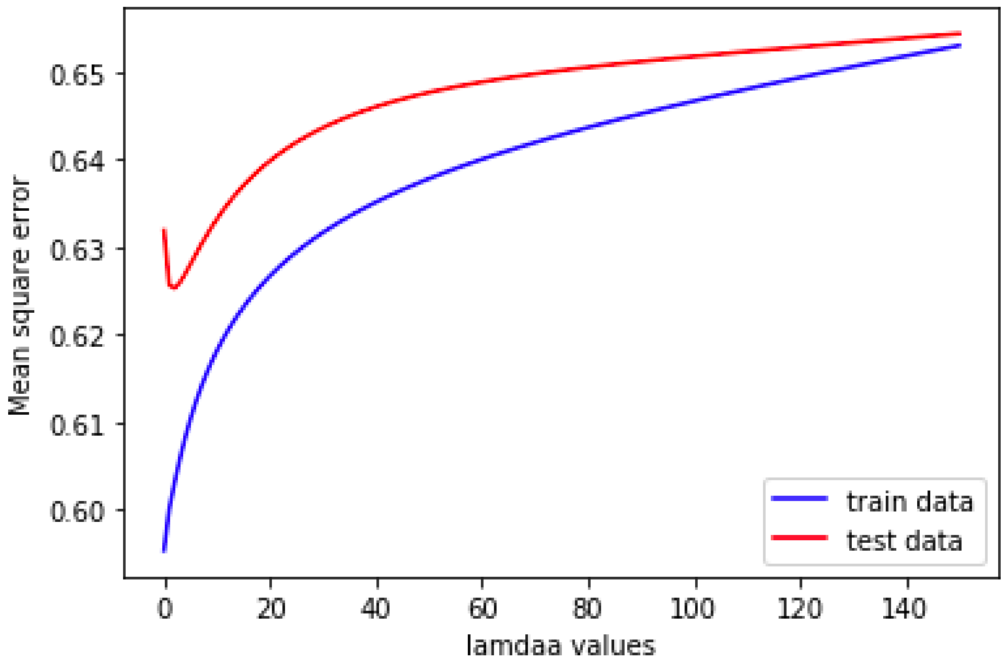
As seen in the graph, Test MSE first decreases for lambda 0 to 30 and then increases . Here the give MSE is 0.557 which will be achieved near lambda 25

For data : (crime)



Here , for the crime dataset , for increasing lambda values the Test-MSE is continuously decreasing . Hence here the best lambda that can be taken is 150

For data: (wine)



Here, for the wine dataset, test MSE decreases for the first few values of lambda and then continuously increases. Hence here the value of lambda should be around 3.

Explanation of such behavior:

for crime data :
Number of examples in train data = 298
Number of examples in test data = 1695
Number of features in every example = 100

for wine data:
Number of examples in train data = 342
Number of examples in test data = 4556
Number of features in every example = 11

for 100-10 data :
Number of examples in train data = 100
Number of examples in test data = 1000
Number of features in every example = 10

for 100-100 data :
Number of examples in train data =100
Number of examples in test data = 1000
Number of features in every example = 100

for 1000-100 data :
Number of examples in train data =1000
Number of examples in test data = 1000
Number of features in every example = 100

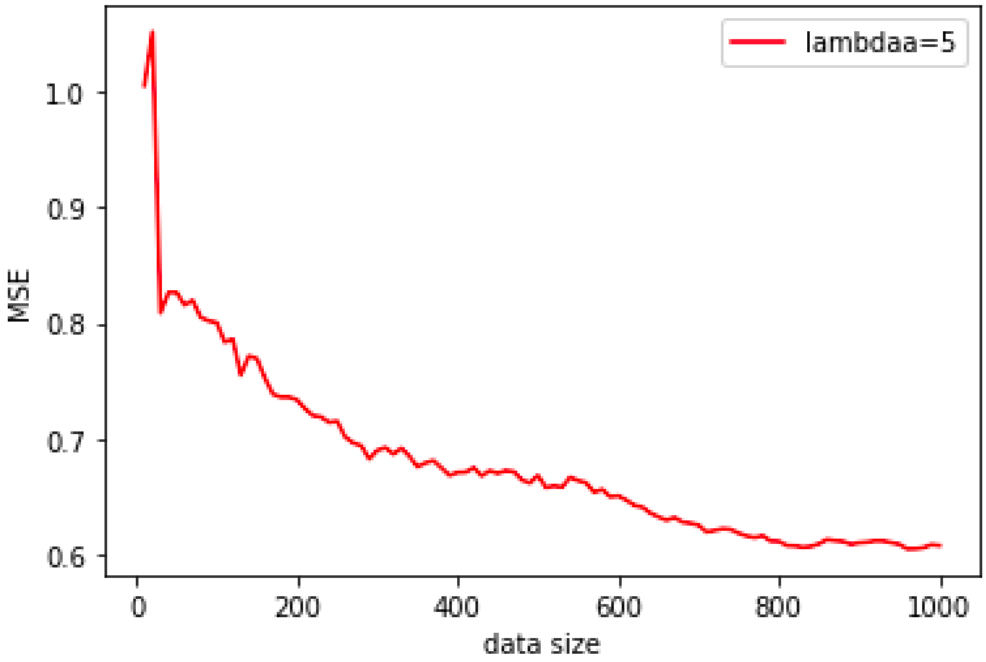
Training MSE cannot be used for selecting the value of lambda as it will be very small as the data used to train, the same data will be used for calculating MSE which will be very optimistic and hence we will get very low value of lambda which is not true.
Lambda tries to minimize the test error if lambda chosen is appropriate as lambda is the regularization parameter which helps us from overfitting the model . Hence we have to select lambda with lowest test MSE . This differs on dataset , number of examples in the training set, number of examples in test and number of features in the dataset. All this taken into account results in selection of different lambda values for different datasets.
Hence from the above graphs we can say that as the number of examples increases train-MSE increases for increasing values of lambda.

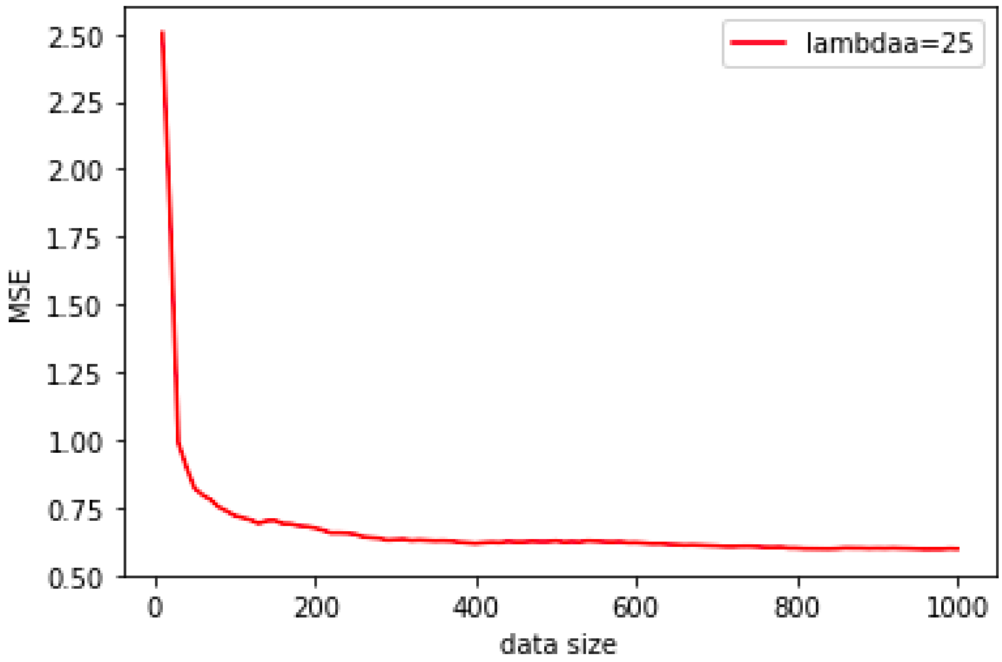
PART-2 :

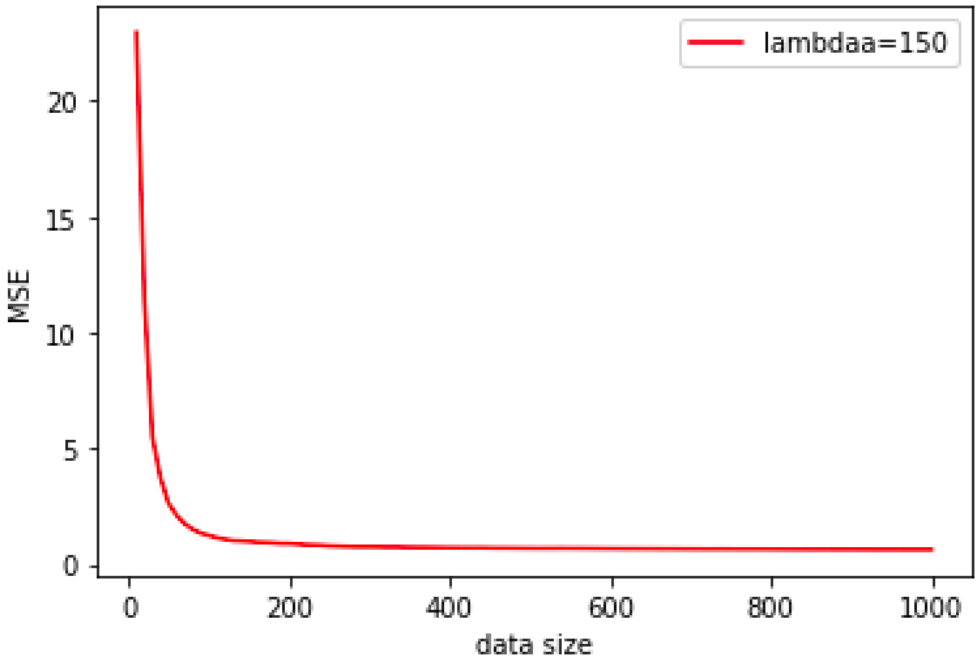
(Input format is : (part2.py)

procedure :
for this example as mentioned in the question we have to take three values of lambda : 1) 5 (too small) 2) 25 (just right), 3)150 (too large).
Here for all three values of lambda , firstly the model is trained and w is calculated. After calculating the W , model is trained using the test data , test label and w for the particular values of lambda.
The above procedure is repeated for varying number of test data size which is randomly selected. i.e.(1-10, 1-20, 1-30, 1-40, …… 1-990 , 1-1000)

Results :


lambdaa=5
1.0
0.9
0.8
0.7
0.6
MSE
0
200
400
600
800
1000
data size


lambdaa=25
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.75
0.50
MSE
0
200
400
600
800
1000
data size



Here from the above plots we can infer that as the value of lambda increases the test MSE also increases but with increase in the data size ,test MSE decreases.
Here for small training test sizes , the data error is the maximum as the model that is trained has very less examples and problem, of overfitting arises as it becomes more specific instead of our model to be generalized hence the MSE is highest with lowest data size and it increases with increase in lambda value as that lambda value will make our model more specific instead of generalizing it . I.e. our model for very less data and high lambda has the highest MSE and it gradually decreases with increase in data set size.

Part - 3.1

Input format: (part3_1.py) (train-data) (train-data-label) (test-data) (test-data-label))

Procedure :
Here, in this example, to get the best value of lambda we implement the algorithm given and calculate the average MSE for each lambda value. The one with lowest MSE we will return its index which is our best lambda.

Now after getting our best lambda we will train our model and get the value of w and test our data and calculate test MSE.

Algorithm for finding lambda value :

To select parameter a of algorithm A(a) over an enumerated range a $\in$ $V_1, \ldots, V_K$ using dataset D we do the following:

1. Split the data D into 10 disjoint portions. 2. For each value of a in $V_1,\ldots,V_K$:

(a) For each i in 1...10
i. Train A(a) on all portions but i and test on i recording the error on portion i

(b) Record the average performance of a on the 10 folds. 3. Pick the value of a with the best average performance.

Now, in the above, D only includes the training set and the parameter is chosen without knowledge of the test data. We then retrain on the entire train set D using the chosen value and evaluate the result on the test set.

Results :

For data 100-10 :

```
In [853]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_1.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
best value of lambdaa is  15
test MSE for the following lambda is  6.214438800288887
running time is 0.35221419599838555
```

For part-1:
lambda is : 13
MSE is 6.1

For data 100-100 :

```
In [823]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_1.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
best value of lambdaa is  18
test MSE for the following lambda is  0.720278805652723
running time is 0.6190462719823699
```

For part-1:

lambda is : 15
MSE is 0.65

For data 1000-100

```
In [825]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_1.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
best value of lambdaa is  23
test MSE for the following lambda is  0.5970023803034488
running time is 2.3543972820043564
```

For part-1:
lambda is : 25
MSE is 0.585

For wine

```
In [826]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_1.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
best value of lambdaa is  2
test MSE for the following lambda is  0.6253088423047022
running time is 0.704653031018097
```

For part-1 :
lambda is : 2
MSE is 0.625

For Crime

```
In [827]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_1.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
best value of lambdaa is  150
test MSE for the following lambda is  0.3923389920343811
running time is 0.9523139880038798
```

For part-1:
lambda is : 150
MSE is 0.39

As we can see from the comparison we can tell that lambda and MSE are very much similar to that obtained in part 1 .

PART-3_2:

Input format: (part3_2.py) (train-data-for-1000-100) (train-data-label-for-1000-100) (test-data-for-1000-100) (test-data-label-for-1000-100)

Procedure :

Here , we calculate , Sn, Mn, gamma , Alpha and beta and iterate it and update it for 100 iterations . Before starting with iterations we randomly assign alpha and beta in the range of 1-10. We calculate lambda by dividing alpha by beta . And for that particular value of lambda we train our model and calculate test MSE.

Results :

For data 100-10 :

```
In [828]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_2.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
here the lambda is   5.340684382677252
test MSE for the following lambda is  6.08795694201548
running time is 0.04510176400071941
```

For part-1:
lambda is : 13
MSE is 6.1

For data 100-100 :

```
In [830]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_2.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
here the lambda is   1.6341681242830735
test MSE for the following lambda is  1.0635085541857139
running time is 0.3720771509979386
```

For part-1:
lambda is : 15
MSE is 0.65

For data 1000-100 :

```
In [831]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_2.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
here the lambda is   5.529076370320971
test MSE for the following lambda is  0.6083085922060236
running time is 0.6423246850026771
```

For part-1:
lambda is : 25
MSE is 0.585

For data crime :

```
In [832]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_2.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
here the lambda is   130.9503952612452
test MSE for the following lambda is  0.39110230747343366
running time is 0.44379750001826324
```

For part-1:
lambda is : 150
MSE is 0.39

For data wine :

```
In [833]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_2.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
here the lambda is   3.8290133914778743
test MSE for the following lambda is  0.626746238570825
running time is 0.14680013799807057
```

For part-1 :
lambda is : 2
MSE is 0.625

Here , the results obtained are different for many cases to that obtained in part1.

Part 3_3 : Comparison

Comparison for lambda , test-MSE and running time .

->For data 100-10 :

-According to cross validation

```
In [853]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_1.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
best value of lambdaa is  15
test MSE for the following lambda is  6.214438800288887
running time is 0.35221419599838555
```

-According to evidence maximization :

```
In [828]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_2.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
here the lambda is   5.340684382677252
test MSE for the following lambda is  6.08795694201548
running time is 0.04510176400071941
```

->For data 100-100 :

-According to cross validation

```
In [823]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_1.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
best value of lambdaa is  18
test MSE for the following lambda is  0.720278805652723
running time is 0.6190462719823699
```

-According to evidence maximization :

```
In [830]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_2.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
here the lambda is   1.6341681242830735
test MSE for the following lambda is  1.0635085541857139
running time is 0.3720771509979386
```

->For data 1000 -100 :

-According to cross validation

```
In [825]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_1.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
best value of lambdaa is  23
test MSE for the following lambda is  0.5970023803034488
running time is 2.3543972820043564
```

-According to evidence maximization :

```
In [831]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_2.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
here the lambda is   5.529076370320971
test MSE for the following lambda is  0.6083085922060236
running time is 0.6423246850026771
```

->For data crime :

-According to cross validation

```
In [827]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_1.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
best value of lambdaa is  150
test MSE for the following lambda is  0.3923389920343811
running time is 0.9523139880038798
```

-According to evidence maximization :

```
In [832]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_2.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
here the lambda is   130.9503952612452
test MSE for the following lambda is  0.39110230747343366
running time is 0.44379750001826324
```

->For data wine :

-According to cross validation

```
In [826]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_1.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
best value of lambdaa is  2
test MSE for the following lambda is  0.6253088423047022
running time is 0.704653031018097
```

-According to evidence maximization :

```
In [833]: runfile('/Users/vanshsmacpro/Desktop/sem1/pp2data/part3_2.py', wdir='/Users/
vanshsmacpro/Desktop/sem1/pp2data')
here the lambda is   3.8290133914778743
test MSE for the following lambda is  0.626746238570825
running time is 0.14680013799807057
```

Here , from the data above the running time of evidence maximization is significantly less compared to cross validation for all cases.
The test MSE for cross validation is greater then test MSE for evidence approximation for : wine-data , 1000-100 , 100-100 : so in this case we will chose cross validation method
The test MSE for cross validation is less then test MSE for evidence approximation for : crime-data , 100-10 : so in this case we will chose evidence maximization method.