Satinitigan_Karl_HW2

Karl Satinitigan

2/2/2020

Homework 2 - Classification Methods

The Bayes Classifier

library(tidyverse) library(broom) library(modelr) library(knitr) library(patchwork)

Setting random number generator seed

set.seed(200) theme_set(theme_minimal())

Simulating dataset

sim_bayes <- tibble(x1 = runif(200, -1, 1), x2 = runif(200, -1, 1), y = x1 + x1^2 + x2 + x2^2)

LDA and QDA

If linear

sim_linear <- tibble(x1 = runif(1000, -1, 1), x2 = runif(1000, -1, 1), y = x1 + x2)

split <- initial_split(sim_LDA, prop = .7) train <- training(split) test <- testing(split)

(lda_linear <- MASS::lda(sim_linear ~ x1 + x2, data = train)) (qda_linear <- MASS::qda(sim_linear ~ x1 + x2, data = train))

If non-linear

sim_nonlinear <- tibble(x1 = runif(1000, -1, 1), x2 = runif(1000, -1, 1), y = x1 + x1^2 + x2 + x2^2)

split <- initial_split(sim_LDA, prop = .7) train <- training(split) test <- testing(split)

(lda_linear <- MASS::lda(sim_nonlinear ~ x1 + x1^2 + x2 + x2^2, data = train)) (qda_linear <- MASS::qda(sim_nonlinear ~ x1 + x1^2 + x2 + x2^2, data = train))

Modeling voter turnout

mental_health <- read_csv("mental_health.csv") mental_health <- na.omit(mental_health)

split <- initial_split(mental_health, prop = .7) train <- training(split) test <- testing(split)

mental_logit <- glm(vote96 ~ ., data = train, family = binomial) (mental_logit_error <- augment(mental_logit, newdata = test) %>% as_tibble() %>% mutate(.prob = logit2prob(.fitted), .pred = factor(round(.prob))) %>% accuracy(truth = vote96, estimate = .pred))

(mental_lda <- MASS::lda(vote96 ~ mhealth_sum + age + educ + black + female + married + inc10, data = train)) (mental_qda <- MASS::qda(vote96 ~ mhealth_sum + age + educ + black + female + married + inc10, data = train))