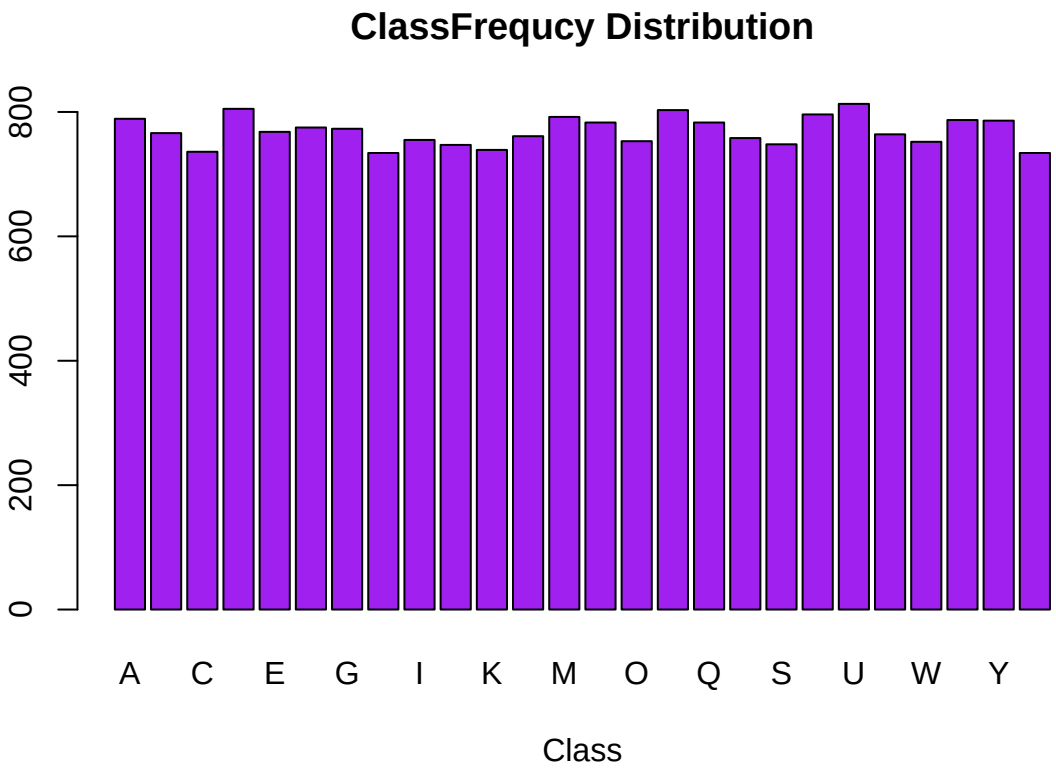
Character Classificatin UCI ML

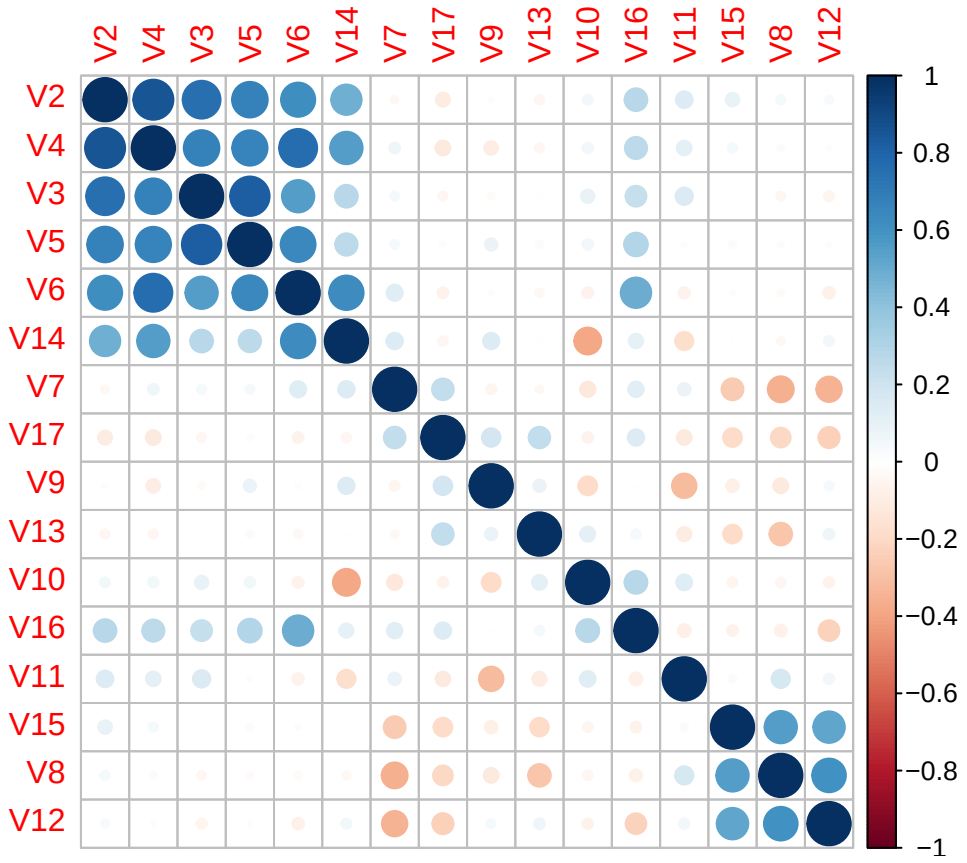
rm(list=ls()) Loading and Preprocessing : removeing near zero variance predictors and high correlated predictors

```
Data <- read.csv("./letter-recognition.data",sep=",",header=FALSE)
#Class Frequency Distribution
y <- Data[,1]
barplot(table(y),xlab="Class",col="purple",main="ClassFrequcy Distribution")
library(corrplot)
library(caret)
```

```
## Loading required package: lattice
## Loading required package: ggplot2
```




```
x <- Data[,-1]
#Correlations between predictors.
correlations <-cor(x)
corrplot(correlations, order = "hclust")
```




```
x<-x[,-findCorrelation(correlations,cutoff=0.75)]
nearZeroVar(x)
```

```
## integer(0)
```

```
smp_size <- floor(0.75 * nrow(x))

## set the seed to make your partition reproductible
set.seed(123)
train_ind <- sample(seq_len(nrow(x)), size = smp_size)

trainX <- x[train_ind, ]
trainY <- y[train_ind]
testX <-  x[-train_ind, ]
testY<-y[-train_ind]
```

Linear Models

```
ldaModel<- train(x = trainX,y = trainY,method = "lda2",metric = "Kappa",preProc = c("center", "scale"))
```

```
## Loading required package: MASS
```

```
ldaModel
```

```
## Linear Discriminant Analysis
##
## 15000 samples
##    13 predictor
##    26 classes: 'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J', 'K', 'L', 'M', 'N', 'O', 'P', 'Q',
##
## Pre-processing: centered, scaled
## Resampling: Bootstrapped (25 reps)
##
## Summary of sample sizes: 15000, 15000, 15000, 15000, 15000, 15000, ...
##
## Resampling results across tuning parameters:
##
##   dimen  Accuracy   Kappa      Accuracy SD  Kappa SD
##    1     0.1685297  0.1355126  0.008434755  0.008705092
##    2     0.3497718  0.3236751  0.006712291  0.006945867
##    3     0.4289124  0.4060080  0.015855319  0.016455423
##    4     0.5390830  0.5205938  0.007712305  0.008013558
##    5     0.5833730  0.5666671  0.006934202  0.007205122
##    6     0.6151006  0.5996624  0.008165075  0.008487258
##    7     0.6250795  0.6100264  0.009064244  0.009421349
##    8     0.6726983  0.6595546  0.006922812  0.007192980
##    9     0.6816396  0.6688550  0.006811382  0.007074126
##   10     0.6904524  0.6780292  0.006649750  0.006907253
##   11     0.6867831  0.6742183  0.007103359  0.007377300
##   12     0.6925854  0.6802618  0.007167037  0.007444857
##   13     0.6915845  0.6792207  0.006972897  0.007243935
##
## Kappa was used to select the optimal model using  the largest value.
## The final value used for the model was dimen = 12.
```

```
ldaPred <- predict(ldaModel,testX)
confusionMatrix(ldaPred,testY)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   A   B   C   D   E   F   G   H   I   J   K   L   M   N   O   P
##          A 156   0   0   1   0   0   6   0   0   1   0   7   2   0   3   0
##          B   1 132   3   5  16  12   7   5   1   4   6   8   1   0   0   5
##          C   0   0 142   0  22   0  27   0   0   0   2   3   0   0   0   0
##          D   0   6   0 143   0   3   3  16   2   7   4   0   0   5  12   1
##          E   0   0   9   0  78   0   3   0   3   0   3   5   0   0   0   0
##          F   0   0   4   0   0 121   0   1   1   2   0   0   0   0   0  15
##          G   0   1   9   1  29   2  91   0   0   0   2   8   0   0   7   3
##          H   1   7   0   3   0   2   7  83   0   0   3   1   8   7  34   3
##          I   0   2   0   2   1   1   0   1 169   9   0   2   0   0   0   0
##          J  10   0   0   3   0   0   0   0   4 132   0  10   0   2   0   0
##          K   2   7  12   0  10   2  16  20   0   0 111   3   0   2   5   2
##          L   0   0   0   0   0   0   0   0   0   0   0 139   0   0   0   0
##          M   5   0   1   0   0   0   1   1   0   0   0   0 201   9   1   0
##          N   0   0   0   5   0   1   1  10   0   0   2   0   5 134   0   1
##          O   6   1   4   8   1   0   1  12   0   2   1   0   1   3 123   2
##          P   0   0   0   0   0  20   0   5   2   8   0   0   0   0   0 142
```

```
##          Q   0   7   4   0  10   0  15   2   0   1   1   5   0   0   0   1
##          R   0  13   0   6   2   1   7  14   1   0  22   0   1   0   0   0
##          S   6  11   4   6  10  10   6   3   9  16   0   3   0   0   0   0
##          T   1   0   0   0   2   7   0   0   0   0   0   0   0   0   0   4
##          U   3   0   0   0   2   1   0   7   0   0   4   0   1   2   0   0
##          V   0   0   0   0   0   1   0   0   0   0   3   0   0   0   0   2
##          W   2   0   0   0   0   1   5   0   0   0   1   0   8  13   8   1
##          X   3  10   0   8  19   2   5   6   2   3   6   1   0   2   5   0
##          Y   2   0   0   0   0   1   0   0   0   0   0   0   0   0   0   3
##          Z   0   0   1   0   6   0   0   0   0   0   0   0   0   0   0   0
##           Reference
## Prediction   Q   R   S   T   U   V   W   X   Y   Z
##          A   3   0   8   0   0   0   0   0   0   0
##          B   7  12  20   4   0   1   0   2   0   1
##          C   0   0   0   0   2   0   0   0   0   0
##          D   0  11   0   0   1   0   0  10   1   0
##          E   2   0   1   2   0   0   0   3   0  10
##          F   0   0   6  22   0   8   0   0  14   0
##          G   6   0   2   4   0   0   0   0   0   0
##          H   0   7   0   8  12   3  10   0   4   0
##          I   0   0   5   0   0   0   0   6   0   4
##          J   0   0   2   0   0   0   0   0   0   2
##          K   6  13   0   6   2   2   1   4   0   0
##          L   3   0   0   0   1   0   0   0   0   0
##          M   0   0   0   0   3   0   2   0   0   0
##          N   1   2   0   0  10   0   7   0   1   0
##          O  22   0   0   0   8   3   0   1   1   1
##          P   0   0   0   3   0   0   0   0   0   0
##          Q 108   0   0   0   0   0   0   4   5   0
##          R   0 141   9   0   0   3   0   3   0   2
##          S   8   0  64   2   0   0   0  12   8  26
##          T   0   0   0 129   0   0   0   1   9   1
##          U   0   0   0   0 171   0   0  10   2   0
##          V   3   0   0   0   0 164   0   0  17   0
##          W   2   1   0   0   2   6 177   0   0   0
##          X   0   6  12   3   0   0   0 138   1   5
##          Y   0   0   4   5   0   4   0   6 108   0
##          Z   1   0  46   0   0   0   0   1   0 144
##
## Overall Statistics
##
##                Accuracy : 0.6882
##                  95% CI : (0.6752, 0.701)
##     No Information Rate : 0.0456
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.6757
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E Class: F
## Sensitivity            0.7879   0.6701   0.7358   0.7487   0.3750   0.6436
## Specificity            0.9935   0.9748   0.9884   0.9829   0.9914   0.9848
```

```
## Pos Pred Value         0.8342   0.5217   0.7172   0.6356   0.6555   0.6237
## Neg Pred Value         0.9913   0.9863   0.9894   0.9899   0.9734   0.9861
## Prevalence             0.0396   0.0394   0.0386   0.0382   0.0416   0.0376
## Detection Rate         0.0312   0.0264   0.0284   0.0286   0.0156   0.0242
## Detection Prevalence   0.0374   0.0506   0.0396   0.0450   0.0238   0.0388
## Balanced Accuracy      0.8907   0.8224   0.8621   0.8658   0.6832   0.8142
##                      Class: G Class: H Class: I Class: J Class: K Class: L
## Sensitivity            0.4527   0.4462   0.8711   0.7135   0.6491   0.7128
## Specificity            0.9846   0.9751   0.9931   0.9931   0.9762   0.9992
## Pos Pred Value         0.5515   0.4089   0.8366   0.8000   0.4912   0.9720
## Neg Pred Value         0.9772   0.9785   0.9948   0.9890   0.9874   0.9885
## Prevalence             0.0402   0.0372   0.0388   0.0370   0.0342   0.0390
## Detection Rate         0.0182   0.0166   0.0338   0.0264   0.0222   0.0278
## Detection Prevalence   0.0330   0.0406   0.0404   0.0330   0.0452   0.0286
## Balanced Accuracy      0.7187   0.7107   0.9321   0.8533   0.8127   0.8560
##                      Class: M Class: N Class: O Class: P Class: Q Class: R
## Sensitivity            0.8816   0.7486   0.6212   0.7676   0.6279   0.7306
## Specificity            0.9952   0.9905   0.9838   0.9921   0.9886   0.9825
## Pos Pred Value         0.8973   0.7444   0.6119   0.7889   0.6626   0.6267
## Neg Pred Value         0.9943   0.9907   0.9844   0.9911   0.9868   0.9891
## Prevalence             0.0456   0.0358   0.0396   0.0370   0.0344   0.0386
## Detection Rate         0.0402   0.0268   0.0246   0.0284   0.0216   0.0282
## Detection Prevalence   0.0448   0.0360   0.0402   0.0360   0.0326   0.0450
## Balanced Accuracy      0.9384   0.8695   0.8025   0.8798   0.8083   0.8565
##                      Class: S Class: T Class: U Class: V Class: W Class: X
## Sensitivity            0.3575   0.6862   0.8066   0.8454   0.8985   0.6866
## Specificity            0.9710   0.9948   0.9933   0.9946   0.9896   0.9794
## Pos Pred Value         0.3137   0.8377   0.8424   0.8632   0.7797   0.5823
## Neg Pred Value         0.9760   0.9878   0.9915   0.9938   0.9958   0.9868
## Prevalence             0.0358   0.0376   0.0424   0.0388   0.0394   0.0402
## Detection Rate         0.0128   0.0258   0.0342   0.0328   0.0354   0.0276
## Detection Prevalence   0.0408   0.0308   0.0406   0.0380   0.0454   0.0474
## Balanced Accuracy      0.6643   0.8405   0.9000   0.9200   0.9440   0.8330
##                      Class: Y Class: Z
## Sensitivity            0.6316   0.7347
## Specificity            0.9948   0.9886
## Pos Pred Value         0.8120   0.7236
## Neg Pred Value         0.9871   0.9892
## Prevalence             0.0342   0.0392
## Detection Rate         0.0216   0.0288
## Detection Prevalence   0.0266   0.0398
## Balanced Accuracy      0.8132   0.8616
```

```
plsdaModel <- train(x = trainX,y = trainY,method = "pls",tuneGrid = expand.grid(.ncomp = 1:2),preProc =
```

```
## Loading required package: pls
##
## Attaching package: 'pls'
##
## The following object is masked from 'package:caret':
##
##     R2
##
## The following object is masked from 'package:corrplot':
```

```
##
##     corrplot
##
## The following object is masked from 'package:stats':
##
##     loadings
```

```
plsdaModel
```

```
## Partial Least Squares
##
## 15000 samples
##    13 predictor
##    26 classes: 'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J', 'K', 'L', 'M', 'N', 'O', 'P', 'Q',
##
## Pre-processing: centered, scaled
## Resampling: Bootstrapped (25 reps)
##
## Summary of sample sizes: 15000, 15000, 15000, 15000, 15000, 15000, ...
##
## Resampling results across tuning parameters:
##
##   ncomp  Accuracy    Kappa      Accuracy SD  Kappa SD
##   1      0.07420849  0.0372415  0.003271222  0.002700414
##   2      0.17892449  0.1469015  0.010474731  0.010607809
##
## Kappa was used to select the optimal model using  the largest value.
## The final value used for the model was ncomp = 2.
```

```
plsdaPred <- predict(plsdaModel,testX)
confusionMatrix(plsdaPred,testY)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   A   B   C   D   E   F   G   H   I   J   K   L   M   N   O   P
##          A 159  41   9  67  11   3  54  53  17  61  45  95   4   2  85   4
##          B   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##          C   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##          D   0   0   0   1   0   0   0   0   0   0   1   0   0   0   0   0
##          E   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##          F   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##          G   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##          H   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##          I   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##          J   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##          K   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##          L   0  47   4  25  30   2  52   2   6  21  23  51   0   0   4   0
##          M  28  36  16  34  12  10  37  91   3   5  47   3 214  81  76  10
##          N   0   1   1   0   0   0   0   0   0   0   0   0   0   0   3   0
##          O   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##          P   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##          Q   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0   0
```

```
##          R   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##          S   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##          T   0   1  64  12  17 146   5  10  13   8  18   1   0   0   1  74
##          U   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##          V   0   0   0   0   0   0   0   2   0   0   1   0   0   0   0   0
##          W  11   3   1  12   6  12   9  12   0   1   8   6  10  96  21  25
##          X   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
##          Y   0   0   1   9   0  11   0   3   0   0   4   0   0   0   3  65
##          Z   0  68  97  30 132   4  44  13 155  89  24  39   0   0   5   7
##           Reference
## Prediction   Q   R   S   T   U   V   W   X   Y   Z
##          A  51 125  22   2   0   0   0  50   2  14
##          B   0   0   0   0   0   0   0   0   0   0
##          C   0   0   0   0   0   0   0   0   0   0
##          D   0   0   0   0   0   0   0   0   0   0
##          E   0   0   0   0   0   0   0   0   0   0
##          F   0   0   0   0   0   0   0   0   0   0
##          G   0   0   0   0   0   0   0   0   0   0
##          H   0   0   0   0   0   0   0   0   0   0
##          I   0   0   0   0   0   0   0   0   0   0
##          J   0   0   0   0   0   0   0   0   0   0
##          K   0   0   0   0   0   0   0   0   0   0
##          L  10  14  15   1   0   0   0  14   0   8
##          M  72  50   5  11  48  10  37   2   8   3
##          N   0   0   1   0   0   0   0   0   0   1
##          O   0   0   0   0   0   0   0   0   0   0
##          P   0   0   0   0   0   0   0   0   0   0
##          Q   0   0   0   0   0   0   0   0   0   0
##          R   0   0   0   0   0   0   0   0   0   0
##          S   0   0   0   0   0   0   0   0   0   0
##          T   0   0  24 154  57   0   0  32  82   9
##          U   0   0   0   0   0   0   0   0   0   0
##          V   0   0   0   0   0  12   0   0   3   0
##          W  36   2   4   2  78  72 160   1  30   0
##          X   0   0   0   0   0   0   0   0   0   0
##          Y   2   0   3  14  29 100   0   1  46   4
##          Z   1   2 105   4   0   0   0 101   0 157
##
## Overall Statistics
##
##                Accuracy : 0.1908
##                  95% CI : (0.18, 0.202)
##     No Information Rate : 0.0456
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.1571
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E Class: F
## Sensitivity            0.8030   0.0000   0.0000 0.005236   0.0000   0.0000
## Specificity            0.8299   1.0000   1.0000 0.999792   1.0000   1.0000
## Pos Pred Value         0.1629      NaN      NaN 0.500000      NaN      NaN
```

```
## Neg Pred Value         0.9903   0.9606   0.9614 0.961985   0.9584   0.9624
## Prevalence             0.0396   0.0394   0.0386 0.038200   0.0416   0.0376
## Detection Rate         0.0318   0.0000   0.0000 0.000200   0.0000   0.0000
## Detection Prevalence   0.1952   0.0000   0.0000 0.000400   0.0000   0.0000
## Balanced Accuracy      0.8164   0.5000   0.5000 0.502514   0.5000   0.5000
##                      Class: G Class: H Class: I Class: J Class: K Class: L
## Sensitivity            0.0000   0.0000   0.0000    0.000   0.0000   0.2615
## Specificity            1.0000   1.0000   1.0000    1.000   1.0000   0.9421
## Pos Pred Value            NaN      NaN      NaN      NaN      NaN   0.1550
## Neg Pred Value         0.9598   0.9628   0.9612    0.963   0.9658   0.9692
## Prevalence             0.0402   0.0372   0.0388    0.037   0.0342   0.0390
## Detection Rate         0.0000   0.0000   0.0000    0.000   0.0000   0.0102
## Detection Prevalence   0.0000   0.0000   0.0000    0.000   0.0000   0.0658
## Balanced Accuracy      0.5000   0.5000   0.5000    0.500   0.5000   0.6018
##                      Class: M Class: N Class: O Class: P Class: Q Class: R
## Sensitivity            0.9386   0.0000   0.0000    0.000   0.0000   0.0000
## Specificity            0.8460   0.9985   1.0000    1.000   0.9998   1.0000
## Pos Pred Value         0.2255   0.0000      NaN      NaN   0.0000      NaN
## Neg Pred Value         0.9965   0.9641   0.9604    0.963   0.9656   0.9614
## Prevalence             0.0456   0.0358   0.0396    0.037   0.0344   0.0386
## Detection Rate         0.0428   0.0000   0.0000    0.000   0.0000   0.0000
## Detection Prevalence   0.1898   0.0014   0.0000    0.000   0.0002   0.0000
## Balanced Accuracy      0.8923   0.4993   0.5000    0.500   0.4999   0.5000
##                      Class: S Class: T Class: U Class: V Class: W Class: X
## Sensitivity            0.0000   0.8191   0.0000  0.06186   0.8122   0.0000
## Specificity            1.0000   0.8807   1.0000  0.99875   0.9046   1.0000
## Pos Pred Value            NaN   0.2115      NaN  0.66667   0.2589      NaN
## Neg Pred Value         0.9642   0.9920   0.9576  0.96347   0.9916   0.9598
## Prevalence             0.0358   0.0376   0.0424  0.03880   0.0394   0.0402
## Detection Rate         0.0000   0.0308   0.0000  0.00240   0.0320   0.0000
## Detection Prevalence   0.0000   0.1456   0.0000  0.00360   0.1236   0.0000
## Balanced Accuracy      0.5000   0.8499   0.5000  0.53030   0.8584   0.5000
##                      Class: Y Class: Z
## Sensitivity            0.2690   0.8010
## Specificity            0.9484   0.8085
## Pos Pred Value         0.1559   0.1458
## Neg Pred Value         0.9734   0.9901
## Prevalence             0.0342   0.0392
## Detection Rate         0.0092   0.0314
## Detection Prevalence   0.0590   0.2154
## Balanced Accuracy      0.6087   0.8048
```

Non - Linear models

```
knnfit<- train(x = trainX,y = trainY,method = "knn",metric = "Kappa",preProc = c("center", "scale"),tun
library(class)
myknn<-knn(trainX,testX,as.factor(trainY), k = 3, l = 0, prob = FALSE, use.all = TRUE)
confusionMatrix(myknn,testY)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   A   B   C   D   E   F   G   H   I   J   K   L   M   N   O   P
```

```
##          A 196   0   0   0   0   0   0   0   0   1   0   0   0   1   0   0
##          B   0 187   0   0   0   0   2   4   0   0   2   0   2   0   0   1
##          C   1   0 186   0   2   0   0   1   0   0   0   0   0   0   0   0
##          D   1   0   0 187   0   1   1   8   0   0   0   0   1   2   1   1
##          E   0   0   0   0 196   0   3   1   0   0   0   1   0   0   0   0
##          F   0   0   0   1   0 175   0   0   1   0   0   0   0   0   0   8
##          G   0   0   3   0   5   1 189   1   0   0   0   1   3   0   0   0
##          H   0   2   1   3   0   0   2 159   0   0   3   0   0   3   0   1
##          I   0   0   0   0   0   1   0   0 190  12   0   0   0   0   0   1
##          J   0   0   0   0   0   0   0   0   3 171   0   1   0   0   0   0
##          K   0   0   1   0   1   0   0   3   0   0 157   0   0   1   0   0
##          L   0   0   0   0   1   0   0   0   0   0   0 190   0   1   0   0
##          M   0   0   0   0   0   0   0   1   0   0   0   0 221   0   0   0
##          N   0   0   0   0   0   0   0   1   0   0   0   0   0 167   0   0
##          O   0   0   1   0   0   0   0   0   0   0   0   1   0   0 193   0
##          P   0   0   0   0   0   7   0   0   0   0   0   0   0   0   0 170
##          Q   0   0   0   0   0   0   0   0   0   0   0   0   0   0   3   0
##          R   0   2   0   0   0   0   1   5   0   0   4   1   0   3   0   1
##          S   0   0   0   0   0   0   1   1   0   0   0   0   0   0   0   0
##          T   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0   0
##          U   0   0   0   0   0   0   1   0   0   0   1   0   1   0   0   1
##          V   0   6   1   0   0   1   1   1   0   0   0   0   0   1   0   0
##          W   0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   1
##          X   0   0   0   0   0   0   0   0   0   1   4   0   0   0   0   0
##          Y   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0   0
##          Z   0   0   0   0   3   0   0   0   0   0   0   0   0   0   0   0
##           Reference
## Prediction   Q   R   S   T   U   V   W   X   Y   Z
##          A   0   0   0   0   1   0   0   0   0   0
##          B   0   4   0   2   0   4   0   0   0   0
##          C   0   0   0   0   0   0   0   0   0   0
##          D   0   0   0   0   0   0   0   2   0   0
##          E   0   0   0   0   0   0   0   2   0   3
##          F   0   0   3   1   0   0   0   0   1   0
##          G   0   0   0   0   0   0   1   0   0   0
##          H   0   2   1   1   1   0   0   0   0   0
##          I   0   0   0   0   0   0   0   0   0   1
##          J   0   0   1   0   0   0   0   0   0   0
##          K   0   1   0   0   0   0   0   1   0   0
##          L   0   0   0   0   0   0   0   0   0   0
##          M   0   0   0   0   0   0   0   0   0   0
##          N   0   3   0   0   0   0   1   0   0   0
##          O   7   0   0   0   0   0   1   1   0   0
##          P   1   0   0   0   0   0   0   0   0   0
##          Q 163   0   0   0   0   0   0   0   0   4
##          R   1 183   2   0   0   0   0   1   0   0
##          S   0   0 172   1   0   0   0   0   0   0
##          T   0   0   0 182   0   0   0   0   0   0
##          U   0   0   0   0 209   0   0   0   0   0
##          V   0   0   0   0   1 188   0   0   2   0
##          W   0   0   0   0   0   1 193   0   0   0
##          X   0   0   0   0   0   0   0 193   1   0
##          Y   0   0   0   1   0   1   1   1 167   0
##          Z   0   0   0   0   0   0   0   0   0 188
```

```
##
## Overall Statistics
##
##                Accuracy : 0.9544
##                  95% CI : (0.9482, 0.96)
##     No Information Rate : 0.0456
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9526
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E Class: F
## Sensitivity            0.9899   0.9492   0.9637   0.9791   0.9423   0.9309
## Specificity            0.9994   0.9956   0.9992   0.9963   0.9979   0.9969
## Pos Pred Value         0.9849   0.8990   0.9789   0.9122   0.9515   0.9211
## Neg Pred Value         0.9996   0.9979   0.9985   0.9992   0.9975   0.9973
## Prevalence             0.0396   0.0394   0.0386   0.0382   0.0416   0.0376
## Detection Rate         0.0392   0.0374   0.0372   0.0374   0.0392   0.0350
## Detection Prevalence   0.0398   0.0416   0.0380   0.0410   0.0412   0.0380
## Balanced Accuracy      0.9946   0.9724   0.9814   0.9877   0.9701   0.9639
##                      Class: G Class: H Class: I Class: J Class: K Class: L
## Sensitivity            0.9403   0.8548   0.9794   0.9243   0.9181   0.9744
## Specificity            0.9969   0.9958   0.9969   0.9990   0.9983   0.9996
## Pos Pred Value         0.9265   0.8883   0.9268   0.9716   0.9515   0.9896
## Neg Pred Value         0.9975   0.9944   0.9992   0.9971   0.9971   0.9990
## Prevalence             0.0402   0.0372   0.0388   0.0370   0.0342   0.0390
## Detection Rate         0.0378   0.0318   0.0380   0.0342   0.0314   0.0380
## Detection Prevalence   0.0408   0.0358   0.0410   0.0352   0.0330   0.0384
## Balanced Accuracy      0.9686   0.9253   0.9881   0.9616   0.9582   0.9870
##                      Class: M Class: N Class: O Class: P Class: Q Class: R
## Sensitivity            0.9693   0.9330   0.9747   0.9189   0.9477   0.9482
## Specificity            0.9998   0.9990   0.9977   0.9983   0.9986   0.9956
## Pos Pred Value         0.9955   0.9709   0.9461   0.9551   0.9588   0.8971
## Neg Pred Value         0.9985   0.9975   0.9990   0.9969   0.9981   0.9979
## Prevalence             0.0456   0.0358   0.0396   0.0370   0.0344   0.0386
## Detection Rate         0.0442   0.0334   0.0386   0.0340   0.0326   0.0366
## Detection Prevalence   0.0444   0.0344   0.0408   0.0356   0.0340   0.0408
## Balanced Accuracy      0.9845   0.9660   0.9862   0.9586   0.9731   0.9719
##                      Class: S Class: T Class: U Class: V Class: W Class: X
## Sensitivity            0.9609   0.9681   0.9858   0.9691   0.9797   0.9602
## Specificity            0.9994   0.9998   0.9992   0.9971   0.9994   0.9987
## Pos Pred Value         0.9829   0.9945   0.9812   0.9307   0.9847   0.9698
## Neg Pred Value         0.9985   0.9988   0.9994   0.9987   0.9992   0.9983
## Prevalence             0.0358   0.0376   0.0424   0.0388   0.0394   0.0402
## Detection Rate         0.0344   0.0364   0.0418   0.0376   0.0386   0.0386
## Detection Prevalence   0.0350   0.0366   0.0426   0.0404   0.0392   0.0398
## Balanced Accuracy      0.9801   0.9839   0.9925   0.9831   0.9895   0.9795
##                      Class: Y Class: Z
## Sensitivity            0.9766   0.9592
## Specificity            0.9990   0.9994
## Pos Pred Value         0.9709   0.9843
## Neg Pred Value         0.9992   0.9983
```

```
## Prevalence             0.0342   0.0392
## Detection Rate         0.0334   0.0376
## Detection Prevalence   0.0344   0.0382
## Balanced Accuracy      0.9878   0.9793
```

improved Knn. For the problem of nearest neighbor classification, a simpler approach called "leave-out-one" cross-validation can be used, and this is provided by the knn.cv function. Using this technique, the observation itself is ignored when looking for its neighbors.

```
myknn<-knn.cv(x, y, k = 3, l = 0, prob = FALSE, use.all = TRUE)
confusionMatrix(myknn, y)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   A   B   C   D   E   F   G   H   I   J   K   L   M   N   O   P
##          A 784   0   0   0   0   0   0   1   0   2   0   0   1   1   0   0
##          B   0 728   0   5   1   4   5  15   0   0   1   0   5   0   1   3
##          C   1   0 712   0   5   0   3   1   0   0   0   0   0   0   2   0
##          D   0   1   0 775   0   0   6  20   0   0   0   0   0   4   8   2
##          E   0   3   4   0 731   1   9   1   0   0   6   5   0   0   0   0
##          F   0   0   0   0   2 730   0   1   0   1   0   0   0   0   0  24
##          G   0   1   7   0   7   0 738   5   0   0   1   4   2   0   0   0
##          H   0   2   2  12   0   1   1 651   0   1  18   1   1   6   0   4
##          I   0   0   0   0   0   1   0   0 721  24   0   0   0   0   0   1
##          J   0   0   0   0   0   2   0   0  29 712   0   1   0   0   0   0
##          K   0   0   0   0   3   0   0  15   0   0 685   1   0   0   0   0
##          L   1   0   1   0   3   0   0   0   0   0   0 745   0   1   0   1
##          M   0   0   0   0   0   0   0   1   0   0   0   0 770   1   0   0
##          N   0   0   0   0   0   2   0   2   1   1   0   0   1 751   1   0
##          O   0   0   3   7   0   0   4   2   0   0   0   0   1   5 728   0
##          P   0   0   0   1   1  18   0   0   0   0   0   0   0   1   0 757
##          Q   0   0   1   0   2   0   0   0   0   0   0   1   0   0   9   2
##          R   0  11   0   2   0   0   1  15   0   0  13   1   0  10   1   2
##          S   0   4   0   1   1   3   0   0   1   0   0   1   0   0   0   0
##          T   1   0   2   0   0  10   0   0   0   0   0   0   0   0   0   0
##          U   0   1   2   0   0   0   0   1   0   2   2   0   0   0   1   1
##          V   0  15   2   0   0   2   3   2   0   0   0   0   6   3   0   3
##          W   0   0   0   0   1   0   2   1   0   0   1   0   5   0   2   1
##          X   0   0   0   0   4   0   1   0   1   2  12   1   0   0   0   0
##          Y   1   0   0   0   0   1   0   0   1   0   0   0   0   0   0   2
##          Z   1   0   0   2   7   0   0   0   1   2   0   0   0   0   0   0
##           Reference
## Prediction   Q   R   S   T   U   V   W   X   Y   Z
##          A   0   0   1   0   1   2   0   1   1   0
##          B   0  10   1   5   1  12   0   1   1   0
##          C   0   0   0   1   1   0   0   0   0   0
##          D   0   2   0   1   1   0   0   2   0   0
##          E   0   0   2   0   0   0   0   9   0   4
##          F   0   0   1   2   0   2   0   0   0   0
##          G   4   1   0   0   0   2   0   0   0   0
##          H   0   8   1   0   3   1   0   1   0   0
##          I   0   0   0   0   0   0   0   0   0   0
```

```
##          J   0   0   0   1   1   0   0   0   0   1
##          K   0   6   1   0   0   0   1   5   0   0
##          L   1   1   0   0   0   0   0   0   0   1
##          M   0   0   0   0   4   1   2   0   1   0
##          N   0   5   0   0   0   1   1   0   0   0
##          O  19   0   0   0   0   0   3   2   0   0
##          P   0   0   0   0   0   3   0   0   0   0
##          Q 756   1   0   0   0   0   0   0   2   5
##          R   1 724   3   0   0   0   1   2   0   1
##          S   1   0 738   2   0   0   0   0   0   2
##          T   0   0   0 772   0   0   0   0   5   0
##          U   0   0   0   0 799   1   0   0   2   0
##          V   1   0   0   0   1 733   0   0  10   0
##          W   0   0   0   0   0   2 744   0   1   0
##          X   0   0   0   1   0   0   0 764   1   2
##          Y   0   0   0  10   1   4   0   0 762   0
##          Z   0   0   0   1   0   0   0   0   0 718
##
## Overall Statistics
##
##                Accuracy : 0.9614
##                  95% CI : (0.9586, 0.964)
##     No Information Rate : 0.0406
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.9599
##  Mcnemar's Test P-Value : NA
##
## Statistics by Class:
##
##                      Class: A Class: B Class: C Class: D Class: E Class: F
## Sensitivity           0.99366  0.95039   0.9674  0.96273  0.95182  0.94194
## Specificity           0.99943  0.99631   0.9993  0.99755  0.99771  0.99828
## Pos Pred Value        0.98616  0.91114   0.9807  0.94282  0.94323  0.95675
## Neg Pred Value        0.99974  0.99802   0.9988  0.99844  0.99808  0.99766
## Prevalence            0.03945  0.03830   0.0368  0.04025  0.03840  0.03875
## Detection Rate        0.03920  0.03640   0.0356  0.03875  0.03655  0.03650
## Detection Prevalence  0.03975  0.03995   0.0363  0.04110  0.03875  0.03815
## Balanced Accuracy     0.99655  0.97335   0.9833  0.98014  0.97477  0.97011
##                      Class: G Class: H Class: I Class: J Class: K Class: L
## Sensitivity           0.95472  0.88692  0.95497  0.95315  0.92693  0.97898
## Specificity           0.99823  0.99673  0.99865  0.99818  0.99834  0.99948
## Pos Pred Value        0.95596  0.91176  0.96519  0.95315  0.95537  0.98675
## Neg Pred Value        0.99818  0.99570  0.99823  0.99818  0.99720  0.99917
## Prevalence            0.03865  0.03670  0.03775  0.03735  0.03695  0.03805
## Detection Rate        0.03690  0.03255  0.03605  0.03560  0.03425  0.03725
## Detection Prevalence  0.03860  0.03570  0.03735  0.03735  0.03585  0.03775
## Balanced Accuracy     0.97648  0.94183  0.97681  0.97566  0.96263  0.98923
##                      Class: M Class: N Class: O Class: P Class: Q Class: R
## Sensitivity            0.9722  0.95913  0.96680  0.94271  0.96552   0.9551
## Specificity            0.9995  0.99922  0.99761  0.99875  0.99880   0.9967
## Pos Pred Value         0.9872  0.98042  0.94057  0.96927  0.97047   0.9188
## Neg Pred Value         0.9989  0.99834  0.99870  0.99761  0.99860   0.9982
## Prevalence             0.0396  0.03915  0.03765  0.04015  0.03915   0.0379
```

```
## Detection Rate         0.0385  0.03755  0.03640  0.03785  0.03780   0.0362
## Detection Prevalence   0.0390  0.03830  0.03870  0.03905  0.03895   0.0394
## Balanced Accuracy      0.9859  0.97918  0.98220  0.97073  0.98216   0.9759
##                      Class: S Class: T Class: U Class: V Class: W Class: X
## Sensitivity            0.9866   0.9698  0.98278  0.95942   0.9894  0.97078
## Specificity            0.9992   0.9991  0.99932  0.99750   0.9992  0.99870
## Pos Pred Value         0.9788   0.9772  0.98399  0.93854   0.9789  0.96831
## Neg Pred Value         0.9995   0.9988  0.99927  0.99839   0.9996  0.99880
## Prevalence             0.0374   0.0398  0.04065  0.03820   0.0376  0.03935
## Detection Rate         0.0369   0.0386  0.03995  0.03665   0.0372  0.03820
## Detection Prevalence   0.0377   0.0395  0.04060  0.03905   0.0380  0.03945
## Balanced Accuracy      0.9929   0.9845  0.99105  0.97846   0.9943  0.98474
##                      Class: Y Class: Z
## Sensitivity            0.9695   0.9782
## Specificity            0.9990   0.9993
## Pos Pred Value         0.9744   0.9809
## Neg Pred Value         0.9988   0.9992
## Prevalence             0.0393   0.0367
## Detection Rate         0.0381   0.0359
## Detection Prevalence   0.0391   0.0366
## Balanced Accuracy      0.9842   0.9887
```