Biostat 203B Homework 5 / Logistic regression with enet

Due Mar 20 @ 11:59PM

AUTHOR
Ningke Zhang 705834790

1. Load libraries

```
library(tidymodels)
```

```
── Attaching packages ─────────────────────────────────────── tidymodels 1.3.0 ──

✔ broom        1.0.7     ✔ recipes      1.1.1
✔ dials        1.4.0     ✔ rsample      1.2.1
✔ dplyr        1.1.4     ✔ tibble       3.2.1
✔ ggplot2      3.5.1     ✔ tidyr        1.3.1
✔ infer        1.0.7     ✔ tune         1.3.0
✔ modeldata    1.4.0     ✔ workflows    1.2.0
✔ parsnip      1.3.1     ✔ workflowsets 1.1.0
✔ purrr        1.0.4     ✔ yardstick    1.3.2

── Conflicts ────────────────────────────────────────── tidymodels_conflicts() ──
✖ purrr::discard() masks scales::discard()
✖ dplyr::filter()  masks stats::filter()
✖ dplyr::lag()     masks stats::lag()
✖ recipes::step()  masks stats::step()
```

```
library(dplyr)
library(recipes)
library(workflows)
library(tune)
library(glmnet)
```

```
Loading required package: Matrix


Attaching package: 'Matrix'

The following objects are masked from 'package:tidyr':

    expand, pack, unpack

Loaded glmnet 4.1-8
```

```
library(vip)
```

```
Attaching package: 'vip'

The following object is masked from 'package:utils':

    vi
```

2.Data preprocessing and feature engineering.

```
# read data
mimiciv_icu_cohort <- readRDS("../hw4/mimiciv_shiny/mimic_icu_cohort.rds") |>
  select(-c(intime,
            outtime,
            admittime,
            dischtime,
            deathtime,
            admit_provider_id,
            edregtime,
            edouttime,
            anchor_age,
            anchor_year,
            anchor_year_group,
            last_careunit,
            discharge_location,
            hospital_expire_flag,
            dod,
            los)
         ) |>
  mutate(los_long = as.factor(los_long)) |>
  print(width = Inf)
```

```
# A tibble: 94,458 × 26
   subject_id  hadm_id  stay_id first_careunit
        <int>    <int>    <int> <fct>
 1   10000032 29079034 39553978 Medical Intensive Care Unit (MICU)
 2   10000690 25860671 37081114 Medical Intensive Care Unit (MICU)
 3   10000980 26913865 39765666 Medical Intensive Care Unit (MICU)
 4   10001217 24597018 37067082 Surgical Intensive Care Unit (SICU)
 5   10001217 27703517 34592300 Surgical Intensive Care Unit (SICU)
 6   10001725 25563031 31205490 Medical/Surgical Intensive Care Unit (MICU/SICU)
 7   10001843 26133978 39698942 Medical/Surgical Intensive Care Unit (MICU/SICU)
 8   10001884 26184834 37510196 Medical Intensive Care Unit (MICU)
 9   10002013 23581541 39060235 Cardiac Vascular Intensive Care Unit (CVICU)
10   10002114 27793700 34672098 Other
   admission_type             admission_location     insurance language
   <fct>                      <fct>                  <chr>     <chr>
 1 EW EMER.                   EMERGENCY ROOM         Medicaid  English
 2 EW EMER.                   EMERGENCY ROOM         Medicare  English
 3 EW EMER.                   EMERGENCY ROOM         Medicare  English
 4 EW EMER.                   EMERGENCY ROOM         Private   Other
 5 Other                      PHYSICIAN REFERRAL     Private   Other
```

```
   6 EW EMER.                   Other                 Private   English
   7 URGENT                     TRANSFER FROM HOSPITAL Medicare  English
   8 OBSERVATION ADMIT          EMERGENCY ROOM        Medicare  English
   9 SURGICAL SAME DAY ADMISSION PHYSICIAN REFERRAL    Medicare  English
  10 OBSERVATION ADMIT          PHYSICIAN REFERRAL    Medicaid  English
     marital_status race  gender intime_age hematocrit bicarbonate   wbc
     <chr>          <chr> <chr>       <int>      <dbl>       <dbl> <dbl>
   1 WIDOWED        WHITE F             52       41.1          25   6.9
   2 WIDOWED        WHITE F             86       36.1          26   7.1
   3 MARRIED        BLACK F             76       27.3          21   5.3
   4 MARRIED        WHITE F             55       38.1          22  15.7
   5 MARRIED        WHITE F             55       37.4          30   5.4
   6 MARRIED        WHITE F             46       NA            NA  NA
   7 SINGLE         WHITE M             76       31.4          28  10.4
   8 MARRIED        BLACK F             77       39.7          30  12.2
   9 SINGLE         Other F             57       34.9          24   7.2
  10 <NA>           Other M             56       34.3          18  16.8
     creatinine chloride sodium potassium glucose respiratory_rate
          <dbl>    <dbl>  <dbl>     <dbl>   <dbl>            <dbl>
   1        0.7       95    126       6.7     102               24
   2        1        100    137       4.8      85               27
   3        2.3      109    144       3.9      89               24
   4        0.6      108    142       4.2     112               18
   5        0.5      104    142       4.1      87               17
   6       NA         98    139       4.1      NA               19
   7        1.3       97    138       3.9     131               17
   8        1.1       88    130       4.5     141               16
   9        0.9      102    137       3.5     288               14
  10        3.1       NA    125       6.5      95               22
     non_invasive_blood_pressure_diastolic heart_rate temperature_fahrenheit
                                     <dbl>      <dbl>                  <dbl>
   1                                    48         91                   98.7
   2                                    63         80                   97.7
   3                                   127         77                   98
   4                                    90         86                   98.5
   5                                    97         96                   97.6
   6                                    56         86                   97.7
   7                                    85        131                   97.9
   8                                    49         60                   98.1
   9                                    70         80                   97.2
  10                                    80        111                   97.9
     non_invasive_blood_pressure_systolic los_long
                                    <dbl> <fct>
   1                                   84 FALSE
   2                                  107 TRUE
   3                                  158 FALSE
   4                                  151 FALSE
   5                                  167 FALSE
   6                                   73 FALSE
   7                                  112 FALSE
   8                                  180 TRUE
```

```
 9                                  104 FALSE
10                                  112 TRUE
# ℹ 94,448 more rows
```

3.Data split

```
set.seed(203)

mimiciv_icu_cohort <- mimiciv_icu_cohort |>
  arrange(subject_id, hadm_id, stay_id) |>
  select(-c(subject_id, hadm_id, stay_id))
mimiciv_icu_cohort <- mimiciv_icu_cohort |> drop_na()

data_split <- initial_split(mimiciv_icu_cohort,
                            strata = "los_long",
                            prop = 0.5)

icu_other <- training(data_split)
icu_test <- testing(data_split)
```

4.Train logistic regression with elasticnet regularization.

```
# Define the recipe
logit_recipe <-
  recipe(los_long ~ ., data = icu_other) |>

  step_impute_median(all_numeric_predictors()) |>
  step_impute_mode(all_nominal_predictors()) |>
  step_unknown(all_nominal_predictors()) |>
  step_dummy(all_nominal_predictors()) |>
  step_nzv(all_predictors()) |>
  step_normalize(all_numeric_predictors(), -all_outcomes())

# Define the model
logit_mod <- logistic_reg(penalty = tune(), mixture = tune()) |>
  set_engine("glmnet", standardize = FALSE) |>
  print()
```

```
Logistic Regression Model Specification (classification)

Main Arguments:
  penalty = tune()
  mixture = tune()

Engine-Specific Arguments:
  standardize = FALSE

Computational engine: glmnet
```

```
# Define the workflow
logit_wf <- workflow() |>
  add_recipe(logit_recipe) |>
  add_model(logit_mod) |>
  print()
```

```
══ Workflow ════════════════════════════════════════════════════════════════════
Preprocessor: Recipe
Model: logistic_reg()

── Preprocessor ────────────────────────────────────────────────────────────────
6 Recipe Steps

• step_impute_median()
• step_impute_mode()
• step_unknown()
• step_dummy()
• step_nzv()
• step_normalize()

── Model ───────────────────────────────────────────────────────────────────────
Logistic Regression Model Specification (classification)

Main Arguments:
  penalty = tune()
  mixture = tune()

Engine-Specific Arguments:
  standardize = FALSE

Computational engine: glmnet
```

```
# Define the grid
param_grid <- grid_regular(
  penalty(range = range(-6, 2)),
  mixture(range = range(0, 1)),
  levels = c(100, 5)
) |>
  print()
```

```
# A tibble: 500 × 2
      penalty mixture
        <dbl>   <dbl>
 1 0.000001         0
 2 0.00000120       0
 3 0.00000145       0
 4 0.00000175       0
 5 0.00000210       0
 6 0.00000254       0
```

```
 7 0.00000305       0
 8 0.00000368       0
 9 0.00000443       0
10 0.00000534       0
# ℹ 490 more rows
```

5.Cross-validation

```
set.seed(203)

folds <- vfold_cv(icu_other, v = 5, strata = los_long)

# fit cross-validation
logit_fit <- logit_wf |>
  tune_grid(
    resamples = folds,
    grid = param_grid,
    metrics = metric_set(roc_auc, accuracy),
    control = control_grid(save_pred = TRUE, verbose = TRUE)
  )
```

```
i Fold1: preprocessor 1/1

✓ Fold1: preprocessor 1/1

i Fold1: preprocessor 1/1, model 1/5

✓ Fold1: preprocessor 1/1, model 1/5

i Fold1: preprocessor 1/1, model 1/5 (extracts)

i Fold1: preprocessor 1/1, model 1/5 (predictions)

i Fold1: preprocessor 1/1, model 2/5

✓ Fold1: preprocessor 1/1, model 2/5

i Fold1: preprocessor 1/1, model 2/5 (extracts)

i Fold1: preprocessor 1/1, model 2/5 (predictions)

i Fold1: preprocessor 1/1, model 3/5

✓ Fold1: preprocessor 1/1, model 3/5

i Fold1: preprocessor 1/1, model 3/5 (extracts)

i Fold1: preprocessor 1/1, model 3/5 (predictions)

i Fold1: preprocessor 1/1, model 4/5
```

```
    ✓ Fold1: preprocessor 1/1, model 4/5

    i Fold1: preprocessor 1/1, model 4/5 (extracts)

    i Fold1: preprocessor 1/1, model 4/5 (predictions)

    i Fold1: preprocessor 1/1, model 5/5

    ✓ Fold1: preprocessor 1/1, model 5/5

    i Fold1: preprocessor 1/1, model 5/5 (extracts)

    i Fold1: preprocessor 1/1, model 5/5 (predictions)

    i Fold2: preprocessor 1/1

    ✓ Fold2: preprocessor 1/1

    i Fold2: preprocessor 1/1, model 1/5

    ✓ Fold2: preprocessor 1/1, model 1/5

    i Fold2: preprocessor 1/1, model 1/5 (extracts)

    i Fold2: preprocessor 1/1, model 1/5 (predictions)

    i Fold2: preprocessor 1/1, model 2/5

    ✓ Fold2: preprocessor 1/1, model 2/5

    i Fold2: preprocessor 1/1, model 2/5 (extracts)

    i Fold2: preprocessor 1/1, model 2/5 (predictions)

    i Fold2: preprocessor 1/1, model 3/5

    ✓ Fold2: preprocessor 1/1, model 3/5

    i Fold2: preprocessor 1/1, model 3/5 (extracts)

    i Fold2: preprocessor 1/1, model 3/5 (predictions)

    i Fold2: preprocessor 1/1, model 4/5

    ✓ Fold2: preprocessor 1/1, model 4/5

    i Fold2: preprocessor 1/1, model 4/5 (extracts)

    i Fold2: preprocessor 1/1, model 4/5 (predictions)

    i Fold2: preprocessor 1/1, model 5/5
```

```
    ✓ Fold2: preprocessor 1/1, model 5/5

    i Fold2: preprocessor 1/1, model 5/5 (extracts)

    i Fold2: preprocessor 1/1, model 5/5 (predictions)

    i Fold3: preprocessor 1/1

    ✓ Fold3: preprocessor 1/1

    i Fold3: preprocessor 1/1, model 1/5

    ✓ Fold3: preprocessor 1/1, model 1/5

    i Fold3: preprocessor 1/1, model 1/5 (extracts)

    i Fold3: preprocessor 1/1, model 1/5 (predictions)

    i Fold3: preprocessor 1/1, model 2/5

    ✓ Fold3: preprocessor 1/1, model 2/5

    i Fold3: preprocessor 1/1, model 2/5 (extracts)

    i Fold3: preprocessor 1/1, model 2/5 (predictions)

    i Fold3: preprocessor 1/1, model 3/5

    ✓ Fold3: preprocessor 1/1, model 3/5

    i Fold3: preprocessor 1/1, model 3/5 (extracts)

    i Fold3: preprocessor 1/1, model 3/5 (predictions)

    i Fold3: preprocessor 1/1, model 4/5

    ✓ Fold3: preprocessor 1/1, model 4/5

    i Fold3: preprocessor 1/1, model 4/5 (extracts)

    i Fold3: preprocessor 1/1, model 4/5 (predictions)

    i Fold3: preprocessor 1/1, model 5/5

    ✓ Fold3: preprocessor 1/1, model 5/5

    i Fold3: preprocessor 1/1, model 5/5 (extracts)

    i Fold3: preprocessor 1/1, model 5/5 (predictions)

    i Fold4: preprocessor 1/1
```

```
    ✓ Fold4: preprocessor 1/1

    i Fold4: preprocessor 1/1, model 1/5

    ✓ Fold4: preprocessor 1/1, model 1/5

    i Fold4: preprocessor 1/1, model 1/5 (extracts)

    i Fold4: preprocessor 1/1, model 1/5 (predictions)

    i Fold4: preprocessor 1/1, model 2/5

    ✓ Fold4: preprocessor 1/1, model 2/5

    i Fold4: preprocessor 1/1, model 2/5 (extracts)

    i Fold4: preprocessor 1/1, model 2/5 (predictions)

    i Fold4: preprocessor 1/1, model 3/5

    ✓ Fold4: preprocessor 1/1, model 3/5

    i Fold4: preprocessor 1/1, model 3/5 (extracts)

    i Fold4: preprocessor 1/1, model 3/5 (predictions)

    i Fold4: preprocessor 1/1, model 4/5

    ✓ Fold4: preprocessor 1/1, model 4/5

    i Fold4: preprocessor 1/1, model 4/5 (extracts)

    i Fold4: preprocessor 1/1, model 4/5 (predictions)

    i Fold4: preprocessor 1/1, model 5/5

    ✓ Fold4: preprocessor 1/1, model 5/5

    i Fold4: preprocessor 1/1, model 5/5 (extracts)

    i Fold4: preprocessor 1/1, model 5/5 (predictions)

    i Fold5: preprocessor 1/1

    ✓ Fold5: preprocessor 1/1

    i Fold5: preprocessor 1/1, model 1/5

    ✓ Fold5: preprocessor 1/1, model 1/5

    i Fold5: preprocessor 1/1, model 1/5 (extracts)
```

```
i Fold5: preprocessor 1/1, model 1/5 (predictions)

i Fold5: preprocessor 1/1, model 2/5

✓ Fold5: preprocessor 1/1, model 2/5

i Fold5: preprocessor 1/1, model 2/5 (extracts)

i Fold5: preprocessor 1/1, model 2/5 (predictions)

i Fold5: preprocessor 1/1, model 3/5

✓ Fold5: preprocessor 1/1, model 3/5

i Fold5: preprocessor 1/1, model 3/5 (extracts)

i Fold5: preprocessor 1/1, model 3/5 (predictions)

i Fold5: preprocessor 1/1, model 4/5

✓ Fold5: preprocessor 1/1, model 4/5

i Fold5: preprocessor 1/1, model 4/5 (extracts)

i Fold5: preprocessor 1/1, model 4/5 (predictions)

i Fold5: preprocessor 1/1, model 5/5

✓ Fold5: preprocessor 1/1, model 5/5

i Fold5: preprocessor 1/1, model 5/5 (extracts)

i Fold5: preprocessor 1/1, model 5/5 (predictions)
```

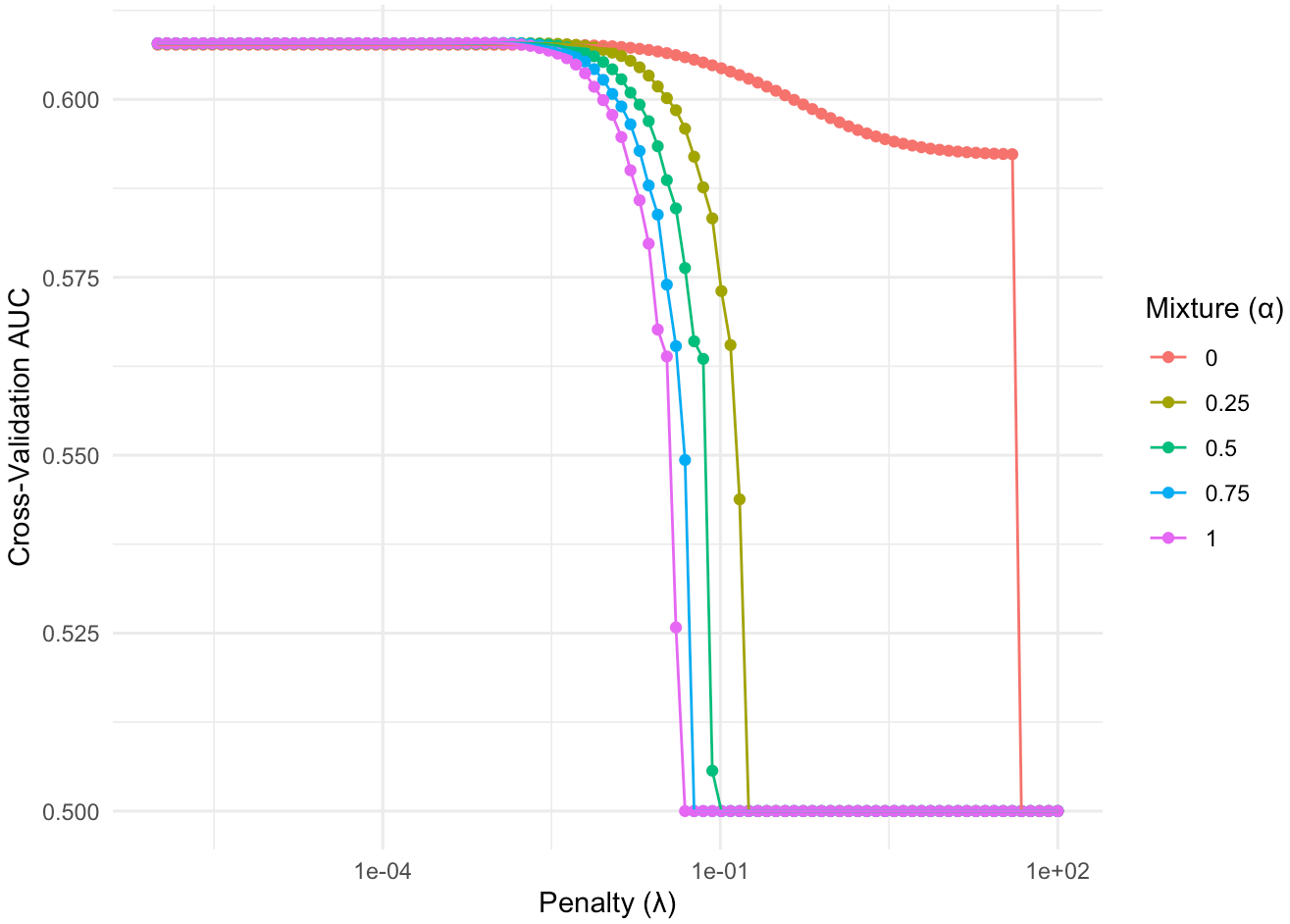
```
logit_fit
```

```
# Tuning results
# 5-fold cross-validation using stratification
# A tibble: 5 × 5
  splits               id    .metrics             .notes           .predictions
  <list>               <chr> <list>               <list>           <list>
1 <split [29744/7437]> Fold1 <tibble [1,000 × 6]> <tibble [0 × 3]> <tibble>
2 <split [29744/7437]> Fold2 <tibble [1,000 × 6]> <tibble [0 × 3]> <tibble>
3 <split [29744/7437]> Fold3 <tibble [1,000 × 6]> <tibble [0 × 3]> <tibble>
4 <split [29746/7435]> Fold4 <tibble [1,000 × 6]> <tibble [0 × 3]> <tibble>
5 <split [29746/7435]> Fold5 <tibble [1,000 × 6]> <tibble [0 × 3]> <tibble>
```

```
#visualize CV results
logit_fit |>
  collect_metrics() |>
```

```
  filter(.metric == "roc_auc") |>
  print(width = Inf) |>
  ggplot(mapping = aes(x = penalty, y = mean, color = factor(mixture),
                       group = factor(mixture))) +
  geom_point() +
  geom_line() +
  labs(x = "Penalty (λ)", y = "Cross-Validation AUC", color = "Mixture (α)") +
  scale_x_log10() +
  theme_minimal()
```

```
# A tibble: 500 × 8
      penalty mixture .metric .estimator  mean     n std_err
        <dbl>   <dbl> <chr>   <chr>      <dbl> <int>   <dbl>
 1 0.000001         0 roc_auc binary     0.608     5 0.00290
 2 0.00000120       0 roc_auc binary     0.608     5 0.00290
 3 0.00000145       0 roc_auc binary     0.608     5 0.00290
 4 0.00000175       0 roc_auc binary     0.608     5 0.00290
 5 0.00000210       0 roc_auc binary     0.608     5 0.00290
 6 0.00000254       0 roc_auc binary     0.608     5 0.00290
 7 0.00000305       0 roc_auc binary     0.608     5 0.00290
 8 0.00000368       0 roc_auc binary     0.608     5 0.00290
 9 0.00000443       0 roc_auc binary     0.608     5 0.00290
10 0.00000534       0 roc_auc binary     0.608     5 0.00290
   .config
   <chr>
 1 Preprocessor1_Model001
 2 Preprocessor1_Model002
 3 Preprocessor1_Model003
 4 Preprocessor1_Model004
 5 Preprocessor1_Model005
 6 Preprocessor1_Model006
 7 Preprocessor1_Model007
 8 Preprocessor1_Model008
 9 Preprocessor1_Model009
10 Preprocessor1_Model010
# ℹ 490 more rows
```



6. Model evaluation

```
# select the best model
best_logit <- logit_fit |> select_best(metric = "roc_auc")
print(best_logit)
```

```
# A tibble: 1 × 3
   penalty mixture .config
     <dbl>   <dbl> <chr>
1 0.000811       1 Preprocessor1_Model437
```

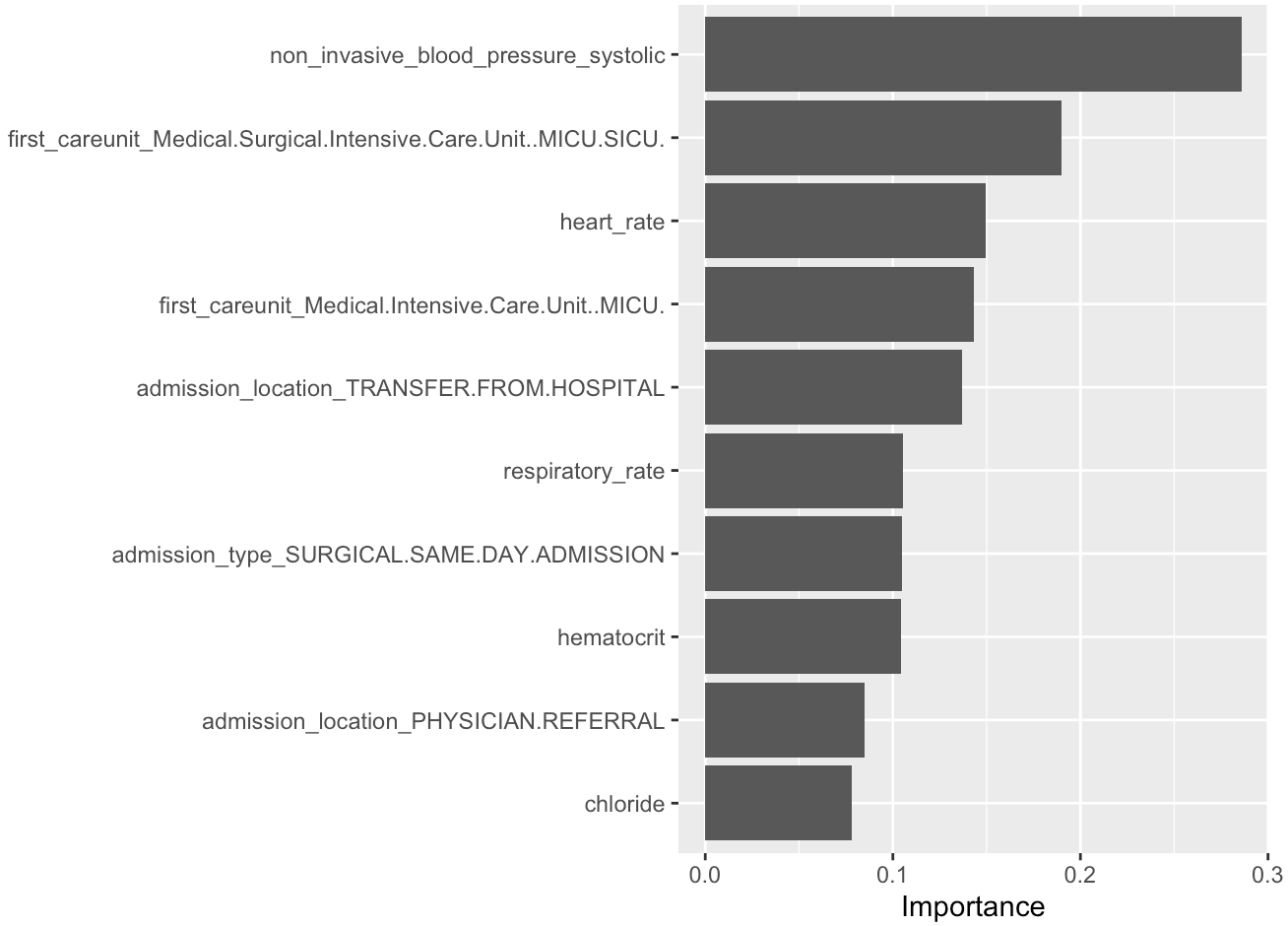
```
# finalize the workflow/fit
final_logit_wf <- finalize_workflow(logit_wf, best_logit)

final_logit_fit <- final_logit_wf |> last_fit(data_split)

saveRDS(final_logit_fit, "final_fit_logistic_lastfit.rds")

final_logit_model <- final_logit_fit |> extract_workflow() |>
  extract_fit_parsnip()

final_logit_model |> vip()
```




```
saveRDS(final_logit_model, "final_fit_logistic.rds")
```