Task 1

Inside the split() function I take the dataframe generated from the csv file and from it I create two new dataframes one using only the feature columns, (Task 1.1, Line 12), and one using the target columns (Task 1.2, Line 13). I then further extract the heating target from the cooling target in order to use .max() and .min() to find the minimum and maximum heating loads in each dataframe. (Task 1.3, Lines 15→26).

Task 2

I created a function to calculate the estimated target vector (Task 2.1, Lines 32→55).

I created a function to find the correct size for the parameter vector given a degree. (Task 2.2, Lines 51→64)

Task 3

Task 3.1 – See comments on lines 69 & 74

Task 3.2 – See comments on line 76

Task 4

Task 4.1 – See comments on line 84

Task 4.2 – See comments on line 85

Task 5

Task 5.1 – See comments on lines 94 & 98

Task 5.2 – See comments on line 99

Task 5.3 – See comments on line 100

Task 6

In modelSelection() I setup two 7different cross validation procedures, one for heating and one for cooling. (Task 6.1, Lines 108 & 109) I calculated the difference between the prediction and test data set. (Task 6.2, Lines 128 & 144) I then append this difference to a list in order to calculate and print the mean difference across all folds for this particular degree (Task 6.3, Lines 129→132 & Lines 145→148) I loop this for each degree from 0 to 2, appending these mean split results to a further list, which I then get the minimum of the find the best performing degree. (Task 6.4, Lines 111→154)

Task 7

In evaluate() I estimated the model parameters using the function from task 5 for both the heating and cooling loads, passing in the best performing degree respectively. (Task 7.1, Lines 161 & 162) I then calculated the estimated loads for both heating and cooling using the entire dataset. (Task 7.2, Lines 164 & 165) Next I plotted these estimates vs the actual true loads. (Task 7.3, Lines 167→180) Finally, I calculated and printed the absolute difference between the prediction loads using the best degree and the actual loads, and then found the mean of that to get the mean absolute difference between the two. (Task 7.4, Lines 182→186)