ML Project

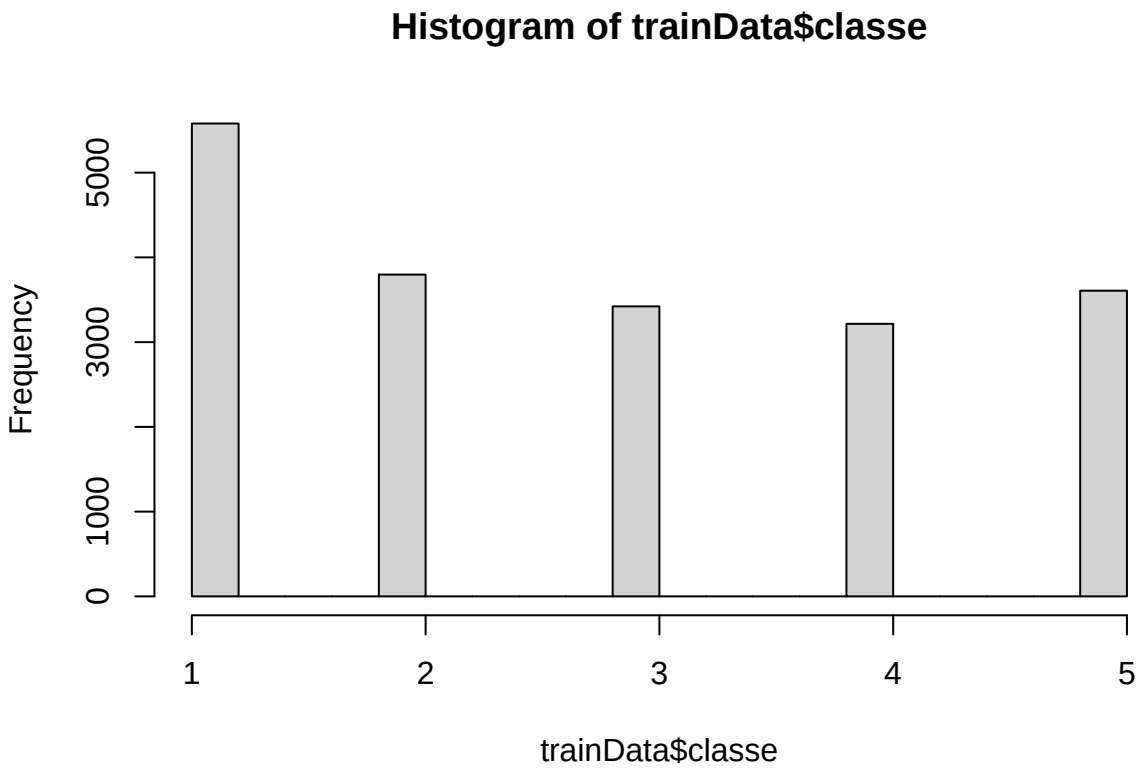
Cody

2023-08-07

ML Project

Goal: predict good exercise habits with accelerometer indicators

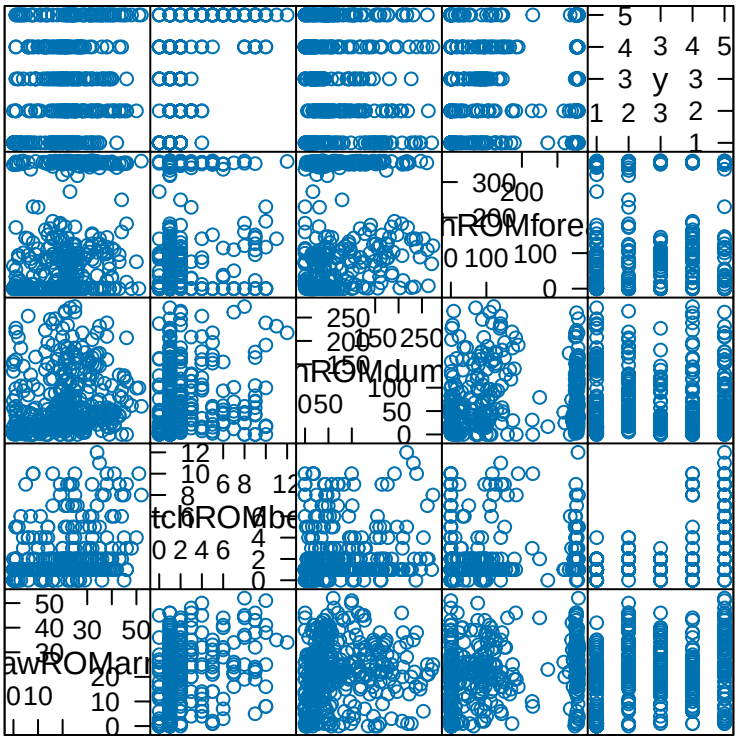
First question: does each quality level of the exercise have enough data around it?



Each quality level does appear to have enough data

Feature Plot

Looking for some kind of relationship between the range of motion and exercise quality



Training algorithm

Finding variables in our data set that have a correlation > 0.9 with another variable

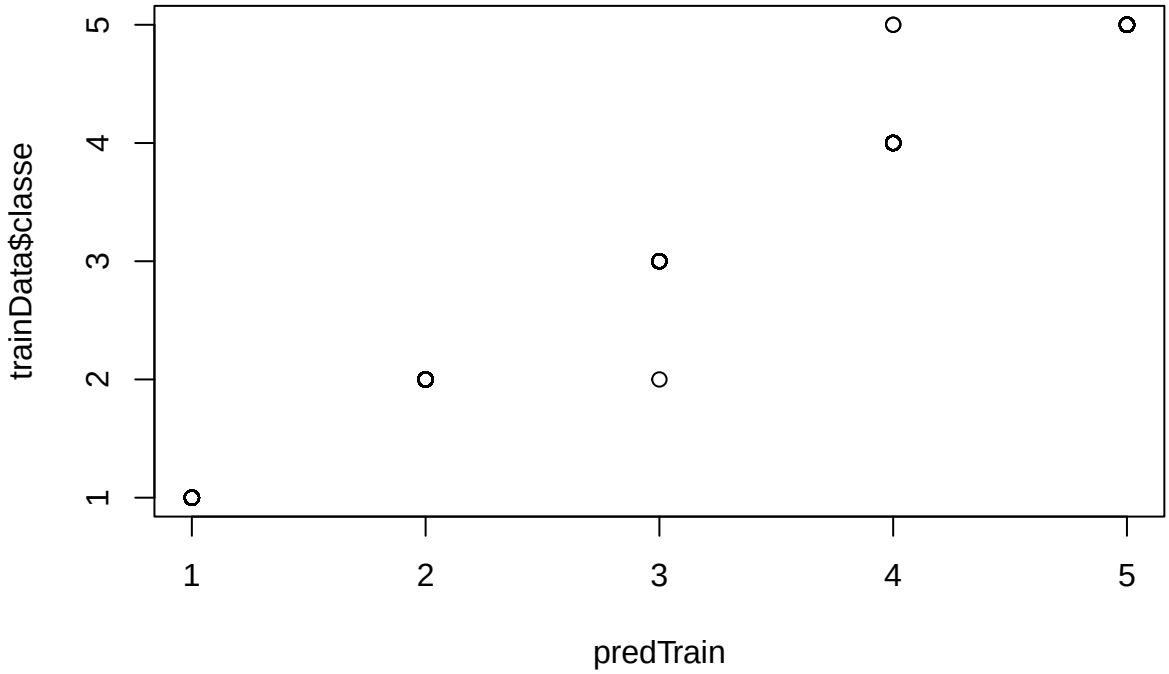
```
##                   row col
## classe             57   1
## total_accel_belt    8   5
## accel_belt_y       13   5
## accel_belt_z       14   5
## accel_belt_x       12   6
## magnet_belt_x      15   6
## roll_belt           5   8
## accel_belt_y       13   8
## accel_belt_z       14   8
## pitch_belt          6  12
## magnet_belt_x      15  12
## roll_belt           5  13
## total_accel_belt    8  13
## accel_belt_z       14  13
## roll_belt           5  14
## total_accel_belt    8  14
## accel_belt_y       13  14
## pitch_belt          6  15
## accel_belt_x       12  15
## gyros_arm_y        23  22
## gyros_arm_x        22  23
```

```
## X                   1  57
```

We should use Principal Component Analysis to process this data because there are a lot of very strong relationships

The predicitions vs the actual values, a clear linear relationship is shown

Plotting how well our prediction algorithm did in the training data set



Confusion matrix from the training dataset

Printing the actual confusion matrix information from our training data set

```
## Confusion Matrix and Statistics
##
##            Reference
## Prediction   1   2   3   4   5
##          1 109   0   0   0   0
##          2   0  78   1   0   0
##          3   0   0  70   0   0
##          4   0   0   0  69   0
##          5   0   0   0   2  77
##
## Overall Statistics
##
##                Accuracy : 0.9926
##                  95% CI : (0.9786, 0.9985)
```

```
##     No Information Rate : 0.2685
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9907
##
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: 1 Class: 2 Class: 3 Class: 4 Class: 5
## Sensitivity            1.0000   1.0000   0.9859   0.9718   1.0000
## Specificity            1.0000   0.9970   1.0000   1.0000   0.9939
## Pos Pred Value         1.0000   0.9873   1.0000   1.0000   0.9747
## Neg Pred Value         1.0000   1.0000   0.9970   0.9941   1.0000
## Prevalence             0.2685   0.1921   0.1749   0.1749   0.1897
## Detection Rate         0.2685   0.1921   0.1724   0.1700   0.1897
## Detection Prevalence   0.2685   0.1946   0.1724   0.1700   0.1946
## Balanced Accuracy      1.0000   0.9985   0.9930   0.9859   0.9970
```