Abed Ahmed

ASA190005

Machine Learning

Homework 2

Data Exploration using C++

a. copy/paste runs of your code showing the output

```
Opening File Boston.csv
Reading line 1
heading: rm,medv

The sum of the vector rm is: 3128.05
The sum of the vector medv is: 8949.8
The mean of the vector rm is: 6.18192
The mean of the vector medv is: 17.6874
The median of the vector rm is: 12.417
The median of the vector medv is: 42.4
The range of the vector rm is: 5.219
The range of the vector medv is: 45
The covariance between the vectors rm and medv is: 4.99214
The covariance between the vectors rm and medv is: 0.676131


Process finished with exit code 0
```

b. Describing your experience using built-in functions in R versus coding your own functions in C++

Coding this program up in C++ definitely highlighted the value of R. One can tell that R has a specific purpose, and as a result, would have taken me only %10 of the time to code a program that does the same thing!

c. Describe the descriptive statistical measures mean, median, and range, and how these values might be useful in data exploration prior to machine learning

Mean: The Average of a set of two or more numbers

Median: The middle value(s) of a sorted set of numbers

Range: The difference between the largest and smallest numbers in a set of numbers.

The values definitely provide some potentially very useful information. The average will give us a more balanced look at the data. The median compared to the average will also provide insight regarding how balanced the data is, and so will the range.

d. Describe the covariance and correlation statistics, and what information they give about two attributes. How might this information be useful in machine learning?

Covariance: Measures how changes in one variable are associated with changes in the second variable.

Correlation: Covariance that is scaled from [-1,1].

In Machine Learning, Covariance essentially indicates the direction of linear relationship between variables. Correlation measures the strength and direction of the same relationship.