Valari Graham

CS 4375

Dr. Mazidi

C++ Code Overview

1. Output:

// Opening file Boston.csv

Reading line 1

Heading: rm,medv

New length: 506

Closing file Boston.csv

Number of records: 506

Stat results for rm:

Sum: 3180.03

Mean: 6.28463

Median: 6.2085

Range: 8.78

Stat results for medv:

Sum: 11401.6

Mean: 22.5328

Median: 21.2

Range: 50

Covariance = 4.49345

Correlation = 0.69536

Program termianted.

2. I find coding in C++ to be quite fun. I enjoy using the built-in libraries to make coding simpler. Below is a list of built-in function that were used in the program. Honestly, coding in C++ does make me more appreciative of how interpretive the R language is. R has so many built in functions that are way more useful for data science applications in machine learning.
3. C++ Built-in Function
    a. <cmath>
        i. sqrt()
    b. <iostream>
        i. cin()
        ii. cout()
    c. <algorithm>
        i. sort()
    d. <vector>
        i. vector<>
    e. <string>
        i. begin()
        ii. end()
4. Descriptive measures like mean are used to convey the average of all the values. The median is the middle value of the vector when all the values are put into ascending order. If the vector is an even number, then there will be two middle numbers. To get the median, we take the mean of those two numbers. The range of the data tells us the max value minus the min value. In machine learning, the mean is useful to tell us the average of a vector. It summarizes the vector into one value. The median is useful as it represents the midpoint of the data and can help with categorizing the data. For example, the midpoint of temperature for days could be useful to split the data into hot days and cold days. The range is useful for normalization in scaling down the data set.
5. Covariance is the measure at which two random variables change together. On the other hand, correlation is how strongly two variables are related to each other. Covariance can be described as a measure of correlation and tells us whether two variables vary in the same direction. Correlation values reside between -1 and 1. Both are a measure of linear relationships between two variables. By finding the covariance and correlation between two variables, the machine learning algorithms can use this information to make future inferences on new data and learn how other variables could affect these two attributes.