Exercise 1

Sarah McLeod

November 1, 2017

1 Problem 1

Statistical learning is the process of learning from data. Most of statistical learning can be divided into two categories: **Supervised Learning** and **Unsupervised learning**. In Supervised Learning, you have **inputs** as well as the corresponding **outcomes** that serve to direct the learning process. The goal is to estimate some unknown function f, which serves as the information about the relationship between inputs and outputs. In Unsupervised Learning we observe only the predictors; the responses are not available. The goal of unsupervised learning is to discover something about the data, for example how it groups together. For this reason it is often referred to as Clustering. Supervised Learning is used for **inference** and **prediction**. In Prediction the goal is to estimate some unknown function f so as to accurately predict some output given a new input. This can be done for **quantitative** values, which is known as **regression** (on continuous values), or for **qualitative** (or categorical) values, which is referred to as **classification**. When talking about supervised learning it's also important to discuss **training data** and **test data**. The training data is used, as the name suggests, to train a statistical model. Test data are data that have not been given to the model during training and are used to test the performance of that model. The methods used to to estimate f can be categorized as **parametric** or **non-parametric**. Parametric methods make the assumption that the unknown function is linear, and use linear methods to estimate it. Non-parametric methods make no assumptions about the underlying form of f, and thus can use methods that have many more degrees of freedom.

2 Section 2

To show $E(Y) = argmin_c E[(Y-c)^2]$:

- $E[(Y-c)^2] = E[(Y-c)(Y-c)]$
- $= E[y^2 - 2Yc = c^2]$
- $= E[y^2] - 2cE[y] + c^2$
- to minimize solve for where the gradient is zero: $\frac{d}{dc}E[(Y-c)^2] = 0 + 2E[y] + 2c$
- the value of c where the gradient is zero is therefore: $c = E[y]$
- to verify this is a minimum take $\frac{d^2}{dc^2}$. The result is 2, which is greater than zero, therefore this value is a minimum.

3 Section 3

4 Section 4

(b) There are 111 observations total, 80 in the training set and 31 in the test set.

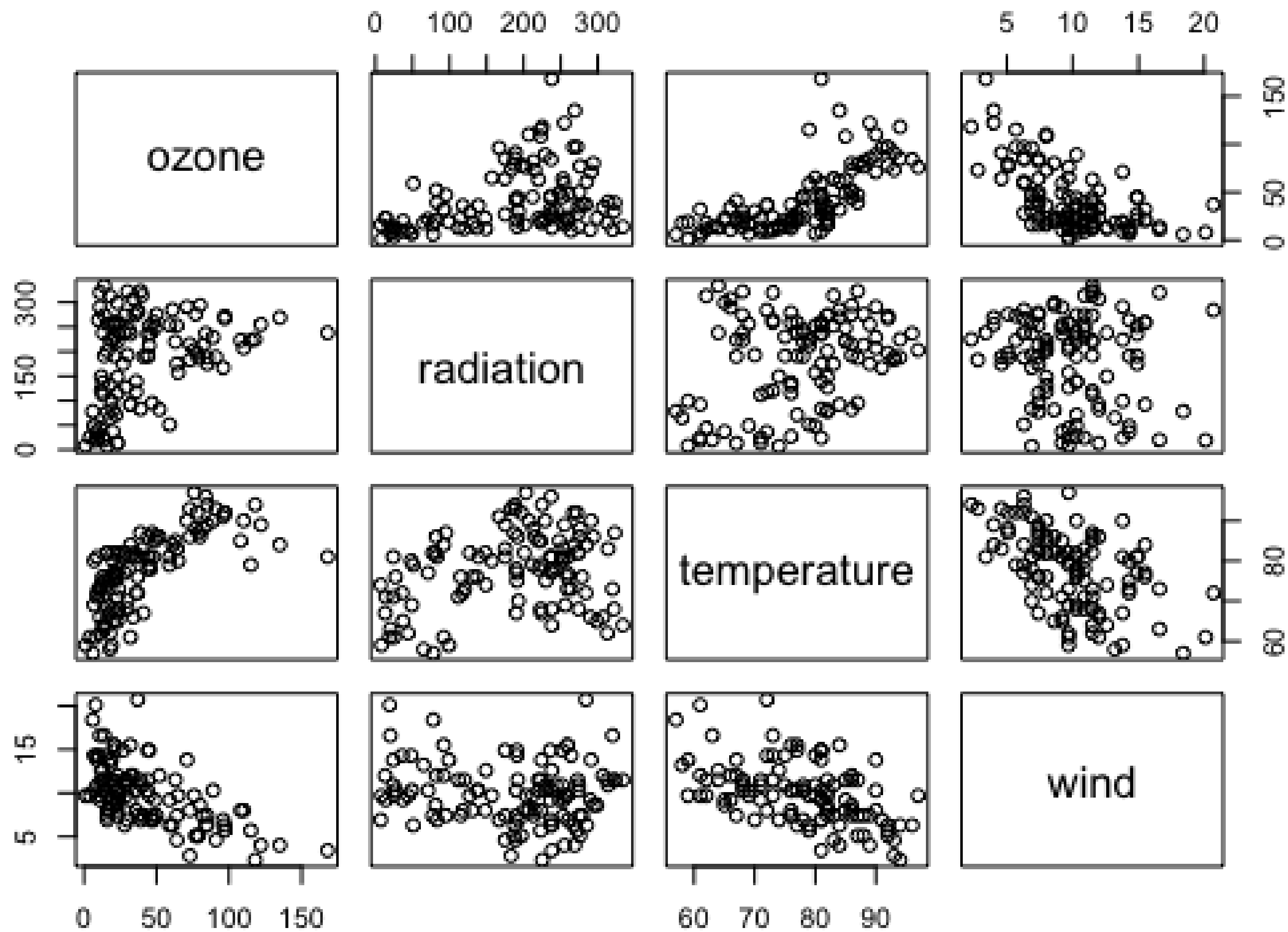
(c)

Variable	Range	Mean	SD
ozone	1 : 165	42.0991	33.27597
radiation	7 : 334	184.8018	91.1523
temperature	57 : 97	77.79279	9.529969
wind	2.3 : 20.7	9.938739	3.559218

(d) For this data r is:

	ozone	radiation	temperature	wind
ozone	1.0000000	0.3483417	0.69854147	-0.6129508
radiation	0.3483417	1.0000000	0.2940876	-0.1273656
temperature	0.6985414	0.2940876	1.0000000	-0.4971459
wind	-0.6129508	-0.1273656	-0.4971459	1.0000000

The range of the Pearson correlation coefficient is -1 to +1. A value of 0 means there is no association between the data. From the table above it looks like ozone, radiation, and temperature are positively correlated, but each of these are negatively correlated with wind. Observing the graph below it's easiest to see a potential positive correlation between temperature and ozone and a potential negative correlation between wind and ozone. In fact, the table above shows these are the largest positive and negative correlations in the data.



(e) See R code for RSS implementation.