Anthony Tummillo

Homework Report #4

Part 1.

(a)

There is one binary attribute and that is attribute 4, Charles River dummy variable

(b)

Correltions:

Attribute 1 and 14: -0.3883
Attribute 2 and 14: 0.3604
Attribute 3 and 14: -0.4837
Attribute 4 and 14: 0.1753
Attribute 5 and 14: -0.4273
Attribute 6 and 14: 0.6954
Attribute 7 and 14: -0.3770
Attribute 8 and 14: 0.2499
Attribute 9 and 14: -0.3816
Attribute 10 and 14: -0.4685
Attribute 11 and 14: -0.5078
Attribute 12 and 14: 0.3335
Attribute 13 and 14: -0.7377

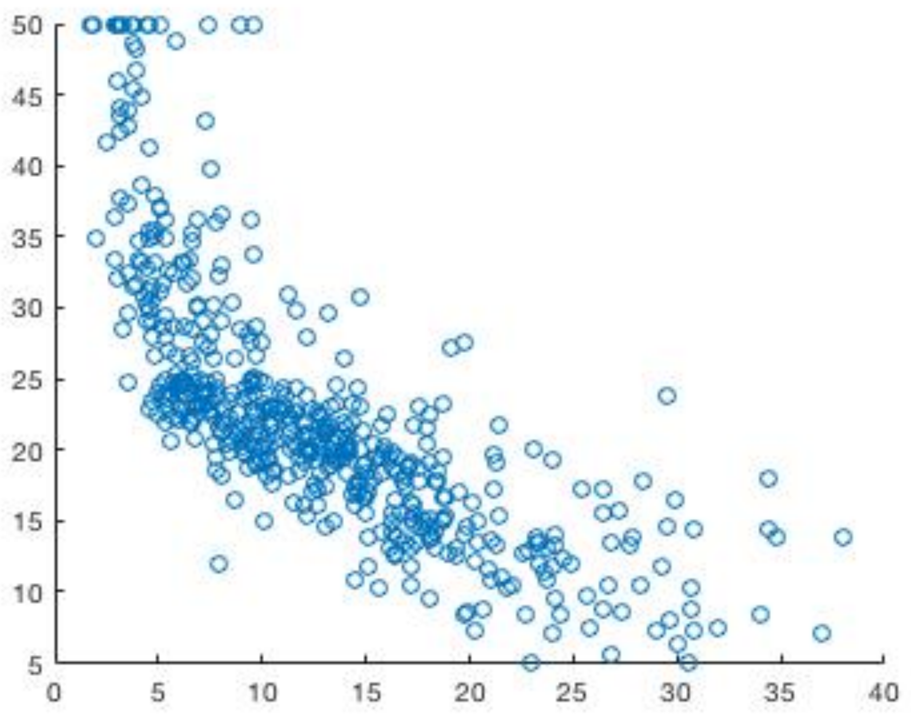
Attribute 6, average number of rooms per dwelling, has the highest positive correlation.

Attribute 13, % lower status of the population, has the highest negative correlation

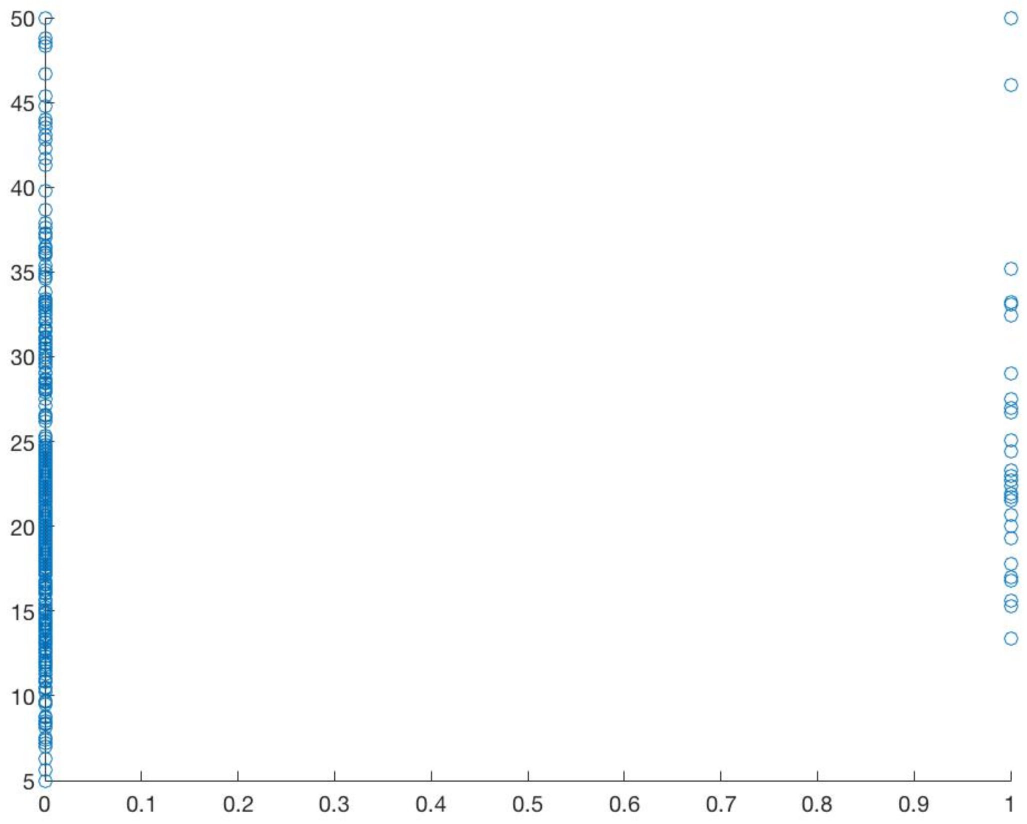
(c)

The plot of attribute 13 against the target attribute looks the most linear. This is because the plot resembles a straight line (without having too many outliers that don't follow the overall linear trend) more than any of the other plots. The plot of attribute 6 against the target attribute was also fairly linear, however, it contained more outliers that did not follow the linear trend than the plot of attribute 13 against the target attribute.

The plot of attribute 4 against the target attribute looks the most nonlinear. This is because attribute 4 is a binary variable. It's plot has very little resemblance to a single linear line. It more resembles two independent vertical lines.



The plot of attribute 13 against the target attribute



The plot of attribute 4 against the target attribute.

(d)
Attribute 9, index of accessibility to radial highways, and attribute 10, full-value property-tax rate per $10,000, have the largest mutual correlation: 0.9102

Part 2.

(a)
Function is written
(b)
Function is written
(c)
Script is written
(d)
Resulting weights:
39.5843
-0.1011

```
0.0459
-0.0027
3.0720
-17.2254
3.7113
0.0072
-1.5990
0.3736
-0.0158
-1.0242
0.0097
-0.5860
```

training mean squared error = 22.0813

testing mean squared error = 22.6383

The mean squared error for the testing set is higher and thus the predictions for the training set were better.

Part 3.

(a)
This procedure was submitted and is named online_gradient_desc.m

(b)
The program was written and submitted.

Train error = 39.8850

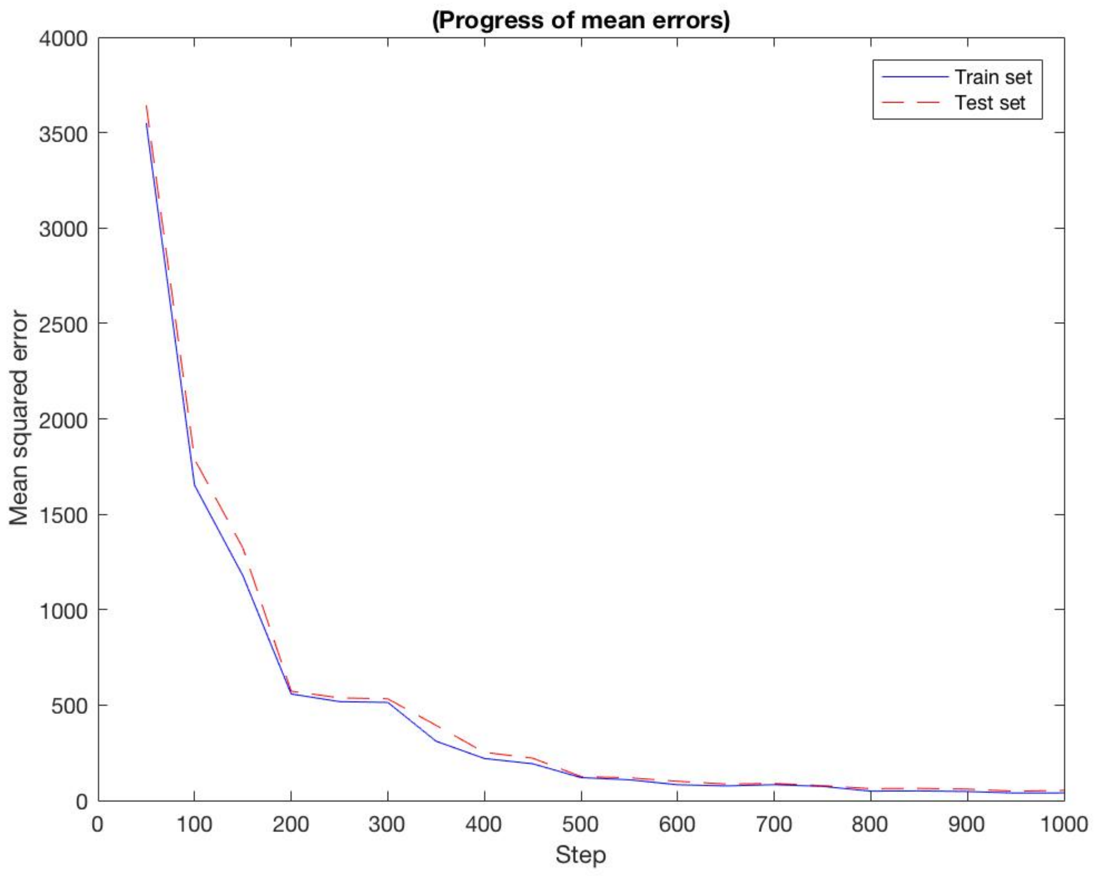
Test error = 53.3013

These results are worse than those that were obtained by solving the regression problem exactly.

(c)
After running online_gradient_desc.m on the untrained housing_train dataset, the weights returned were all of type NaN.

It appears that the weights eventually got so large or so small that they reached either Inf or –Inf. Once they became Inf or –Inf any operation peformed using these values would produce NaN.

(d)



(e)

Learning rate = 0.05, and 500 steps
Train error = 0.6159
Test error = 0.7730

Learning rate = 0.05, and 3000 steps
Train error = 5.6026
Test error = 6.6272

This was interesting because the fit of the model got worse as more steps were added (in comparison to the errors found when using a learning rate of 0.05 and 500 steps).

Learning rate = 0.01, and 500 steps
Train error = 0.3097
Test error = 0.4218

Learning rate = 0.01, and 3000 steps
Train error = 0.2940
Test error = 0.4284

This was interesting because unlike with a learning rate of 0.05 the fit of the model was similar with both 500 steps and 3000 steps.

Learning rate = 2/(500)^(1/2), and 500 steps
Train error = 6.7252e+04
Test error = 7.3922e+04

This was notable because the fit of the model was significantly worse than any others that were seen up until this point.

Part 4.

(a)
Function is written

(b)
The binary attribute resulted in additional attribute values which were either duplicates of those multiplied by it (if the binary attribute was 1 for that row) or 0 (if the binary attribute was 0 for that row) after the transformation.

(c)
Written and submitted.

(d)
Train error = 5.4293

Test error = 36.2601

The train error is far lower for this model than it was for the model in part 2, while the test error is far greater for this model than it was for the model in part 2. This leads me to believe that there is a large amount of overfitting occurring in this model.

I would use the method from part 2 for the prediction on the test data. This is because it produces the lowest error for the test predictions. It is also not nearly as effected by overfitting the train data.