Homework 4 Austin Frownfelter

# Part 1

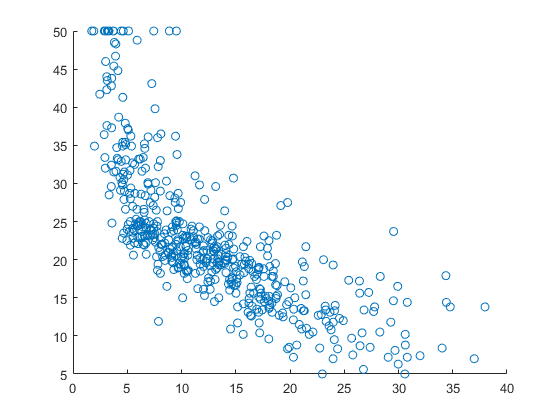
There is 1 binary attribute: the 4th attribute, CHAS

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| -0.3883 | 0.3604 | -0.4837 | 0.1753 | -0.4273 | 0.6954 | -0.377 | 0.2499 | -0.3816 | -0.4685 | -0.5078 | 0.3335 | -0.7377 |

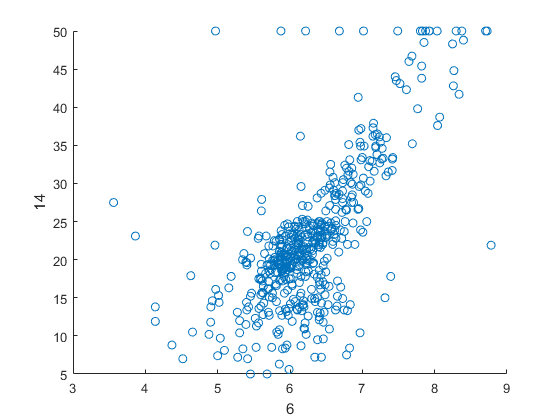
The max correlated attribute was 6 (RM) at 0.6954. The min was 13 (LSTAT) at -0.7377.

## N

The most correlated non-linear graph appears to be attribute 13. The data points are very closely packed, and the shape is not necessarily linear. The shape appears to be maybe quadratic, or an inverse function. It also happens to have the highest (negative) correlation coefficient.



The most correlated linear graph appears to be between 6 and 14. In this case, there are a few outliers, but many points packed in to where a line would fit. It also happens to have the highest correlation coefficient.



9 and 10 have the largest mutual correlation at 0.910228.

5 and 8 have the second largest mutual correlation at 0.76923.

# Part 2

## (D)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| -0.0979 | 0.049 | -0.0254 | 3.4509 | -0.3555 | 5.8165 | -0.0033 | -1.0205 | 0.2266 | -0.0122 | -0.388 | 0.017 | -0.485 |

Mean Squared Error on the test set was 24.2922. The MSE for the training data was 24.4759.

# Part 3

## (B)

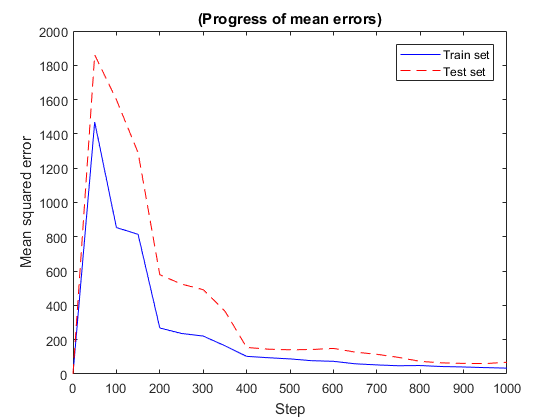
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0.2343 | 2.0068 | 0.4363 | -0.3487 | 8.5354 | 1.358 | -5.4657 | 1.0777 | -4.7237 | 0.4721 | 2.4667 | 2.6942 | -0.3014 |

MSE for the test set was 68.2374; for the training set was 34.2648. In this case, the result was worse than by solving the problem exactly.

## (C)

With unnormalized data, the weights approached infinity instantly. Therefore, weights and errors were not valid.

## (D)



## (E)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Steps | 1000 | 1000 | 1000 | 1000 | 1000 | 1000 | 10000 | 100000 | 1000000 | 64 mil |
| Learning Rate | 0.01 | 0.05 | 1/sqrt(t) | 2/t^2 | 2/t^3 | 2/t | 2/t | 2/t | 2/t | 2/t |
| MSE Train | 0.2895 | 0.9875 | 3.66E+55 | 0.8247 | 0.5577 | 34.2648 | 5.3167 | 1.5551 | 0.6581 | 0.3325 |
| MSE Test | 0.1909 | 0.8354 | 4.27E+55 | 0.9502 | 0.5534 | 68.2374 | 8.4973 | 2.1502 | 0.736 | 0.3115 |
| Weights | -0.0639 | 0.4199 | -6.4E+27 | -0.1014 | -0.0684 | 0.2343 | 0.3515 | 0.1312 | 0.0301 | -0.0387 |
|  | 0.0383 | -0.1958 | 9.07E+26 | -0.1454 | -0.0823 | 2.0068 | 0.2848 | 0.0628 | 0.0266 | 0.0468 |
|  | 0.0038 | 0.2318 | -5.8E+25 | 0.2605 | 0.0154 | 0.4363 | -0.6689 | -0.6397 | -0.4403 | -0.2054 |
|  | 0.0401 | -0.112 | 2.49E+26 | -0.0703 | -0.0452 | -0.3487 | 0.1589 | 0.1617 | 0.14 | 0.1119 |
|  | -0.1707 | -0.3059 | 1.32E+27 | -0.1729 | -0.1091 | 8.5354 | 4.0562 | 1.8411 | 0.7784 | 0.0388 |
|  | 0.2284 | 0.2222 | 1.34E+27 | 1.01 | 0.5398 | 1.358 | 1.1946 | 0.7857 | 0.5552 | 0.3847 |
|  | -0.0005 | 0.2016 | -1.4E+27 | 0.1962 | -0.1313 | -5.4657 | -1.8628 | -0.7714 | -0.3458 | -0.086 |
|  | -0.1767 | -0.2337 | 6.56E+26 | -0.1858 | 0.017 | 1.0777 | 0.5803 | 0.2096 | -0.0531 | -0.2722 |
|  | 0.1175 | 0.0543 | 1.58E+27 | -0.3084 | -0.1616 | -4.7237 | -2.6102 | -1.6003 | -1.036 | -0.482 |
|  | -0.0507 | -0.2419 | 4.12E+26 | -0.2002 | -0.1405 | 0.4721 | 0.9293 | 0.9607 | 0.8309 | 0.519 |
|  | -0.1213 | -0.0677 | -1.4E+26 | -0.2226 | -0.1121 | 2.4667 | 1.345 | 0.5477 | 0.1545 | -0.1127 |
|  | 0.0385 | 0.0768 | -5.9E+25 | 0.084 | 0.0449 | 2.6942 | 0.5061 | 0.2595 | 0.1687 | 0.1063 |
|  | -0.4389 | -0.3127 | 7.67E+26 | -0.2942 | -0.3686 | -0.3014 | -0.2189 | -0.2919 | -0.3539 | -0.4062 |

Keeping the steps constant while changing the learning rate will affect error the most. The smaller the learning rate goes, the lower the error became. It is likely that there is a threshold where, if the number of steps are not enough, a lower rate will have worse error (since the regression never approaches the optimum). If the learning rate is annealed, then the change becomes less noticeable as the function becomes “smaller and smaller” (such as 2/t^2 -> 2/t^3). Now, if the learning function remains the same and the steps increases, the errors go down, albeit with diminishing returns. I basically halved the error every power of 10 I increased. I halted testing 100 million after 8 minutes and it only completed 64 million on an i7-6700K processor.

# Part 4

## (B)

The binary attribute became a “selector” of whether the other variables would affect the function.

## (D)

The MSE for test data was 36.2601, while the training data was 5.4293. I would use the method in part 2 over this method, in this case for two reasons. First, the MSE for the test data was worse here than in part 2. Second, the disparity between the test/training error is concerning, since the system “thinks” it is getting better when it really isn’t; in the case of part 2, the training data’s error was very close to the test’s error.