

Assignment 1

Junbin Wu

2024-03-10

Loading Library

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.2.1
```

Loading data

```
# in this data set, each column was assigned a numeric name 1-281 according to the dataset website.  
blogData_train <- read_csv("data/column name added/blogData_train.csv")
```

```
## Rows: 52397 Columns: 281  
## -- Column specification -----  
## Delimiter: ","  
## dbl (281): 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19...  
##  
## i Use 'spec()' to retrieve the full column specification for this data.  
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
blogData_test_2012_02_01_00_00 <- read_csv("data/column name added/blogData_test-2012.02.01.00_00.csv")
```

```
## Rows: 115 Columns: 281  
## -- Column specification -----  
## Delimiter: ","  
## dbl (281): 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19...  
##  
## i Use 'spec()' to retrieve the full column specification for this data.  
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
#View(blogData_train)
```

Model 1

```
model1 <- lm(`281` ~ `52`, data = blogData_train)
# use `52` as the independent variable, and `281` as the dependent variable.
model1
```

```
##
## Call:
## lm(formula = `281` ~ `52`, data = blogData_train)
##
## Coefficients:
## (Intercept)          `52`
##      2.0580         0.3094
```

model 1 testing

```
predicted <- predict(model1, newdata=blogData_test_2012_02_01_00_00[,52])
mse <- mean((predicted - blogData_test_2012_02_01_00_00$`281`)^2)
mse
```

```
## [1] 914.9985
```