

Banknotes with K-Means

Milica Cudina

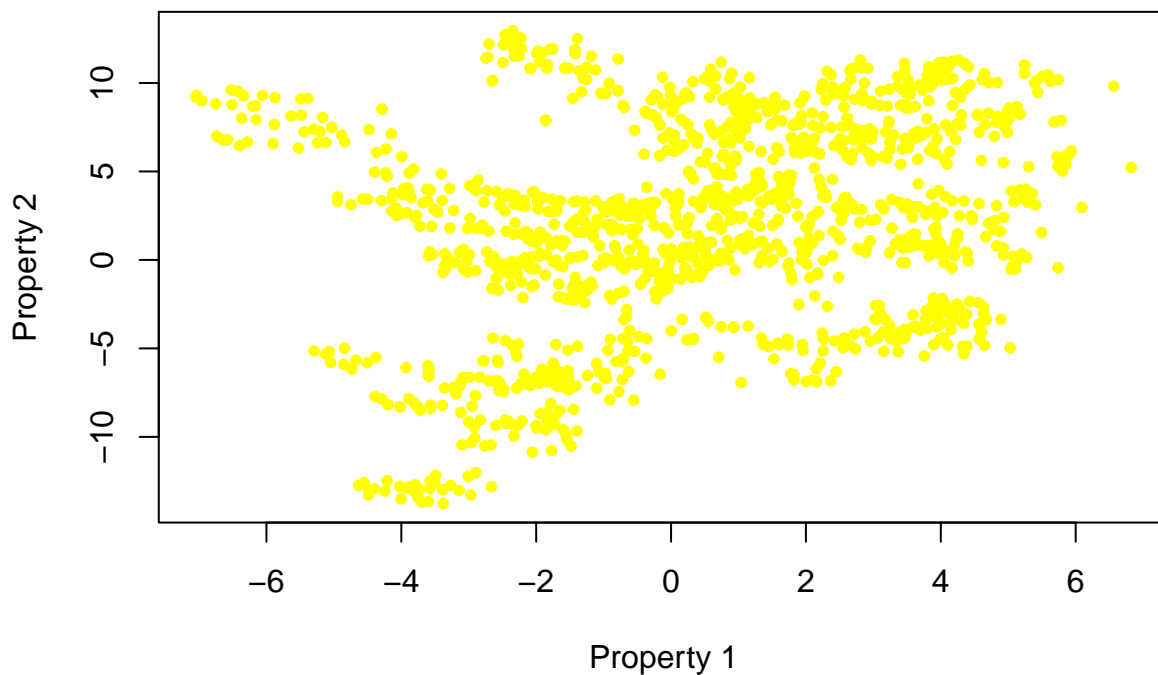
First, we need to import data:

```
notes<-read.csv("banknotes.csv")
head(notes)
```

```
##      V1      V2
## 1 3.62160 8.6661
## 2 4.54590 8.1674
## 3 3.86600 -2.6383
## 4 3.45660 9.5228
## 5 0.32924 -4.4552
## 6 4.36840 9.6718
```

Now, a bit of exploratory data analysis:

```
attach(notes)
plot(V1, V2,
     col="yellow", pch=20,
     xlab="Property 1",
     ylab="Property 2")
```



The function `kmeans()` performs *K*-means clustering in R.

```
km.out <- kmeans(notes, 2, nstart = 20)
km.out
```

```

## K-means clustering with 2 clusters of sizes 589, 783
##
## Cluster means:
##      V1      V2
## 1 -0.1415258 -3.560442
## 2  0.8664667  6.046704
##
## Clustering vector:
##  [1] 2 2 1 2 1 2 2 1 2 2 2 1 2 2 2 1 2 2 1 2 2 2 1 1 2 2 1 1 2 2 2 1 2 2
## [38] 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 1 2 2 1 2 2 1 2 2 2 1 1 2 1
## [75] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 1 2 1 2 1 2 2
## [ reached 'max' / getOption("max.print") -- omitted 1272 entries ]
##
## Within cluster sum of squares by cluster:
## [1] 12687.54 14250.46
## (between_SS / total_SS =  53.8 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
#here are the cluster assignments
km.out$cluster

##  [1] 2 2 1 2 1 2 2 1 2 2 2 1 2 2 2 1 2 2 1 2 2 2 1 1 2 2 1 1 2 2 2 1 2 2
## [38] 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 1 2 2 1 2 2 1 2 2 2 1 1 2 1
## [75] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 1 2 1 2 1 2 1 2 2
## [ reached 'max' / getOption("max.print") -- omitted 1272 entries ]
#here are the cluster centers
km.out$centers

##      V1      V2
## 1 -0.1415258 -3.560442
## 2  0.8664667  6.046704

```

It all works out much better visually.

```

#cloud of points with colors corresponding to cluster
plot(notes, col = (km.out$cluster + 1),
     main = "K-Means Clustering Results with K = 2",
     xlab = "Property 1", ylab = "Property 2", pch = 20, cex = 0.5)
#adding the centers
points(km.out$centers, col="blue", pch=20)

```

K-Means Clustering Results with K = 2

