

Report

2024-03-21

Contents

Load Data and Package	1
MARS	2
Lasso	3

```
knitr::opts_chunk$set(  
  collapse = TRUE,  
  warning = FALSE,  
  message = FALSE,  
  fig.dim = c(10, 5),  
  fig.format = "png")
```

Load Data and Package

```
library(tidyverse)  
library(caret)  
load("./Data/recovery.RData")  
dat = dat |>  
  mutate(across(c(study, gender, hypertension, diabetes, vaccine, severity), ~factor(.)))  
dat |> skimr::skim()
```

Table 1: Data summary

Name	dat
Number of rows	3000
Number of columns	16
Column type frequency:	
factor	8
numeric	8
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
gender	0	1	FALSE	2	0: 1544, 1: 1456
race	0	1	FALSE	4	1: 1967, 3: 604, 4: 271, 2: 158
smoking	0	1	FALSE	3	0: 1822, 1: 859, 2: 319
hypertension	0	1	FALSE	2	0: 1508, 1: 1492
diabetes	0	1	FALSE	2	0: 2537, 1: 463
vaccine	0	1	FALSE	2	1: 1788, 0: 1212

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
severity	0	1	FALSE	2	0: 2679, 1: 321
study	0	1	FALSE	2	A: 2000, B: 1000

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
id	0	1	1500.50	866.17	1.0	750.75	1500.50	2250.25	3000.0	
age	0	1	60.20	4.48	42.0	57.00	60.00	63.00	79.0	
height	0	1	169.90	5.97	147.8	166.00	169.90	173.90	188.6	
weight	0	1	79.96	7.14	55.9	75.20	79.80	84.80	103.7	
bmi	0	1	27.76	2.79	18.8	25.80	27.65	29.50	38.9	
SBP	0	1	130.47	7.97	105.0	125.00	130.00	136.00	156.0	
LDL	0	1	110.45	19.76	28.0	97.00	110.00	124.00	178.0	
recovery_time	0	1	42.17	23.15	2.0	31.00	39.00	49.00	365.0	

MARS

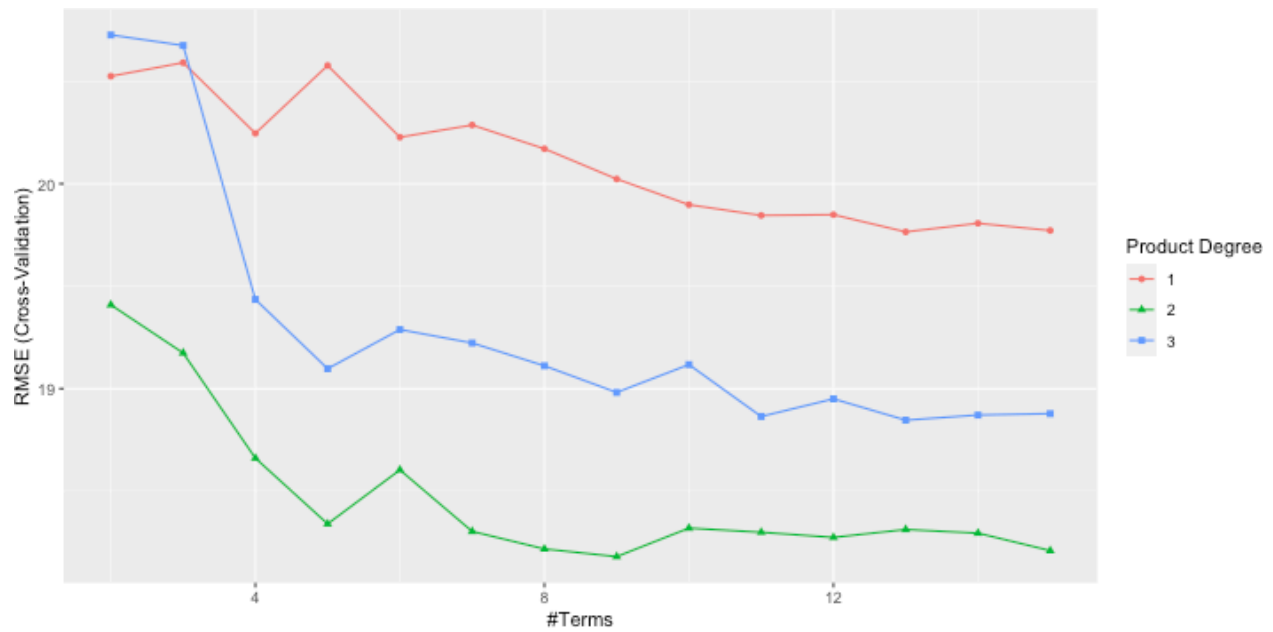
```
# matrix of predictors
x <- model.matrix(recovery_time ~ ., dat |> select(-id))[, -1]
# vector of response
y <- dat$recovery_time

mars_grid <- expand.grid(degree = 1:3,
                        nprune = 2:15)

set.seed(2)
ctrl1 <- trainControl(method = "cv", number = 10)

mars.fit <- train(x, y,
                  method = "earth",
                  tuneGrid = mars_grid,
                  trControl = ctrl1)

ggplot(mars.fit)
```



```

mars.fit$bestTune
##      nprune degree
## 22      9      2

coef(mars.fit$finalModel)
##              (Intercept)              h(30.9-bmi)
##              15.049665              4.250919
##      h(bmi-30.9) * studyB h(height-159.6) * h(bmi-30.9)
##              19.499694              1.810253
##              h(bmi-25.3)              vaccine1
##              5.145345              -5.971222
##      h(weight-85.8) * h(bmi-30.9)              h(bmi-34)
##              -2.245454              45.617060
##              severity1 * studyB
##              14.935697

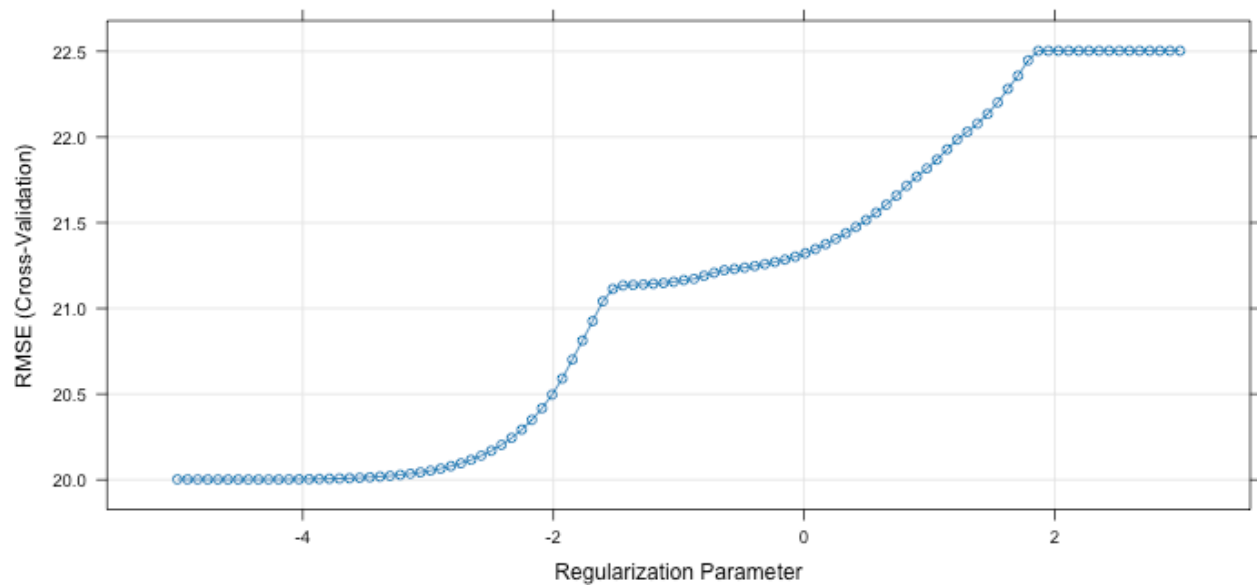
```

Lasso

```

set.seed(2)
lasso.fit <- train(x, y,
  method = "glmnet",
  tuneGrid = expand.grid(alpha = 1,
    lambda = exp(seq(3, -5, length = 100))),
  trControl = ctrl1)
plot(lasso.fit, xTrans = log)

```



```
lasso.fit$bestTune
##   alpha   lambda
## 8      1 0.01186285

# coefficients in the final model
coef(lasso.fit$finalModel, lasso.fit$bestTune$lambda)
## 18 x 1 sparse Matrix of class "dgCMatrix"
##
##          s1
## (Intercept) -2.036035e+03
## age          2.154637e-01
## gender1      -2.928606e+00
## race2         2.032812e+00
## race3        -7.431743e-01
## race4        -7.695697e-01
## smoking1      2.391978e+00
## smoking2      3.387096e+00
## height        1.187517e+01
## weight        -1.290312e+01
## bmi           3.878491e+01
## hypertension1 2.104568e+00
## diabetes1     -1.471818e+00
## SBP           5.964512e-02
## LDL           -3.776767e-02
## vaccine1      -6.375860e+00
## severity1      7.465160e+00
## studyB        4.907405e+00
```