Derrick Martin

CS4375.001

09/08/2022

Data Exploration

```
PS D:\C++ Code> cd "d:\C++ Code\" ; if ($?) { g++ dataexploration.cpp -o dataexploration } ; if ($?) { .\dataexploration }
Sum of rm = 3180.03
Mean of rm = 6.28463
Median of rm = 6.209
Range of rm = 5.219
StdDev of rm = 0.702617
Sum of medv = 11401.6
Mean of medv = 22.5328
Median of medv = 21.2
Range of medv = 45
StdDev of medv = 9.1971
Cov between rm and medv = 4.49345
Corr between rm and medv = 0.69536
```

When writing the code in R vs C++, I found that not only was reading the data in R significantly easier but also calling each of the functions was much easier as well. Writing the functions in C++ was not particularly hard for me since I am sort of familiar with the language, but mostly I found there was not really a way to know that the code I had written was working correctly easily. In R I can be confident the built-in methods are working correctly and not have to worry about it.

The mean is the average of a list of values, the median is the middle point of a list of values and the range is the difference between the largest and smallest values in a list of values. These can be useful when deciding if a data set is good to use for machine learning, for example if the median and mean values differ greatly then that is a sign that the data is skewed in one direction and the range can be expected to be very large in this case as well.

Correlation is a numeric value that represents how well the values between two data sets are correlated and range between -1 and 1 where -1 is a perfect negative correlation and 1 is a perfect positive correlation. This value can be useful when attempting to predict a target value from a data set that follows a linear model.

Covariance is a measure of how changes in values in one set of data are associated with changes in the other set of data. This can be good for deciding which data sets are useful as well since data with little or no covariance probably wouldn’t be good to analyze with a ML algorithm with a linear model.