Homework #11

Henrique Magalhaes Rio

11/19/2019

# Question 1

## Part 1 (a)

average number of registered users in training dataset : 3651.394

## Part 2 (b)

Model 1

Variables: Season log of temperature feel weather

Model 2

Variables: log of temperature feel

Model 3

Variables: Season log of temperature feel weather

Interaction: season\*weather

## Part 3 (c)

Model 1:

In-sample MSE= 1368491

Adjusted =0.4606

Model 2:

In-sample MSE= 1734153

Adjusted =0.3164

Model 3:

In-sample MSE= 1356884

Adjusted =0.3652

## Part 4 (d)

Model 1:

Validation data MSE= 1860843

Model 2:

Validation data MSE= 1946875

Model 3:

Validation data MSE= 1864512

## Part 5 (e)

Model 1 is the best as it has the smallest MSE for the validation data.

## Part 6 (f)

Model 1 MSE for test data :42133772

# Question 2

## Part 1 (a)

Model 1:

variables: log(windspeed) and season.

interaction: log(windspeed)\*season.

Model 2:

variables: weather and temperature

Model 3:

variables: year and temperature.

## Part 2 (b)

Model 1: RMSE=582.3858

Model 2: RMSE=561.3365

Model 3: RMSE=554.9514

## Part 3 (c)

Model 3 is the best due to the smallest cross-validation RMSE.

## Part 4 (d)

MSE=215594.5

# Appendix

# Insert packages you need here  
library(knitr)  
library(ggplot2)  
library(dplyr)  
library(tidyverse)  
library(broom)  
library(splines)  
library(caret)  
  
  
bike <-read.csv("bike\_sharing.csv")  
biket <-read.csv("bike\_sharing\_test.csv")  
  
  
set.seed(2019)  
split\_indices <- createDataPartition(y=bike$registered\_users,  
 p=0.67,  
 times=1)  
  
  
train <- bike[split\_indices$Resample1,]  
validate <- bike[-split\_indices$Resample1,]  
  
mean(train$registered\_users)  
  
mod1 <- lm(registered\_users~season+log(tempfeel)+weather, data=train)  
  
mean((train$registered\_users-fitted(mod1))^2)  
  
mod2 <- lm(registered\_users~log(tempfeel), data=train)  
  
mean((train$registered\_users-fitted(mod2))^2)  
  
mod3 <- lm(registered\_users~season\*weather+log(tempfeel), data=train)  
  
mean((train$registered\_users-fitted(mod3))^2)  
  
  
mean((validate$registered\_users-predict(mod1, newdata =train))^2)  
mean((validate$registered\_users-predict(mod2, newdata =train))^2)  
mean((validate$registered\_users-predict(mod3, newdata =train))^2)  
mean((biket$registered\_users-predict(mod1, newdata =biket))^2)  
  
  
  
fitControl <- trainControl(method = "cv",  
 number = 10)  
  
  
set.seed(1) # Do this each time, with same value  
mod1cv <- train(casual\_users~log(windspeed)\*season,  
 data=bike,  
 method="lm",  
 trControl=fitControl)  
  
mod1cv  
  
  
mod2cv <- train(casual\_users~weather+temp,  
 data=bike,  
 method="lm",  
 trControl=fitControl)  
  
mod2cv  
  
  
  
mod3cv <- train(casual\_users~year+temp,  
 data=bike,  
 method="lm",  
 trControl=fitControl)  
  
mod3cv  
  
  
  
  
  
  
  
  
mean((biket$casual\_users-predict(mod3cv, newdata =biket))^2)