Programming assignment 2

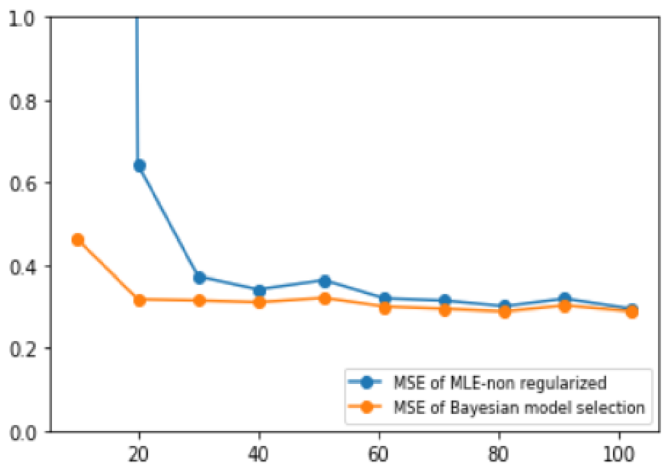
Task1:

1. Using the Model selection algorithm, Values of alpha, beta, and effective lambda for each training size are calculated and displayed as shown below,

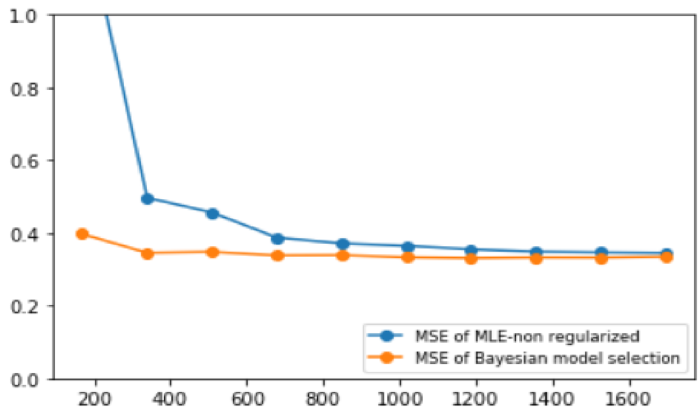
```
1st question Part 1
-----
Housing data
Bayesian model selection parameters value for various train size
train size    alpha         beta          eff-lambda
10            19.68         49.64          0.4
20            16.99         3.62          4.69
30            17.1         4.3          3.97
40            18.41         5.13          3.59
51            17.81         4.41          4.04
61            19.53         4.79          4.08
71            18.62         4.47          4.17
81            18.93         4.59          4.13
91            16.98         4.17          4.07
102            20.41         4.04          5.05


Crime data
Bayesian model selection parameters value for various train size
train size    alpha         beta          eff-lambda
169           188.66         3.08          61.29
339           284.72         2.96          96.05
508           267.21         2.85          93.7
678           280.44         2.85          98.41
847           282.76         2.91          97.01
1017           263.47         2.96          88.94
1186           252.36         3.09          81.77
1356           254.2         3.12          81.36
1525           247.88         3.05          81.35
1695           239.71         3.09          77.69
-----
```

2. MSE for a non-regularized model of MLE and Bayesian model selection with different training sizes,

```
Housing data
1st question part 2 - MSE for non regularization MLE and bayesian model selection
```



```
crime data
1st question part 2 - MSE for non regularization MLE and bayesian model selection
```



Observation based on both plots:

When the training data size is one-tenth of the total training size the MSE is infinity. But as training size increases, the MSE came down very quickly and then converges with the MSE of Bayesian model selection. At the very low size of training data, comparatively, the Bayesian model performed well.

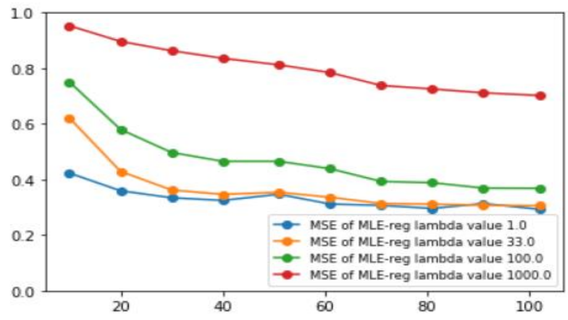
What can you observe w.r.t. their relative performance? Try to explain why the results are as observed and whether they are as expected or not

With a very less training size, the effectiveness of the Maximum likelihood model without any regularization is extremely low compared to the Bayesian model selection. But as the training data size increases, both models result in similar Mean square error, explaining that the accuracy of both models converges as the training data size increases and also the overall accuracy increases as the training data size increases.

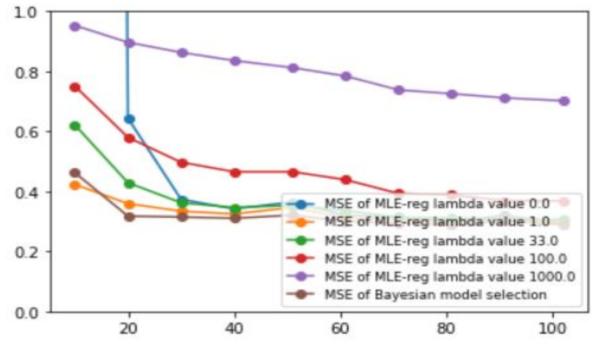
With limited training size, the MLE model would have the overfitting problem, but the same problem is overcome by the Bayesian model with prior knowledge in terms of quantitative beliefs, which helps the model to predict the response data more accurately than the MLE model. With training data size increases, the model developed by both MLE and Bayesian becomes more generalized, so it became effective in predicting the response values, thus the overall MSE reduces. So, the observed results are in the expected lines.

3.

```
Housing data
1st question part 3 - MSE for regularization with values, 1,33,100,1000 and bayesian model selection
```



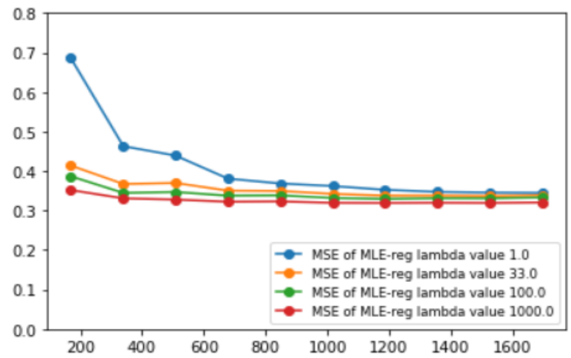
```
Housing data
MSE plot for all the calculated lambda values of MLE and also for MAP of bayesian model selection
```



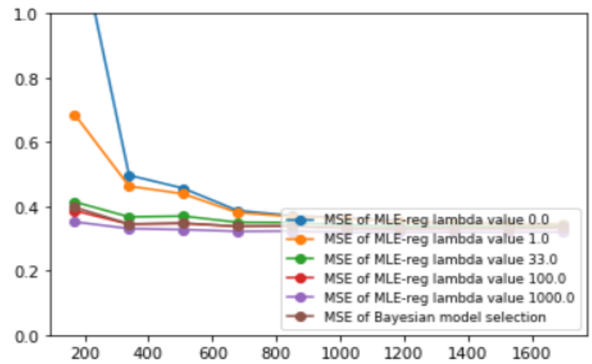
Observation on housing data plots:

For the housing data, the most effective lambda is 1.0. The plot of MLE with lambda equal to 1 has a very similar MSE as the Bayesian model. This explains that the Bayesian model gave comparable or as effective results as regularized MLE model. But as the training data increases, irrespective of lambda values, all the MSE curves converge explaining that the effect of lambda in the large data set is minimal.

```
crime data
1st question part 3 - MSE for regularization with values, 1,33,100,1000 and bayesian model selection
```




```
crime data
MSE plot for all the calculated lambda values of MLE and also for MAP of bayesian model selection
```



Observation on crime data plots:

For crime data, the most effective lambda is 1000 but its effectiveness is just marginally high than the lambda of 100 (which is one-tenth of 1000). But as explained earlier, as the training data size increases, the effect of lambda in the final result seems to be vanishing as irrespective of lambda values, all the models have similar MSE for training data of large size.

Can we use a single universal value for λ for different datasets? Is the Bayesian algorithm successful in selecting a good value? How could one select it otherwise?

By comparing the MSE plots of various lambda of both training data, we can conclude that the same lambda values are not suitable for different data sets. As the best lambda value for housing, data is 1.0, but the same is not effective for crime data. For the crime data, the effective lambda is 1000 and the next best is 100.

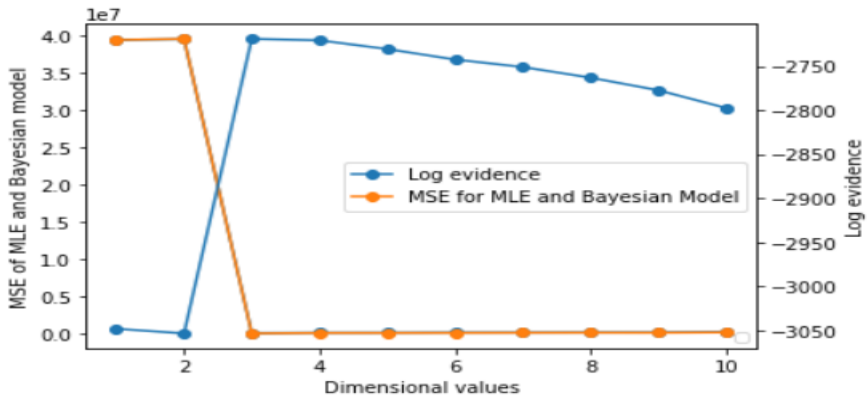
The Bayesian algorithm determined the lambda value for housing data of various sizes is in the range of 4-5 mostly. From the graph, we understand that the best lambda is less than 10 and could be near 1. So for housing data, the Bayesian model gave comparably effective lambda but we can't say that is the best option. Similarly, for the `crime data, the lambda of 100 and 1000 are both greatly reduced with mse, and the MSE of the model with lambda 1000 marginally has higher accuracy. But the Bayesian model predicted the lambda in the range of 80-100.

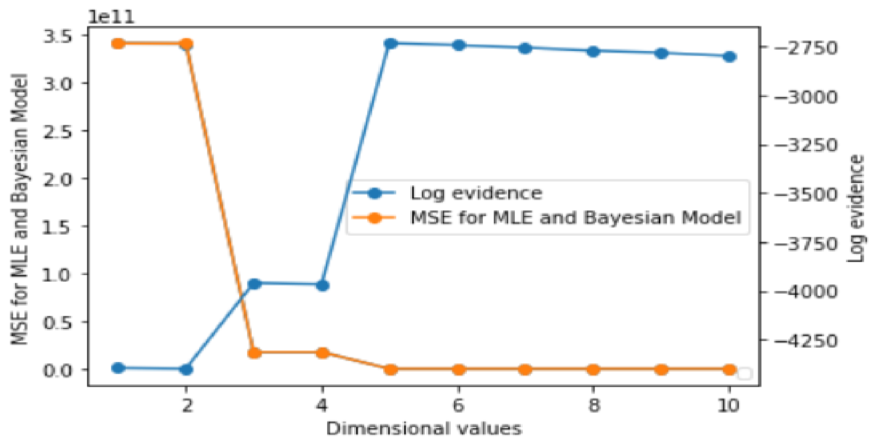
So, overall, the Bayesian model greatly helped in pushing the model towards the nearly good lambda value, but may not be the best one.

The other method could be to use the evidence function or EVIDENCE APPROXIMATION, in which we maximize the function and directly determine the alpha, and beta values and use that for determining the final model to predict the response values.

Task2:

PLOTS for F3 data
MSE of MLE and Bayesian model along with log evidence





Observation on both f3 and f5 plots:

When the data is used with different dimensional values, the evidence function maximizes for the specific dimensional values with corresponding alpha and beta. At the same point, the mse values also drop to the nearly minimum value. So, the log evidence function which is determined solely based on training data would help in optimizing the model to determine the parameters of the model which helps in predicting with greater accuracy.

Can the evidence be used to successfully select α, β and d for the Bayesian method? How does the non-regularized model fare in these runs?

Based on the observation of the plots, we can say that the evidence function helps in determining the optimum dimensionality and hence the alpha and beta corresponding to that where the MSE of the model is least. So, the log evidence function is indeed successful in determining the parameters for the Bayesian model.

The non-regularized model also gave similar accuracy or similar MSE values compared to the Bayesian model. But the Bayesian model gave a marginally better MSE score than the non-regularized model. The MSE of value for different dimensional values is shown below.

F3 Data:

MSE of non-regularized model:

```
{1: 39389142.58553825, 2: 39495762.45900569, 3: 148429.3883649426, 4: 1
79627.46131849606, 5: 186263.56079639218, 6: 211370.6073940488, 7: 2110
30.05221867646, 8: 219229.879478664, 9: 226345.88116710776, 10: 270774.
8068386121}
```

MSE of Bayesian model:

```
{1: 39376476.11601517, 2: 39527146.41872832, 3: 96455.98109742106, 4: 1
54882.81449527247, 5: 157992.413764386, 6: 165209.15501289774, 7: 18289
7.76545493054, 8: 194734.91580312623, 9: 186392.13599227005, 10: 231746
.58034261907}
```

F5 data:

MSE of non-regularized model:

```
{1: 341642661126.0123, 2: 341195638304.2563, 3: 17465602121.93236, 4: 1
7435655143.623474, 5: 61375.345065543755, 6: 79043.0543017258, 7: 10603
1.55252956954, 8: 104840.41212287817, 9: 110065.56808059901, 10: 108832
.5524214696}
```

MSE of Bayesian model:

```
{1: 341402522422.93665, 2: 341325174572.45123, 3: 17502036953.65271, 4:
17478865442.561092, 5: 54311.85449418849, 6: 69754.35175814199, 7: 9076
8.26886716852, 8: 90681.85947563125, 9: 104393.1057899487, 10: 100850.6
1705746669}
```