Nathan Brahmstadt
Jordan Crane
CS 434
Assignment 1

1. This is done in the source code in the function `get_data_arrays()`.
2. Calculated weight vector **with dummy variable**:

```
[ 39.58432122  -0.10113705   0.04589353  -0.00273039   3.0720134
-17.22540718   3.71125235   0.00715862  -1.5990021    0.37362337
-0.01575642  -1.02417703   0.00969321  -0.58596927]
```

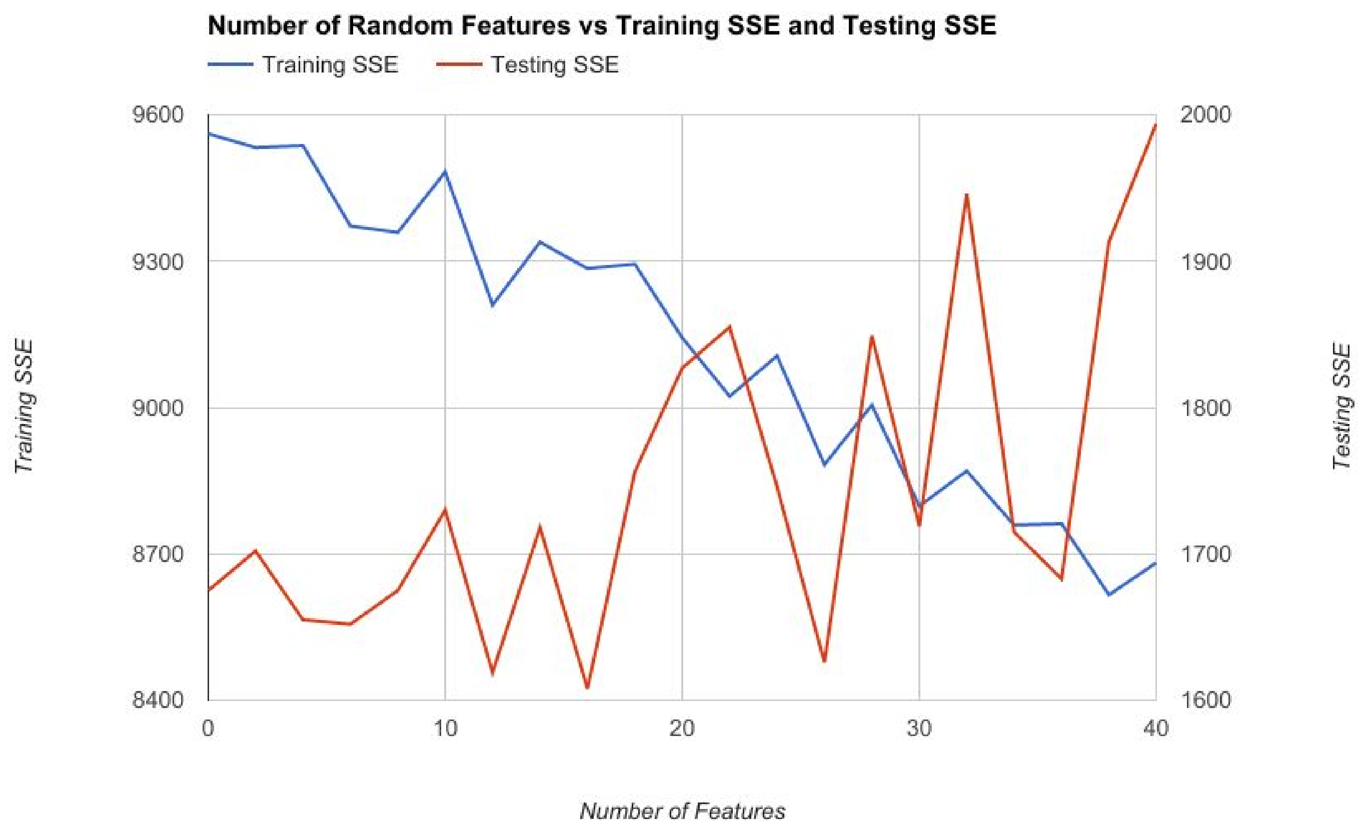
   Note: The first value (`39.58`) is the constant, and the following **w**'s correspond to the 13 features

3. SSE values **with dummy variable**:
   Training Data SSE: `9561.191`
   Test Data SSE: `1675.231`

4. SSE values **without dummy variable**:
   Training Data SSE: `10598.057`
   Test Data SSE: `1797.626`

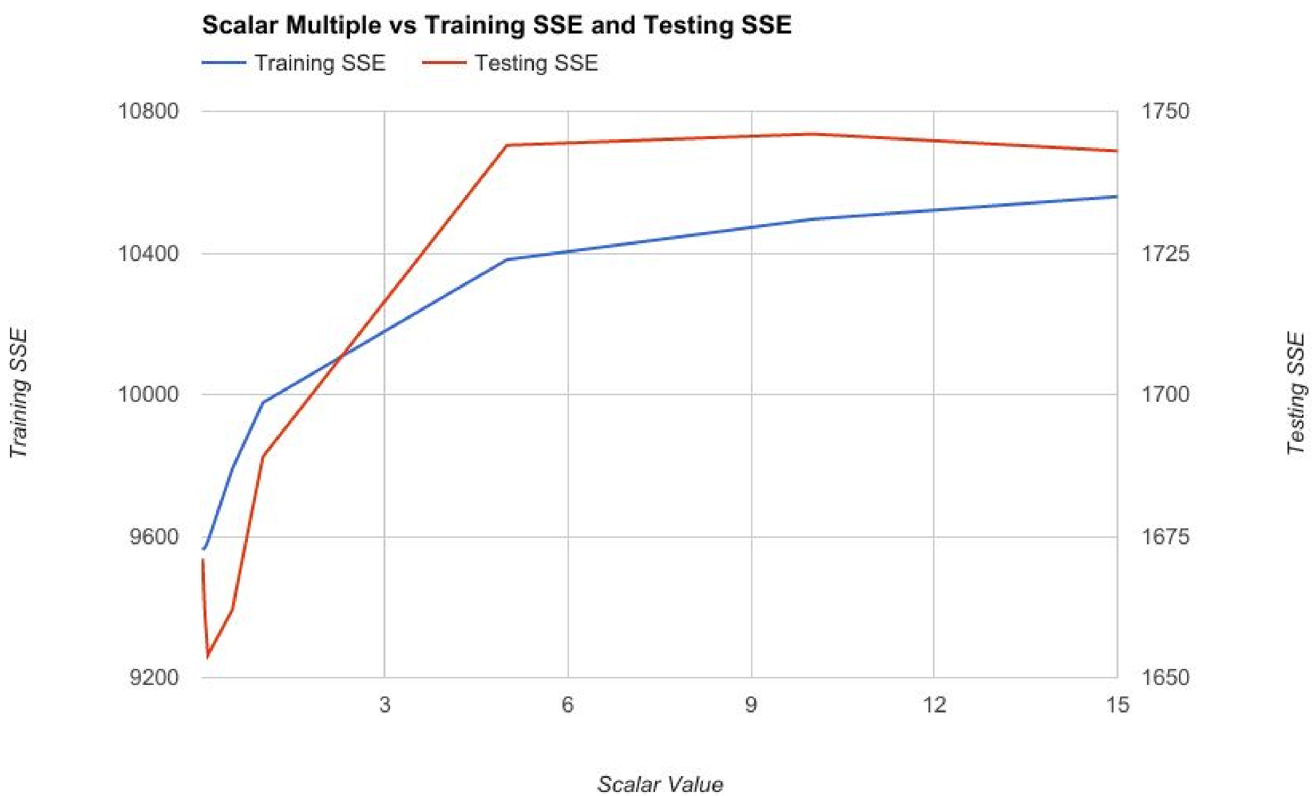
   Both SSE values were larger when removing the dummy variable, which means that the model was more inaccurate.

5. Using 123 as the random seed, the results of adding random features to the data are as follows:
   Note: The dummy variable was used for all sets of data



Going up to 10 added features, there wasn't really a clear trend depending on the randomization, but as we increased the number of random features, trends became clear. We see training SSE decrease and testing SSE increase. With the error increasing on the test data as the number of features increases, we can conclude that introducing more uncorrelated features does not lead to better performance.

6. Here are the results of adding a scalar multiple of the identity matrix to the first term of the weight vector equation:
   Note: The dummy variable was used for all sets of data



   It is apparent from the graph that increasing the scalar value tends to increase both SSE's. This suggests that for this particular problem, the best λ is as small as possible (0 is ideal).

7. As the value of λ increases, the values within the **w** vector decrease.

8. Since λ represents the regularization factor, it is logical that increasing its value would also increase the SSE, especially that of the testing SSE as shown in the graph. Since increasing λ serves to mitigate overfitting, it reduces the fitting of the model to the data, thus resulting in less weight being given to the fitting portion of the optimization equation and increasing the SSE.

   If we were dealing with a situation in which overfitting might be a problem, such as fitting a more complex polynomial model, we would expect to see the training SSE increase, but the testing SSE decrease at the optimal value of λ.

However, since we are using a linear model for this example, overfitting is not an issue, and increasing the amount of regularization just causes the fit of the model to be less accurate, thus increasing the testing SSE as well.