

Neural_Network_bank_note

Nabil Momin

2024-06-12

```
library(corrgram)
library(corrplot)

## corrplot 0.92 loaded

library(caTools)
library(Amelia)

## Loading required package: Rcpp

## ##
## ## Amelia II: Multiple Imputation
## ## (Version 1.8.2, built: 2024-04-10)
## ## Copyright (C) 2005-2024 James Honaker, Gary King and Matthew Blackwell
## ## Refer to http://gking.harvard.edu/amelia/ for more information
## ##

library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(rpart)
library(rpart.plot)
library(randomForest)

## randomForest 4.7-1.1

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'
```

```

## The following object is masked from 'package:dplyr':
##
##      combine

## The following object is masked from 'package:ggplot2':
##
##      margin

library(ISLR)
library(e1071)
library(cluster)
library(tm)

## Loading required package: NLP

##
## Attaching package: 'NLP'

## The following object is masked from 'package:ggplot2':
##
##      annotate

library(twitterR)

##
## Attaching package: 'twitterR'

## The following objects are masked from 'package:dplyr':
##
##      id, location

library(wordcloud)

## Loading required package: RColorBrewer

library(RColorBrewer)
library(neuralnet)

##
## Attaching package: 'neuralnet'

## The following object is masked from 'package:dplyr':
##
##      compute

#### getting the data

df <- read.csv('bank_note_data.csv')

head(df)

##      Image.Var Image.Skew Image.Curt Entropy Class
## 1      3.62160      8.6661      -2.8073 -0.44699      0

```

```
## 2  4.54590      8.1674    -2.4586 -1.46210      0
## 3  3.86600     -2.6383     1.9242  0.10645      0
## 4  3.45660     9.5228     -4.0112 -3.59440      0
## 5  0.32924     -4.4552     4.5718 -0.98880      0
## 6  4.36840     9.6718     -3.9606 -3.16250      0
```

```
View(df)
```

```
#### normalize the data
```

```
var(df[,3])
```

```
## [1] 18.57636
```

```
var(df[,4])
```

```
## [1] 4.414256
```

```
standard.df <- scale(df[1:4])
```

```
var(standard.df[,3])
```

```
## [1] 1
```

```
var(standard.df[,4])
```

```
## [1] 1
```

```
#### adding the last column back in
```

```
new.df <- cbind(standard.df,df[5])
```

```
View(new.df)
```

```
#### train and test
```

```
sample <- sample.split(new.df,SplitRatio = 0.7)
```

```
train <- subset(new.df,sample == TRUE)
```

```
test <- subset(new.df, sample == FALSE)
```

```
#### making model
```

```
model <- neuralnet(Class ~ Image.Var + Image.Skew + Image.Curt +  
Entropy,train, hidden=c(5,3),  
linear.output = FALSE)
```

```
predict.model <- compute(model,test[1:4])
```

```
head(predict.model$net.result)
```

```

##           [,1]
## 2  0.0007897753
## 3  0.0008783631
## 7  0.0007901702
## 8  0.0008577052
## 12 0.0008442159
## 13 0.0007891577

#### rounding to 0 and 1

prediction <- sapply(predict.model$net.result,round)

head(prediction)

## [1] 0 0 0 0 0 0

table(prediction)

## prediction
##    0    1
## 305 244

#### confusion matrix

table(prediction,test$Class)

##
## prediction    0    1
##           0 305    0
##           1   0 244

#### seems too good to be true as its coming out almost without any error or
false positives or false negatives
#### quickly using the random forest to check the accuracy of the neural
network result

df <- read.csv('bank_note_data.csv')

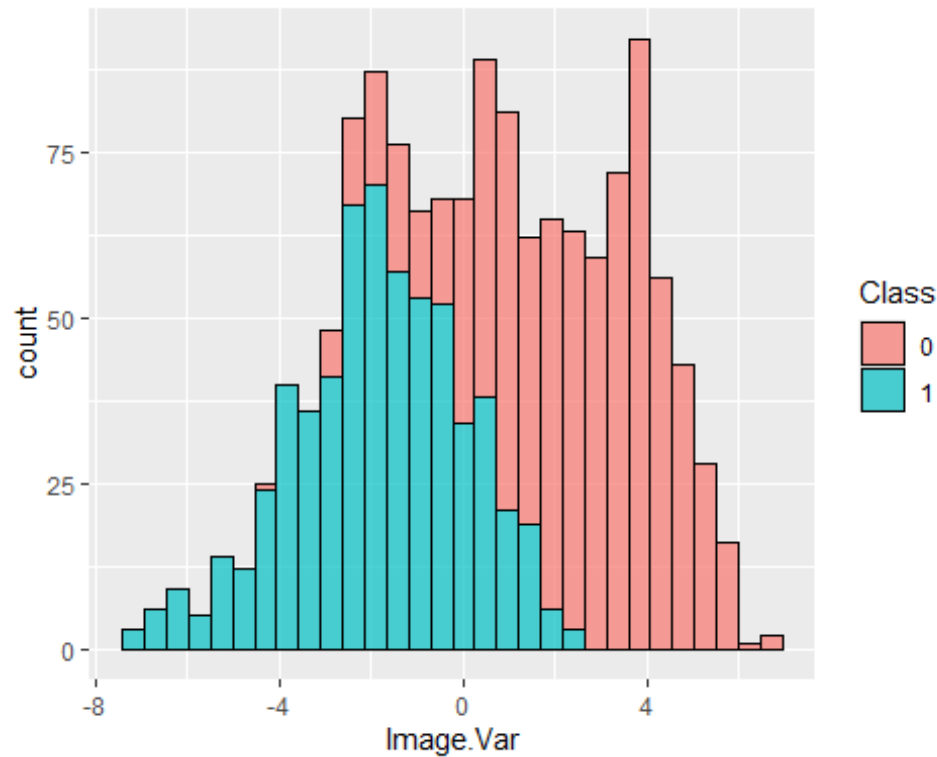
df$Class <- factor(df$Class)

#### EDA time

ggplot(df,aes(Image.Var)) + geom_histogram(aes(fill =
Class),color='black',alpha=0.7)

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

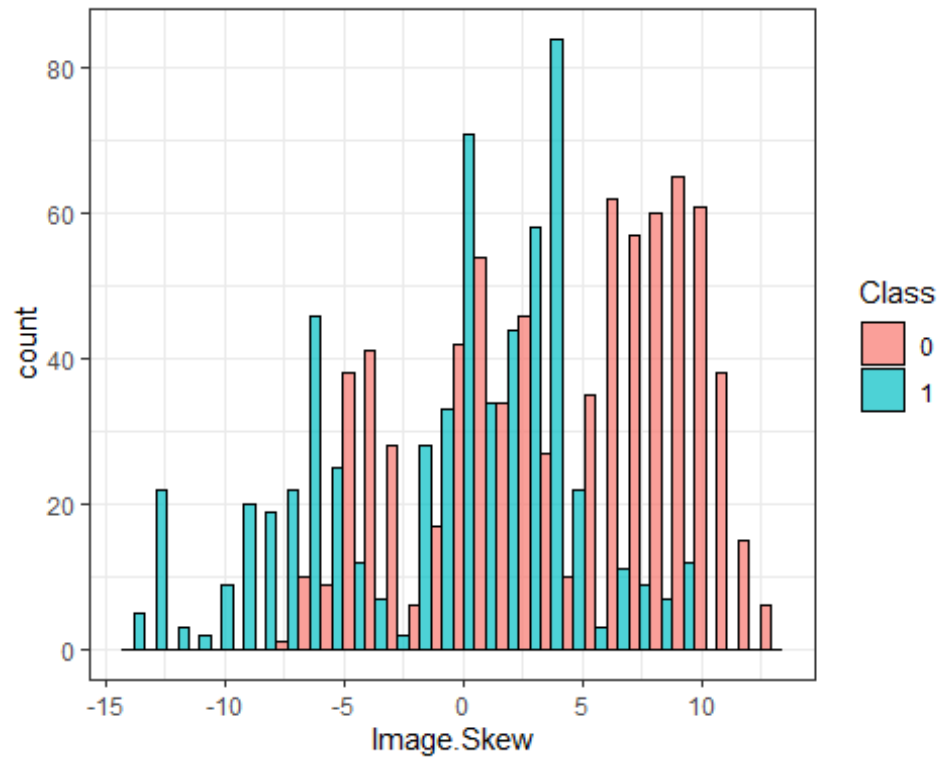
```



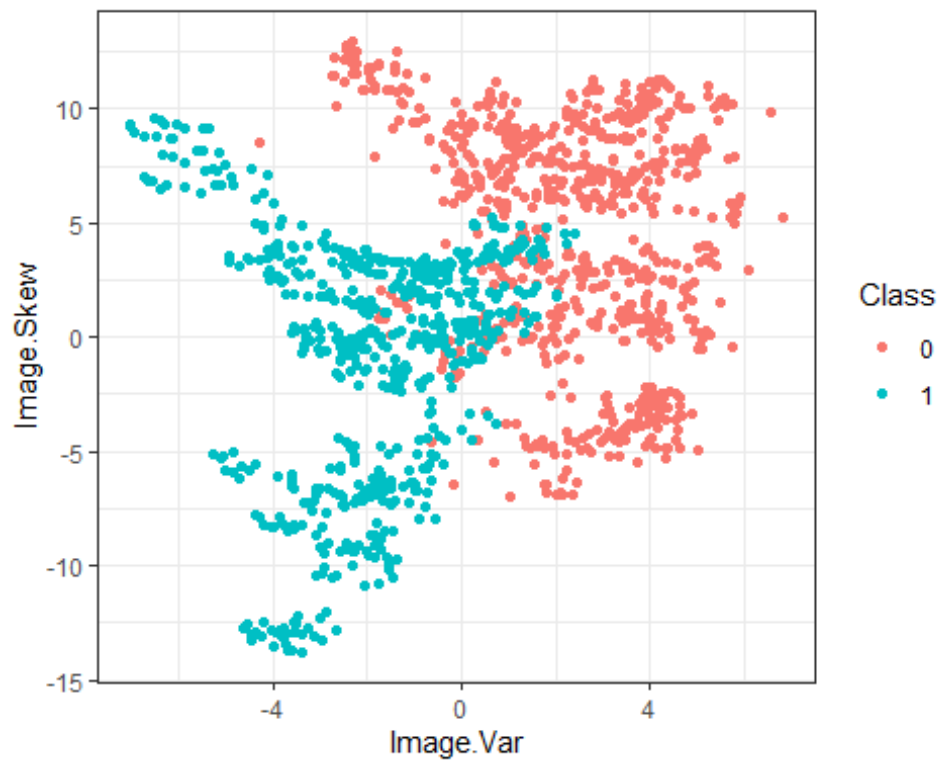
we see from the data plotted above that the image variance really matters to confirm the bank note is real or not, which honestly is not a surprise

```
ggplot(df,aes(Image.Skew)) + geom_histogram(aes(fill =  
Class),color='black',alpha=0.7,position = 'dodge') + theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
ggplot(df,aes(Image.Var,Image.Skew)) + geom_point(aes(color=Class)) +  
theme_bw()
```



the scatter plot clearly shows the higher the variance or skew the higher is the probability of the bill being fake

```
str(df)
```

```
## 'data.frame': 1372 obs. of 5 variables:
## $ Image.Var : num 3.622 4.546 3.866 3.457 0.329 ...
## $ Image.Skew: num 8.67 8.17 -2.64 9.52 -4.46 ...
## $ Image.Curt: num -2.81 -2.46 1.92 -4.01 4.57 ...
## $ Entropy : num -0.447 -1.462 0.106 -3.594 -0.989 ...
## $ Class : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
```

```
sample <- sample.split(df$Class, SplitRatio = 0.7)
```

```
train <- subset(df, sample == TRUE)
```

```
test <- subset(df, sample == FALSE)
```

```
model <- randomForest(Class ~., train)
```

```
table(model$predicted)
```

```
##
## 0 1
## 531 429
```

```
predict.model <- predict(model, test[1:4])
```

```
table(predict.model, test$Class)
```

```
##
## predict.model 0 1
## 0 225 1
## 1 4 182
```

```
table.predict <- cbind(predict.model, test[5])
```

```
print(table.predict)
```

```
##      predict.model Class
## 1           0      0
## 2           0      0
## 6           0      0
## 8           0      0
## 10          0      0
## 14          0      0
## 15          0      0
## 21          0      0
## 22          0      0
## 24          0      0
## 26          0      0
```

## 28	0	0
## 31	1	0
## 35	0	0
## 39	0	0
## 41	0	0
## 43	0	0
## 63	0	0
## 67	0	0
## 73	0	0
## 74	0	0
## 81	0	0
## 82	0	0
## 85	0	0
## 88	0	0
## 100	0	0
## 102	0	0
## 108	0	0
## 112	0	0
## 114	0	0
## 125	0	0
## 132	0	0
## 134	0	0
## 136	0	0
## 147	0	0
## 152	0	0
## 162	0	0
## 165	0	0
## 166	0	0
## 174	0	0
## 179	0	0
## 192	0	0
## 194	0	0
## 197	0	0
## 204	0	0
## 208	0	0
## 213	0	0
## 230	0	0
## 233	0	0
## 235	0	0
## 237	0	0
## 238	0	0
## 241	0	0
## 243	0	0
## 250	0	0
## 255	0	0
## 256	0	0
## 257	0	0
## 259	0	0
## 264	0	0
## 266	0	0

## 269	0	0
## 270	0	0
## 271	0	0
## 274	0	0
## 276	0	0
## 279	0	0
## 284	0	0
## 286	0	0
## 291	0	0
## 293	0	0
## 295	0	0
## 298	0	0
## 303	0	0
## 309	0	0
## 311	0	0
## 313	0	0
## 317	0	0
## 320	0	0
## 322	0	0
## 325	0	0
## 328	0	0
## 331	0	0
## 334	0	0
## 336	0	0
## 342	0	0
## 343	0	0
## 346	0	0
## 350	1	0
## 352	0	0
## 354	0	0
## 355	0	0
## 356	0	0
## 363	0	0
## 364	0	0
## 367	0	0
## 368	0	0
## 371	0	0
## 381	0	0
## 388	0	0
## 389	0	0
## 390	0	0
## 392	0	0
## 393	0	0
## 395	0	0
## 397	0	0
## 399	0	0
## 402	0	0
## 403	0	0
## 407	0	0
## 411	0	0

## 412	0	0
## 416	0	0
## 418	0	0
## 419	0	0
## 420	0	0
## 426	0	0
## 427	0	0
## 431	0	0
## 432	0	0
## 437	0	0
## 439	0	0
## 440	0	0
## 449	0	0
## 452	0	0
## 454	0	0
## 459	0	0
## 460	0	0
## 462	0	0
## 466	0	0
## 467	0	0
## 468	0	0
## 469	0	0
## 471	0	0
## 478	0	0
## 491	0	0
## 493	0	0
## 500	0	0
## 503	0	0
## 507	0	0
## 511	0	0
## 516	0	0
## 517	0	0
## 518	0	0
## 520	0	0
## 523	0	0
## 524	0	0
## 533	0	0
## 538	0	0
## 542	0	0
## 543	0	0
## 546	0	0
## 548	0	0
## 549	0	0
## 550	0	0
## 553	0	0
## 555	0	0
## 556	0	0
## 558	0	0
## 560	0	0
## 562	0	0

## 570	0	0
## 572	0	0
## 573	0	0
## 574	0	0
## 576	0	0
## 577	0	0
## 582	0	0
## 583	0	0
## 584	0	0
## 591	0	0
## 593	0	0
## 594	1	0
## 599	0	0
## 602	0	0
## 614	0	0
## 621	0	0
## 622	0	0
## 624	0	0
## 626	0	0
## 627	0	0
## 638	0	0
## 643	0	0
## 645	0	0
## 648	0	0
## 651	0	0
## 653	0	0
## 656	0	0
## 657	0	0
## 658	0	0
## 665	0	0
## 666	0	0
## 670	0	0
## 672	0	0
## 675	0	0
## 676	0	0
## 679	0	0
## 680	0	0
## 681	0	0
## 682	0	0
## 687	0	0
## 692	0	0
## 693	0	0
## 694	0	0
## 695	0	0
## 696	0	0
## 699	0	0
## 700	0	0
## 702	0	0
## 710	0	0
## 711	0	0

## 718	0	0
## 720	0	0
## 722	0	0
## 726	0	0
## 727	0	0
## 728	0	0
## 730	1	0
## 732	0	0
## 734	0	0
## 735	0	0
## 736	0	0
## 737	0	0
## 738	0	0
## 740	0	0
## 743	0	0
## 745	0	0
## 746	0	0
## 759	0	0
## 770	1	1
## 772	1	1
## 773	1	1
## 779	1	1
## 784	1	1
## 785	1	1
## 795	1	1
## 797	1	1
## 798	1	1
## 799	1	1
## 807	1	1
## 810	1	1
## 811	1	1
## 813	1	1
## 816	1	1
## 823	1	1
## 827	1	1
## 828	1	1
## 831	1	1
## 834	1	1
## 839	1	1
## 840	1	1
## 841	1	1
## 844	1	1
## 848	1	1
## 849	1	1
## 852	1	1
## 856	1	1
## 864	1	1
## 866	1	1
## 867	1	1
## 868	1	1

## 870	1	1
## 875	1	1
## 876	1	1
## 878	1	1
## 879	1	1
## 882	1	1
## 893	1	1
## 895	1	1
## 897	1	1
## 905	1	1
## 908	1	1
## 909	1	1
## 914	1	1
## 915	1	1
## 916	1	1
## 919	1	1
## 922	1	1
## 924	1	1
## 925	1	1
## 929	1	1
## 935	1	1
## 948	1	1
## 952	1	1
## 953	1	1
## 955	1	1
## 958	1	1
## 963	1	1
## 964	1	1
## 967	1	1
## 974	1	1
## 975	1	1
## 978	1	1
## 981	0	1
## 989	1	1
## 997	1	1
## 998	1	1
## 999	1	1
## 1001	1	1
## 1002	1	1
## 1003	1	1
## 1007	1	1
## 1008	1	1
## 1022	1	1
## 1023	1	1
## 1027	1	1
## 1030	1	1
## 1031	1	1
## 1037	1	1
## 1038	1	1
## 1044	1	1

## 1048	1	1
## 1052	1	1
## 1056	1	1
## 1060	1	1
## 1065	1	1
## 1070	1	1
## 1071	1	1
## 1073	1	1
## 1075	1	1
## 1083	1	1
## 1085	1	1
## 1087	1	1
## 1090	1	1
## 1098	1	1
## 1100	1	1
## 1103	1	1
## 1111	1	1
## 1113	1	1
## 1118	1	1
## 1119	1	1
## 1120	1	1
## 1121	1	1
## 1123	1	1
## 1127	1	1
## 1128	1	1
## 1129	1	1
## 1130	1	1
## 1133	1	1
## 1138	1	1
## 1139	1	1
## 1145	1	1
## 1147	1	1
## 1154	1	1
## 1160	1	1
## 1161	1	1
## 1162	1	1
## 1168	1	1
## 1169	1	1
## 1171	1	1
## 1172	1	1
## 1175	1	1
## 1176	1	1
## 1179	1	1
## 1181	1	1
## 1186	1	1
## 1190	1	1
## 1192	1	1
## 1195	1	1
## 1196	1	1
## 1199	1	1

## 1201	1	1
## 1207	1	1
## 1208	1	1
## 1209	1	1
## 1211	1	1
## 1221	1	1
## 1225	1	1
## 1227	1	1
## 1228	1	1
## 1232	1	1
## 1240	1	1
## 1241	1	1
## 1242	1	1
## 1243	1	1
## 1244	1	1
## 1246	1	1
## 1250	1	1
## 1256	1	1
## 1266	1	1
## 1271	1	1
## 1274	1	1
## 1278	1	1
## 1283	1	1
## 1285	1	1
## 1286	1	1
## 1289	1	1
## 1293	1	1
## 1294	1	1
## 1297	1	1
## 1300	1	1
## 1305	1	1
## 1307	1	1
## 1312	1	1
## 1314	1	1
## 1318	1	1
## 1320	1	1
## 1323	1	1
## 1326	1	1
## 1328	1	1
## 1334	1	1
## 1336	1	1
## 1344	1	1
## 1349	1	1
## 1351	1	1
## 1358	1	1
## 1359	1	1
## 1363	1	1
## 1364	1	1
## 1367	1	1

```
## 1369      1      1
## 1372      1      1
```

even the random forest is pretty accurate in this case