PCA and LDA

Code

Jovanni Ochoa

Run PCA on the bank data

c(1, 6, 8, 12, 13)]

Hide

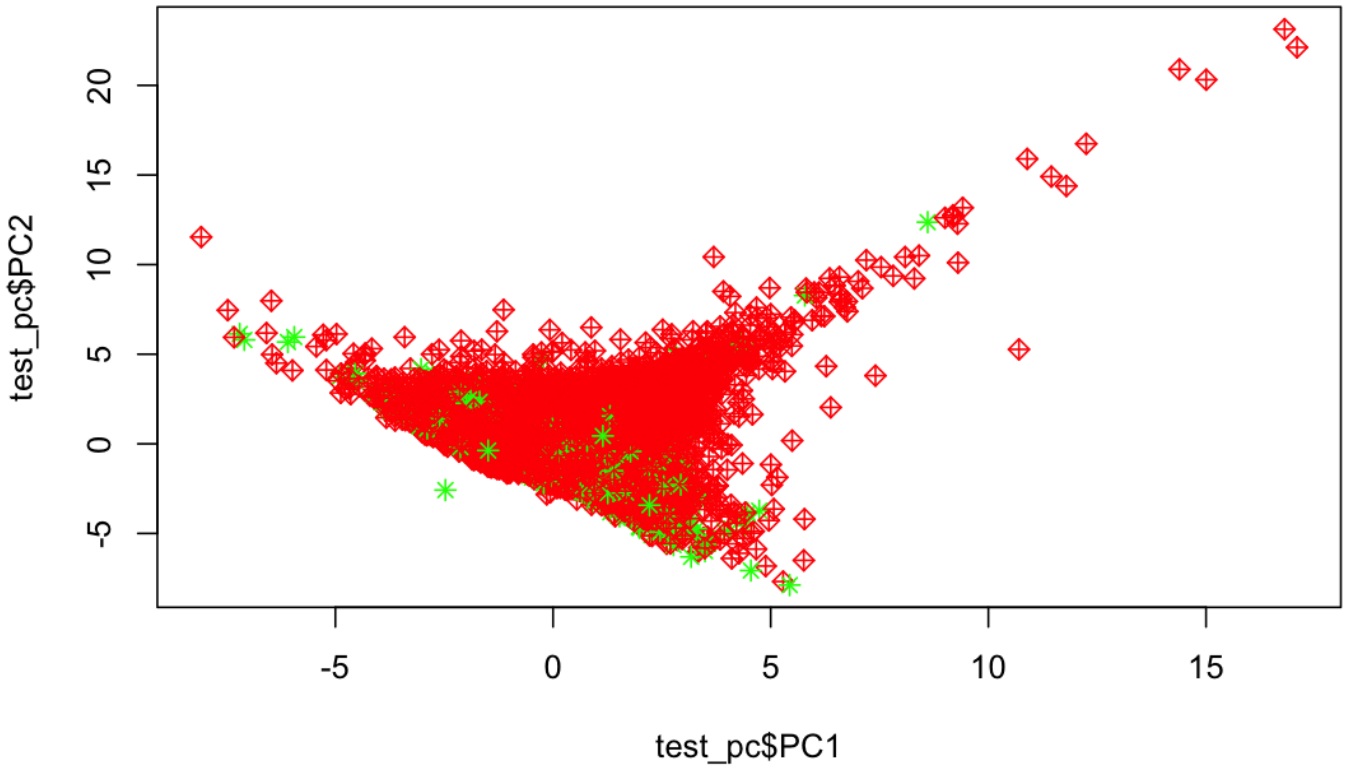
```
library(stats)

# Load data
bank <- read.csv("desktop/bank-full.csv", header=TRUE)

i <- sample(1:15000, 10000, replace=FALSE)
train <- bank[i,]
test <- bank[-i,]

# Apply pre-processing to remaining columns
set.seed(1234)
pca_out <- preProcess(train[c(1, 6, 8, 12, 13)], method=c("center", "scale", "pca"))
pca_out
```

```
Created from 10000 samples and 5 variables

Pre-processing:
  - centered (4)
  - ignored (1)
  - principal component signal extraction (4)
  - scaled (4)

PCA needed 4 components to capture 95 percent of the variance
```

PCA plot

Hide

```
train_pc <- predict(pca_out, train[c(1, 6, 8, 12, 13)])
test_pc <- predict(pca_out, test[,])
plot(test_pc$PC1, test_pc$PC2, pch=ifelse(test_pc$loan == "yes", 8, 9), col=ifelse(test_
pc$loan == "yes", "green", "red"))
```



PCA data in knn

Now let's see if our two principal components can predict class.

Hide

```
train_df <- data.frame(train_pc$PC1, train_pc$PC2, train$loan)
test_df <- data.frame(test_pc$PC1, test_pc$PC2, test$loan)
library(class)
set.seed(1234)
pred <- knn(train=train_df[,1:2], test=test_df[,1:2], cl=train_df[,3], k=3)
mean(pred==test$loan)
```

```
[1] 0.78484
```

With the decison tree we got a little lower accuracy.

LDA

Hide

```
library(MASS)

#check the levels of each column
sapply(test, function(x) length(unique(x)))
```

```
      age       job   marital education   default   balance   housing      loan   contac
t       day     month
       77        12         3         4         2      6534         2         2
3        31        12
 duration  campaign     pdays  previous  poutcome    Target
     1493        41       559        41         4         2
```

Hide

```
lda1 <- lda(loan ~ job + marital + education + default + balance + housing + day + month
+ duration + campaign, data = test)
lda1$means
```

```
    jobblue-collar jobentrepreneur jobhousemaid jobmanagement jobretired jobself-employe
d jobservices  jobstudent
no       0.1878914      0.02846332   0.02792754     0.2322272 0.05588856       0.0373036
9  0.08107022 0.028161940
yes      0.2155946      0.05142109   0.02056844     0.1851159 0.04244577       0.0330964
8  0.10976066 0.001869858
    jobtechnician jobunemployed   jobunknown maritalmarried maritalsingle educationsecon
dary educationtertiary
no      0.1724542    0.03348625 0.0066637645      0.5869805     0.3024478          0.487
3924         0.3306433
yes     0.1851159    0.01645475 0.0005609574      0.6389304     0.2324233          0.589
1922         0.2638369
    educationunknown defaultyes   balance housingyes      day  monthaug    monthdec   mo
nthfeb   monthjan  monthjul
no        0.04259451 0.01208854 1541.8516  0.5029970 16.04735 0.1889629 0.006797710 0.07
564545 0.03988213 0.1251381
yes       0.01608078 0.04188482  816.5361  0.5946148 16.73859 0.1129394 0.002056844 0.07
292446 0.03964099 0.2683246
      monthjun    monthmar  monthmay  monthnov   monthoct    monthsep duration campaign
no  0.06857985 0.015102301 0.2438134 0.1086294 0.02287111 0.018383953 260.6689 2.719854
yes 0.05703067 0.004861631 0.2236350 0.1357517 0.01028422 0.005609574 247.0150 2.887809
```

predict on test

Hide

```
### predict on test
lda_pred <- predict(lda1, newdata=test, type="class")
lda_pred$class
```

```
   [1] no  no  no  no  no  yes no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
  [29] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
  [57] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
  [85] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [113] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [141] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [169] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [197] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  yes no  no  n
o  no  no  no  no  no  no  no
 [225] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [253] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [281] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [309] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [337] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [365] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [393] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [421] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [449] no  no  no  no  yes no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [477] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [505] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [533] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [561] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [589] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [617] no  no  no  no  no  no  no  no  no  yes no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [645] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [673] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  yes no  no  no  no  n
o  no  no  no  no  no  no  no
 [701] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
```

```
 [729] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [757] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [785] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [813] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [841] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [869] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [897] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [925] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [953] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  n
o  no  no  no  no  no  no  no
 [981] no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no  no
 [ reached getOption("max.print") -- omitted 34211 entries ]
Levels: no yes
```

Hide

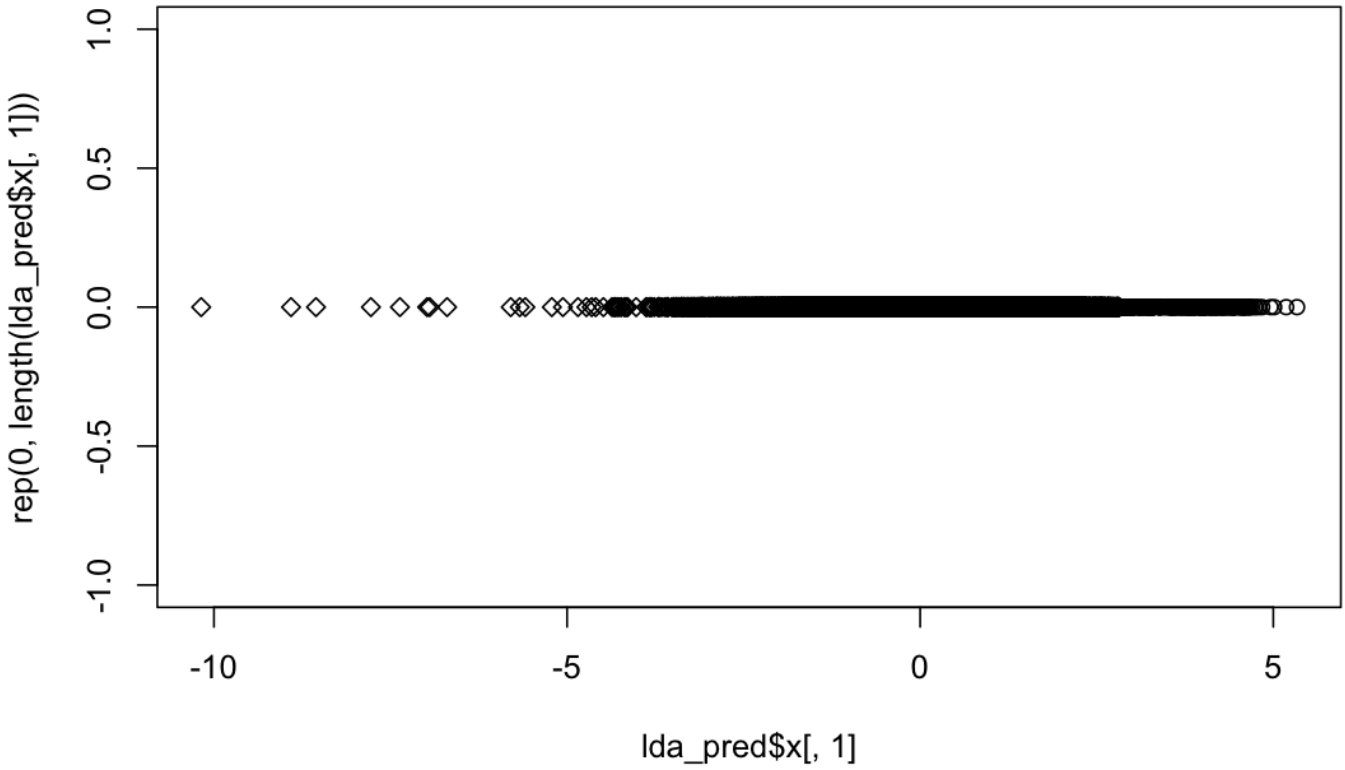
```
mean(lda_pred$class==test$loan)
```

```
[1] 0.84786
```

plot

Hide

```
plot(lda_pred$x[,1], rep(0,length(lda_pred$x[,1])), pch=c(23,21)[unclass(lda_pred$clas
s)], bg=c("red","green")[unclass(test$loan)])
```


rep(0, length(lda_pred$x[, 1]))
1.0
0.5
0.0
-0.5
-1.0
-10
-5
0
5
lda_pred$x[, 1]