

HW 1

Elizabeth Hutton

August 30, 2018

1 2.1 Conceptual

- a. When n is large and p is small, there is a smaller chance that a flexible model will overfit because a large sample is more representative and having a smaller number of predictors will mean less noise.
- b. An inflexible model will perform better on a small sample in general, and is especially the case when p is relatively large. A flexible model will not reduce the bias enough to justify the increase in variance.
- c. If the relationship between predictors and response is nonlinear, a flexible model will perform better than an inflexible, simpler model.
- d. An inflexible model will perform better because the large variance indicates there is noise in the data which will be overfit by a more flexible model.

2 2.9 Applied

- a. Of the 9 variables, only *name*, *year*, and *origin* are qualitative. The other 6 are quantitative.
- b./c. Here are the ranges, means, and standard deviations for each of the quantitative predictors:

Table 1: Summary statistics for Auto data

Statistic	N	Mean	St. Dev.	Min	Max
mpg	392	23.450	7.805	9	47
cylinders	392	5.472	1.706	3	8
displacement	392	194.400	104.600	68	455
horsepower	392	104.500	38.490	46	230
weight	392	2,978.000	849.400	1,613	5,140
acceleration	392	15.540	2.759	8	25

- d. After removing the 10th through 85th samples, the ranges, means, and sd's of each predictor are:

Table 2: Summary statistics for Auto data subset

Statistic	N	Mean	St. Dev.	Min	Max
mpg	316	24.400	7.867	11	47
cylinders	316	5.373	1.654	3	8
displacement	316	187.200	99.680	68	455
horsepower	316	100.700	35.710	46	230
weight	316	2,936.000	811.300	1,649	4,997
acceleration	316	15.730	2.694	8	25

- e./f. The following plots reveal interesting relationships between MPG and many of the other predictors. For example, visual inspection shows that 4 Cylinder cars get higher MPG than any other and cars of American origin (denoted by 1) get worse MPG than cars made in Europe or Japan (2 and 3, respectively). It appears that almost all of the predictors have some correlation with MPG.

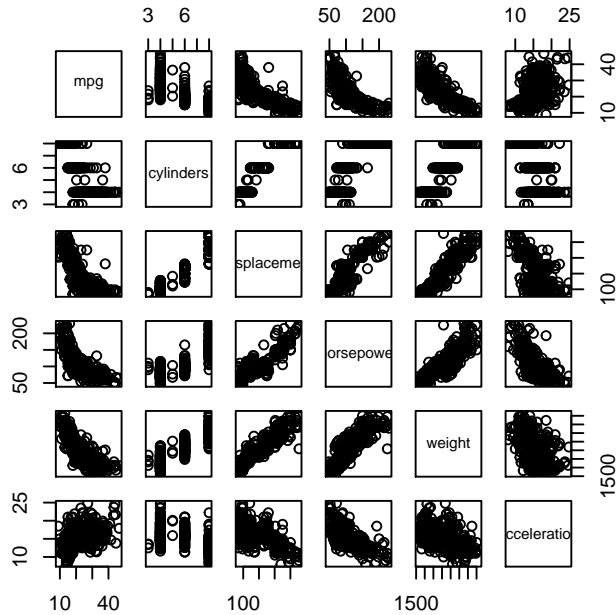
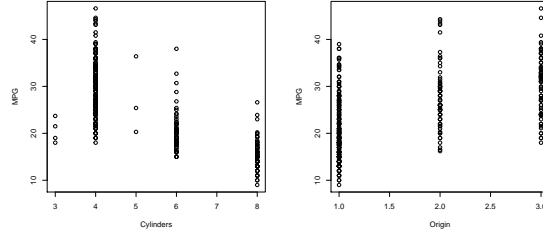


Figure 1: Matrix Plot of Predictors



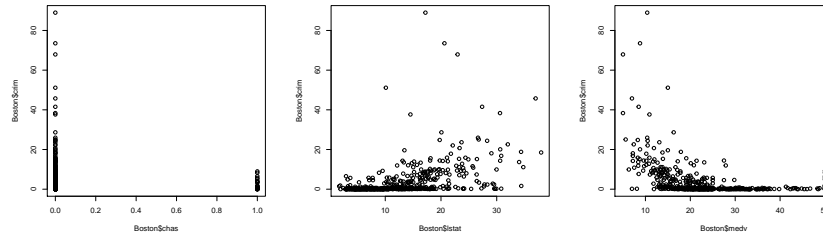
(a) MPG by # Cylinders (b) MPG by Origin

Figure 2: MPG and Related Predictors

3 2.10 Applied

a-c. After exploring the relationship between crime per capita and the other predictors in the Boston dataset, one can see that:

- There is a lower crime rate in areas adjacent to the Charles River.
- Crime rate tends to increase with percent of low status population.
- Crime rate tends to decrease as median home value increases.



(a) Crime by Charles River (1) vs. not (0). (b) Crime by low status population. (c) Crime by median home value.

Figure 3: Crime Rate and Related Predictors

- d. – The area identified in row 381 has the highest crime rate of 88.9762.
- Area 489 has the highest tax rate of 711.
- Area 355 has the highest pupil-teacher ratio of 22.
- e. 35 areas in Boston are adjacent to the Charles River.
- f. The median pupil-teacher ratio is 19.05
- g. The 399'th town has the lowest median value homes.

```
> Boston[399,]
```

```
      crim zn indus chas    nox    rm age  dis rad tax ptratio black lstat medv
399 38.35  0  18.1    0 0.693 5.453 100 1.49  24 666    20.2 396.9 30.59    5
```

Observations about the other predictors for this neighborhood:

- This neighborhood has the minimum of 0% of the residential land zoned for lots over 25,000 sq.ft.
- 100% of the buildings in this area were built before 1940.
- The neighborhood is relatively close to employment centers and radial highways.
- The neighborhood has the highest proportionate Black population in Boston.

h. Number of rooms and other observations.

- 64 suburbs have an average of seven rooms or more.
- 13 suburbs have an average of eight rooms or more.
- Looking at the summary statistics for the subset of neighborhoods with an average of 8 or more rooms, one can see that these neighborhoods have a higher proportionate Black population, their mean median home value lies above the average of the whole population, and they have relatively small low status populations.