Linear Regression

Chanchawat Pakdeesri

2025-05-18

Linear Regression

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```
library(caret)
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

load data

```
data("mtcars")
```

split data

70:30

```
set.seed(42)
n <- nrow(mtcars)
id<- sample(1:n, size = 0.7*n)
train_df <- mtcars[id,]
test_df <- mtcars[-id,]
```

train a linear regression model

```r
set.seed(42)
lm_model <- train(mpg ~ hp+ wt+ am,
                  data = train_df,
                  method = "lm")

summary(lm_model) # result
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.943 -1.457 -0.277  1.116  5.962
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 31.47347    2.75413  11.428  1.1e-09 ***
## hp          -0.03479    0.01060  -3.283  0.00414 **
## wt          -2.29545    0.92365  -2.485  0.02301 *
## am           2.31007    1.45682   1.586  0.13022
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.355 on 18 degrees of freedom
## Multiple R-squared:  0.839,  Adjusted R-squared:  0.8122
## F-statistic: 31.27 on 3 and 18 DF,  p-value: 2.366e-07
```

score

```r
p_test <- predict(lm_model,
                  newdata = test_df)
error <- test_df$mpg - p_test
```

evaluate

```r
mae <- mean(abs(error))
mse <- mean(error**2)
rmse <- sqrt(mse)

list(mae,mse,rmse)
```

```
## [[1]]
## [1] 2.626584
##
## [[2]]
## [1] 9.412172
##
## [[3]]
## [1] 3.067926
```