Linear Regression

Chanchawat Pakdeesri

2025-05-19

K-Nearest Neighbors (K-NN)

full loop k-fold cv

##install.packages("mlbench")

0. load data & library

```
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
```

```
library(mlbench)
data("BostonHousing")
```

1. split data

```
df <- BostonHousing
set.seed(42)
n <- nrow(df)
id<- sample(1:n,0.8*n)
train_data <- df[id,]
test_data <- df[-id,]## check train_data %>% head
```

2. training (with k-fold)

```
set.seed(42)

## hyper parameter tuning
grid_k <- data.frame(k = c(3,5))

## repeated k-fold cross validation
train_ctrl <- trainControl(method = "repeatedcv",#cv #boot #LOOCV
                           number = 5,# k=5
                           repeats = 5,
                           verboseIter = TRUE) #progress bar

knn_model <- train(medv ~ crim + indus + rm + b,
                     data = train_data,
                   method = "knn", #ranger
                   metric = "RMSE",# default is RMSE
                   trControl = train_ctrl,
                   #tuneGrid = grid_k
                   tuneLength = 3)
```

```
## + Fold1.Rep1: k=5
## - Fold1.Rep1: k=5
## + Fold1.Rep1: k=7
## - Fold1.Rep1: k=7
## + Fold1.Rep1: k=9
## - Fold1.Rep1: k=9
## + Fold2.Rep1: k=5
## - Fold2.Rep1: k=5
## + Fold2.Rep1: k=7
## - Fold2.Rep1: k=7
## + Fold2.Rep1: k=9
## - Fold2.Rep1: k=9
## + Fold3.Rep1: k=5
## - Fold3.Rep1: k=5
## + Fold3.Rep1: k=7
## - Fold3.Rep1: k=7
## + Fold3.Rep1: k=9
## - Fold3.Rep1: k=9
## + Fold4.Rep1: k=5
## - Fold4.Rep1: k=5
## + Fold4.Rep1: k=7
## - Fold4.Rep1: k=7
## + Fold4.Rep1: k=9
## - Fold4.Rep1: k=9
## + Fold5.Rep1: k=5
## - Fold5.Rep1: k=5
## + Fold5.Rep1: k=7
## - Fold5.Rep1: k=7
## + Fold5.Rep1: k=9
## - Fold5.Rep1: k=9
## + Fold1.Rep2: k=5
## - Fold1.Rep2: k=5
## + Fold1.Rep2: k=7
```

```
## - Fold1.Rep2: k=7
## + Fold1.Rep2: k=9
## - Fold1.Rep2: k=9
## + Fold2.Rep2: k=5
## - Fold2.Rep2: k=5
## + Fold2.Rep2: k=7
## - Fold2.Rep2: k=7
## + Fold2.Rep2: k=9
## - Fold2.Rep2: k=9
## + Fold3.Rep2: k=5
## - Fold3.Rep2: k=5
## + Fold3.Rep2: k=7
## - Fold3.Rep2: k=7
## + Fold3.Rep2: k=9
## - Fold3.Rep2: k=9
## + Fold4.Rep2: k=5
## - Fold4.Rep2: k=5
## + Fold4.Rep2: k=7
## - Fold4.Rep2: k=7
## + Fold4.Rep2: k=9
## - Fold4.Rep2: k=9
## + Fold5.Rep2: k=5
## - Fold5.Rep2: k=5
## + Fold5.Rep2: k=7
## - Fold5.Rep2: k=7
## + Fold5.Rep2: k=9
## - Fold5.Rep2: k=9
## + Fold1.Rep3: k=5
## - Fold1.Rep3: k=5
## + Fold1.Rep3: k=7
## - Fold1.Rep3: k=7
## + Fold1.Rep3: k=9
## - Fold1.Rep3: k=9
## + Fold2.Rep3: k=5
## - Fold2.Rep3: k=5
## + Fold2.Rep3: k=7
## - Fold2.Rep3: k=7
## + Fold2.Rep3: k=9
## - Fold2.Rep3: k=9
## + Fold3.Rep3: k=5
## - Fold3.Rep3: k=5
## + Fold3.Rep3: k=7
## - Fold3.Rep3: k=7
## + Fold3.Rep3: k=9
## - Fold3.Rep3: k=9
## + Fold4.Rep3: k=5
## - Fold4.Rep3: k=5
## + Fold4.Rep3: k=7
## - Fold4.Rep3: k=7
## + Fold4.Rep3: k=9
## - Fold4.Rep3: k=9
## + Fold5.Rep3: k=5
## - Fold5.Rep3: k=5
## + Fold5.Rep3: k=7
```

```
## - Fold5.Rep3: k=7
## + Fold5.Rep3: k=9
## - Fold5.Rep3: k=9
## + Fold1.Rep4: k=5
## - Fold1.Rep4: k=5
## + Fold1.Rep4: k=7
## - Fold1.Rep4: k=7
## + Fold1.Rep4: k=9
## - Fold1.Rep4: k=9
## + Fold2.Rep4: k=5
## - Fold2.Rep4: k=5
## + Fold2.Rep4: k=7
## - Fold2.Rep4: k=7
## + Fold2.Rep4: k=9
## - Fold2.Rep4: k=9
## + Fold3.Rep4: k=5
## - Fold3.Rep4: k=5
## + Fold3.Rep4: k=7
## - Fold3.Rep4: k=7
## + Fold3.Rep4: k=9
## - Fold3.Rep4: k=9
## + Fold4.Rep4: k=5
## - Fold4.Rep4: k=5
## + Fold4.Rep4: k=7
## - Fold4.Rep4: k=7
## + Fold4.Rep4: k=9
## - Fold4.Rep4: k=9
## + Fold5.Rep4: k=5
## - Fold5.Rep4: k=5
## + Fold5.Rep4: k=7
## - Fold5.Rep4: k=7
## + Fold5.Rep4: k=9
## - Fold5.Rep4: k=9
## + Fold1.Rep5: k=5
## - Fold1.Rep5: k=5
## + Fold1.Rep5: k=7
## - Fold1.Rep5: k=7
## + Fold1.Rep5: k=9
## - Fold1.Rep5: k=9
## + Fold2.Rep5: k=5
## - Fold2.Rep5: k=5
## + Fold2.Rep5: k=7
## - Fold2.Rep5: k=7
## + Fold2.Rep5: k=9
## - Fold2.Rep5: k=9
## + Fold3.Rep5: k=5
## - Fold3.Rep5: k=5
## + Fold3.Rep5: k=7
## - Fold3.Rep5: k=7
## + Fold3.Rep5: k=9
## - Fold3.Rep5: k=9
## + Fold4.Rep5: k=5
## - Fold4.Rep5: k=5
## + Fold4.Rep5: k=7
```

```
## - Fold4.Rep5: k=7
## + Fold4.Rep5: k=9
## - Fold4.Rep5: k=9
## + Fold5.Rep5: k=5
## - Fold5.Rep5: k=5
## + Fold5.Rep5: k=7
## - Fold5.Rep5: k=7
## + Fold5.Rep5: k=9
## - Fold5.Rep5: k=9
## Aggregating results
## Selecting tuning parameters
## Fitting k = 7 on full training set
```

3. score

```
pred_medv <- predict(knn_model,
                     newdata = test_data)
```

4. evaluate

```
error <- pred_medv - test_data$medv
test_rmse <- sqrt(mean(error **2))

knn_model$results$RMSE[1]
```

```
## [1] 6.618946
```

```
test_rmse
```

```
## [1] 6.970845
```