1

Haoran Hu

2019-5-1

```
heart_disease = read_csv("..\\data\\heart.csv") %>%
    mutate(target=ifelse(target==1, "absence", "presence"))
```

```
## Parsed with column specification:
## cols(
##   age = col_double(),
##   sex = col_double(),
##   cp = col_double(),
##   trestbps = col_double(),
##   chol = col_double(),
##   fbs = col_double(),
##   restecg = col_double(),
##   thalach = col_double(),
##   exang = col_double(),
##   oldpeak = col_double(),
##   slope = col_double(),
##   ca = col_double(),
##   thal = col_double(),
##   target = col_double()
## )
```

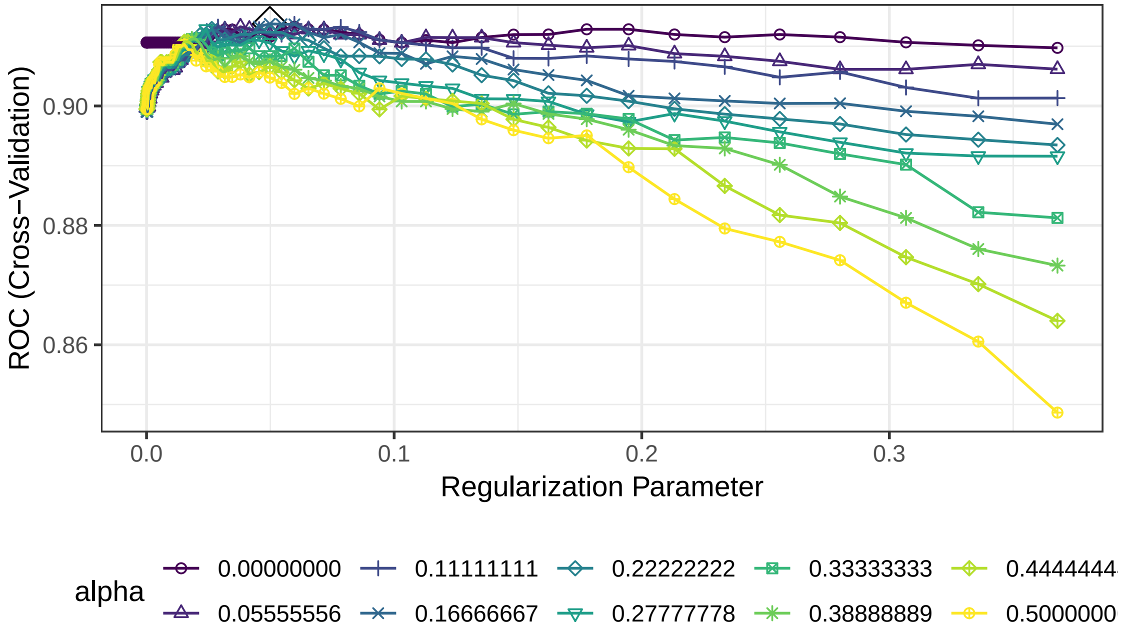
```
set.seed(1)
#trRows = createDataPartition(heart_disease$target, p = .75, list = FALSE)
#train = heart_disease[trRows,]
#test = heart_disease[-trRows,]

heart_disease = heart_disease %>%
    mutate(sex=as.factor(sex),
           cp=as.factor(cp),
           fbs=as.factor(fbs),
           restecg=as.factor(restecg),
           exang=as.factor(exang),
           slope=as.factor(slope),
           thal=as.factor(thal))
model.x <- model.matrix(target~.,heart_disease)[,-1]
model.y <- heart_disease$target

# test = test %>%
#     mutate(sex=as.factor(sex),
#            cp=as.factor(cp),
#            fbs=as.factor(fbs),
#            restecg=as.factor(restecg),
#            exang=as.factor(exang),
#            slope=as.factor(slope),
#            thal=as.factor(thal))
# test.x <- model.matrix(target~.,test)[,-1]
# test.y <- test$target
```

Regularized logistic

```
ctrl = trainControl(method = "cv",
                    classProbs = TRUE,
                    summaryFunction = twoClassSummary)
set.seed(1)

glmnGrid <- expand.grid(.alpha = seq(0, 0.5, length = 10),
                        .lambda = exp(seq(-10,-1, length = 100)))

model.glm <- train(x = model.x,
                   y = model.y,
                   method = "glmnet",
                   tuneGrid = glmnGrid,
                   metric = "ROC",
                   trControl = ctrl)
```

```
ggplot(model.glm, highlight = T)  +
    viridis::scale_color_viridis(discrete = TRUE) +
    scale_shape_manual(values = seq(1,10))
```

```
## Scale for 'colour' is already present. Adding another scale for
## 'colour', which will replace the existing scale.

## Scale for 'shape' is already present. Adding another scale for 'shape',
## which will replace the existing scale.
```



```
model.glm$bestTune
```

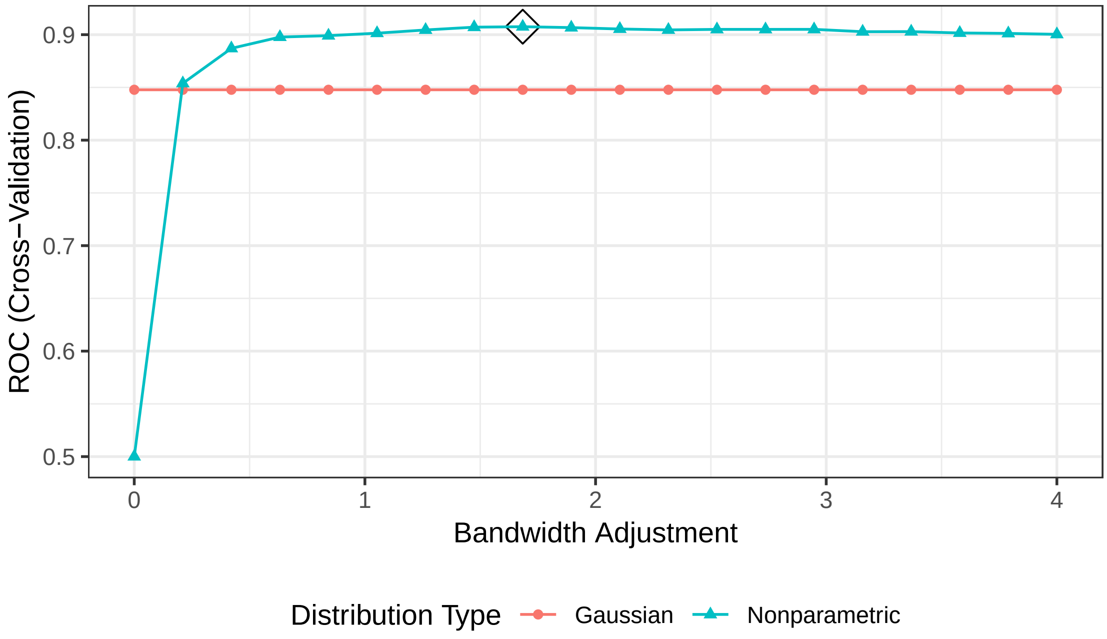
```
##         alpha     lambda
## 378 0.1666667 0.04978707
```

LDA

```
set.seed(1)
model.lda = train(x = model.x,
                   y = model.y,
                  method = "lda",
                  metric = "ROC",
                  trControl = ctrl)
```

Naive bayes

```
set.seed(1)
nbGrid = expand.grid(usekernel = c(FALSE,TRUE),
                     fL = 1, adjust = seq(0, 4, length = 20))
model.bayes = train(x = model.x,
                    y = model.y,
                    method = "nb",
                    tuneGrid = nbGrid,
                    metric = "ROC",
                    trControl = ctrl)

ggplot(model.bayes, highlight = T)
```



```
model.bayes$bestTune
```

```
##    fL usekernel   adjust
## 29  1      TRUE 1.684211
```