

# Qualitative Activity Recognition

*Miguel Pintor*

*30 de junio de 2016*

## Summary

In this document different machine learning algorithms will be trained in order to find the best fit for classifying qualitative activity recognition of weight lifting exercises. The data is included in Weight Lifting Exercise Dataset [1] and was taken from 9 degrees of freedom IMUs mounted on glove, armband, lumbar belt and dumbbell. Weight lifting exercise was performed correctly and with a set of common mistakes so classifying correct or four of mistaken performances is our goal.

## Loading Data

Data is divided in a training and a test set.

```
orTraining <- read.csv("data/pml-training.csv")
orTesting <- read.csv("data/pml-testing.csv")
dim(orTraining)
```

```
## [1] 19622 160
```

```
dim(orTesting)
```

```
## [1] 20 160
```

Up to 160 columns are included in both training and testing data so some previous tidying and variable preprocessing should be done using large datasets techniques.

First of all, column class will be verified:

```
classes <- sapply(orTraining[,1:160], class)
table(classes)
```

```
## classes
## factor integer numeric
##      37      35      88
```

Data from IMUs is in columns from 8 to 159 so our train and test features are reduced to these columns. Also, many factor columns are found so I will coerce this columns to numeric. NAs will be cleared and low variance features removed.

```
# subset columns for features
trainFeatures <- orTraining[,8:159]
testFeatures <- orTesting[,8:159]
# coerce factors to numeric
for(i in seq(1, dim(trainFeatures)[2])){
  if(is.factor(trainFeatures[,i])){
```

```

    trainFeatures[,i] <- as.numeric(as.character(trainFeatures[,i]))
    testFeatures[,i] <- as.numeric(as.character(testFeatures[,i]))
  }
}
# NAs to zero
trainFeatures[is.na(trainFeatures)] <- 0
testFeatures[is.na(testFeatures)] <- 0
# near zero variance features remotion
nzv <- nearZeroVar(trainFeatures)
trainFeatures <- trainFeatures[,-nzv]
testFeatures <- testFeatures[,-nzv]

```

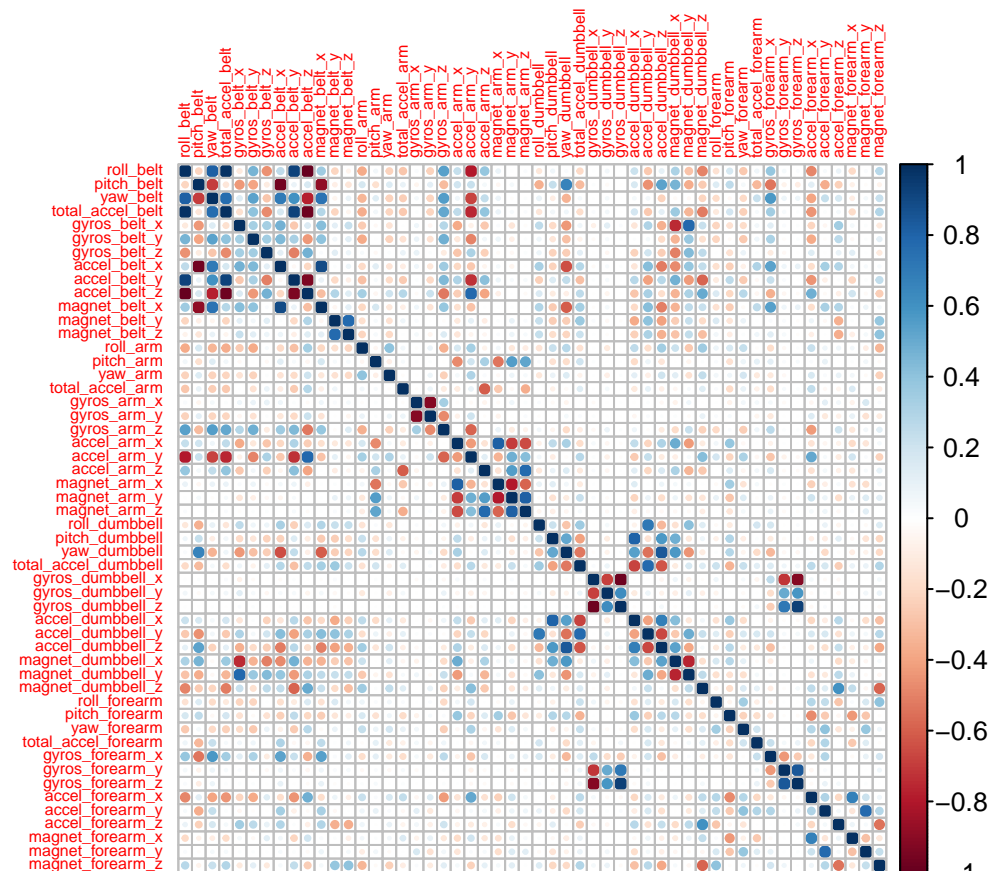
After removing near zero variance features, only 52 are included in analysis.

Let's have a look to correlation matrix for these features.

```

corMatrix <- cor(trainFeatures)
corrplot(corMatrix, method="circle", tl.cex = .5)

```



There are some high correlated features so these will be removed with default cutoff of .9.

```

corFeatures <- findCorrelation(corMatrix)
testFeatures <- testFeatures[, -corFeatures]
trainFeatures <- trainFeatures[, -corFeatures]

```

Finally, only 45 features are included in analysis.

## References

- [1] Qualitative activity recognition of weight lifting exercises E Velloso, A Bulling, H Gellersen, W Ugulino, H Fuks Proceedings of the 4th Augmented Human International Conference, 116-123