

Practical Machine Learning Course Project

Jamin Wong

2022-07-11

Abstract

Load data

```
trainingset <- read.csv("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-training.csv")
testing <- read.csv("https://d396qusza40orc.cloudfront.net/predmachlearn/pml-testing.csv")
```

Create cross validation set with k-fold

```
set.seed(1234)
numOfrow <- dim(trainingset)[1]
index <- sample(1:numOfrow)

training1 <- trainingset[index[sapply(1:numOfrow, function(x) x%%4 %in%
  c(1, 2, 3))], ]
cvd1 <- trainingset[index[sapply(1:numOfrow, function(x) x%%4 ==
  0)], ]

training2 <- trainingset[index[sapply(1:numOfrow, function(x) x%%4 %in%
  c(0, 2, 3))], ]
cvd2 <- trainingset[index[sapply(1:numOfrow, function(x) x%%4 ==
  1)], ]

training3 <- trainingset[index[sapply(1:numOfrow, function(x) x%%4 %in%
  c(0, 1, 3))], ]
cvd3 <- trainingset[index[sapply(1:numOfrow, function(x) x%%4 ==
  2)], ]

training4 <- trainingset[index[sapply(1:numOfrow, function(x) x%%4 %in%
  c(0, 1, 2))], ]
cvd4 <- trainingset[index[sapply(1:numOfrow, function(x) x%%4 ==
  3)], ]
```

dimension of one of the training and cross validation set

```
dim(training1)
```

```
## [1] 14717 160
```

```
dim(cvd1)
```

```
## [1] 4905 160
```

Exploring Data set