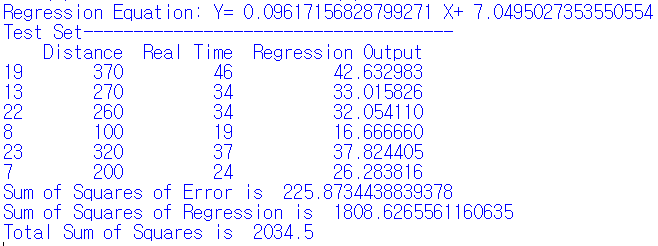
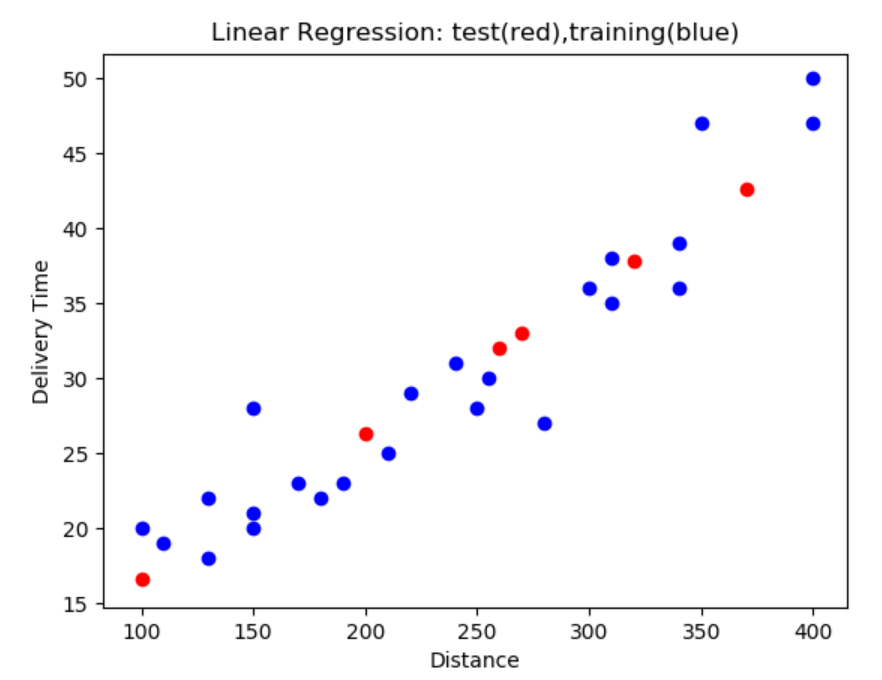
**Data Science lab4 201533661 이승수**

**Problem1: Linear Regression(Hold-out method)**



I make index list in size of data set and randomly shuffle it and devide into testing set 1/5 and training set 4/5. Then make linear regression from testing set and predict test case output and it is almost close with original data.

Then I compute SSE(Sum of Squares of Error) and SSR(Sum of Squares of Regression). Then get SST(Total Sum of Squares) in two ways, SST=SSE+SSR and SST=sum((X-meanY)^2).

Result is two of them are same.

**Problem2: Decision Tree(Hold-out method)**

**Problem3: K-Nearest Neighbors(K-fold cross validation)**

For 75 records, divide it into 5-fold, each subset have 15 records. Try 5 times of training and get mean of standard deviation of each.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Subset0(0~14) | Subset1(15~29) | Subset2(30~44) | Subset3(45~59) | Subset4(60~74) |
| Test set | Training set | Training set | Training set | Training set |
| Training set | Test set | Training set | Training set | Training set |
| Training set | Training set | Test set | Training set | Training set |
| Training set | Training set | Training set | Test set | Training set |
| Training set | Training set | Training set | Training set | Test set |