Dominic DiPasquale
CS1567 Introduction to Machine Learning
Professor Milos Hauskrecht

Problem Assignment 4

Problem 1:

(a) There is one, it is CHAS
(b) **RM** has the greatest positive correlation, **LSTAT** has the greatest negative correlation.
(a) CRIM has a correlation: -0.388305
(b) ZN has a correlation: 0.360445
(c) INDUS has a correlation: -0.483725
(d) CHAS river has a correlation: 0.175260
(e) NOX has a correlation: -0.427321
(f) RM has a correlation: 0.695360
(g) AGE has a correlation: -0.376955
(h) DIS has a correlation: 0.249929
(i) RAD has a correlation: -0.381626
(j) TAX has a correlation: -0.468536
(k) PTRATIO has a correlation: -0.507787
(l) B has a correlation: 0.333461
(m) LSTAT has a correlation: -0.737663
(c) Most linear: RM/LSTAT
Least linear: appears to be the binary variable CHAS
(d) RAD and TAX have the largest mutual correlation with .910228

Problem 2:

(a) N/A
(b) N/A
(c) N/A
(d) Training MSE: 24.4759, Testing MSE: 24.2922. Because it is less the testing MSE is arguably the better one.

Attr	Weight
CRIM	-.0979
ZN	.049
INDUS	-.0254
CHAS	3.4509
NOX	-.3555
RM	5.8165
AGE	-.0033
DIS	-1.0205
RAD	.2266
TAX	-.0122
PTRATIO	-.388
B	.017
LSTAT	-.4850

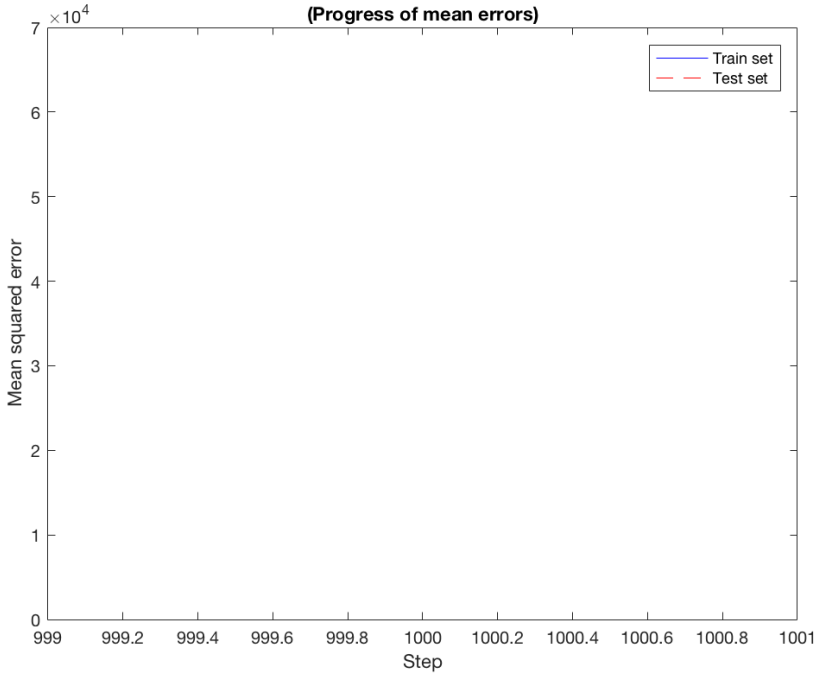
Problem 3:

(a) N/A

(b) Training MSE: 6.1415e+04, Testing MSE: 769.4491. This is much worse than the values we calculated above previously

(c) What I observed is that the weights become too large for Matlab to compute (NaNs) because we are punishing attributes with a small value and aggressively rewarding anything that is comparatively large.

(d) Here is the graph that was plotted:



(e) When I experimented with different parameters I found that my mean square errors got smaller, for example if I used 3000 iterations and .00001 learning rate my MSE dropped to: 600.3757 and 483.8409

Problem 4:

(a) N/A

(b) The binary attribute removed some values because of it's frequent 0 value and left the rest unchanged. It basically highlighted random lines in the dataset by being 1.

(c) N/A

(d) Training: 5.4293, Testing: 36.2601. The training set is represented better in the polynomial than in the plain linear, but as we can see this accuracy does not fully extend to the testing set. In this case it seems that we can represent the data better using a plain linear model, but it is only marginally better.