Cross-Validation-for-Model-Evaluation.R

patriciamaya

2020-12-07

```
#CROSS VALIDATION FOR MODEL EVALUATION

# divide data into k folds
# use 1 fold as test data and remaining data as training data
# record error rate k times.

library(ISLR)
```

```
## Warning: package 'ISLR' was built under R version 4.0.2
```

```
data("Carseats")
attach(Carseats)


High <- as.factor(ifelse(Sales >= 8, "YES", "NO"))
Data <- data.frame(Carseats, High)
Data <- Data[-1]
colnames(Data)[11] <- "Target"

Data <- Data[sample(nrow(Data)), ]


k <- 10
nmethod <- 1
folds <- cut(seq(1,nrow(Data)),breaks=k,labels=FALSE)
model.err <- matrix(-1,k,nmethod,dimnames=list(paste0("Fold", 1:k), c("LogitReg")))


for(i in 1:k)
{
  testindexes <- which(folds==i, arr.ind=TRUE)
  test <- Data[testindexes, ]
  train <- Data[-testindexes, ]


  LogitModel<- glm(Target~., data = train, family = "binomial")
  predicted <- predict(LogitModel, newdata = test, type = "response")
  pred_class <- as.factor(ifelse(predicted >= 0.5,"YES", "NO"))
  model.err[i] <- mean(test$Target != pred_class)
}


mean(model.err)
```

```
## [1] 0.105
```

```
#error of model is the mean of the errors in k-fold CV.
```