

Лабораторная работа №5. Кластерный анализ. Алгоритм K-means

Студент:

Dubovskij Pan

Вариант

13

Чтение данных из файла

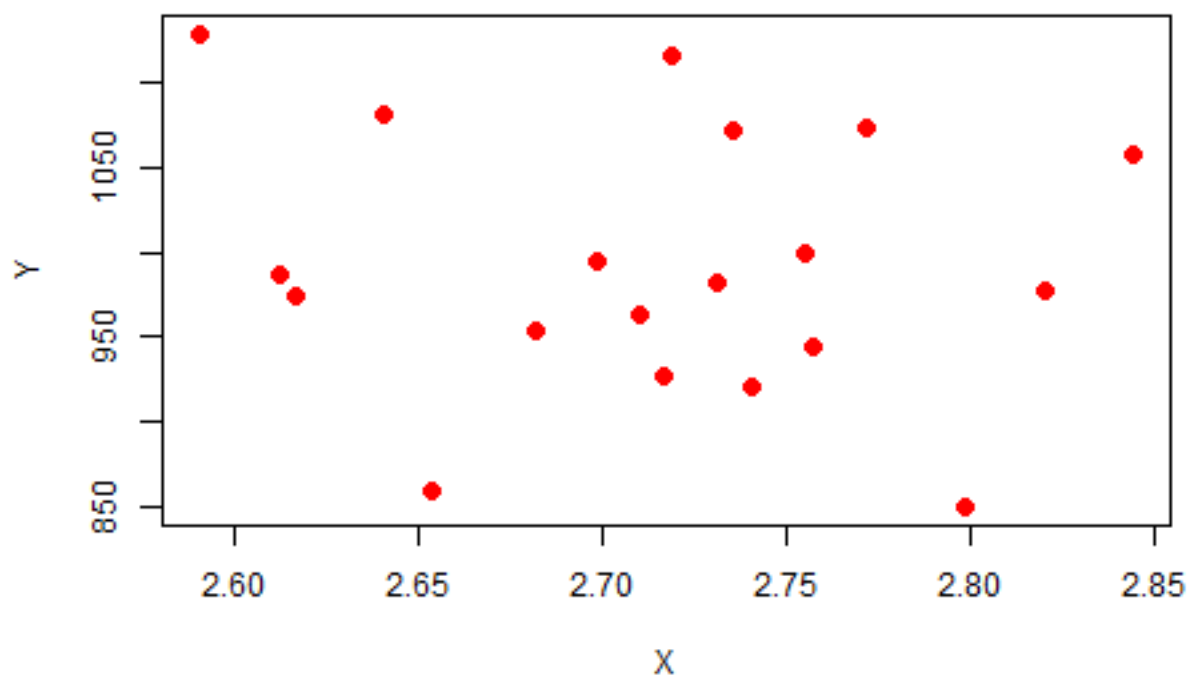
```
dat = read.table(file = "input.txt", dec = ",")
dat
```

```
##      V1      V2
## 1  2.6540  859.0
## 2  2.7982  850.4
## 3  2.7166  926.8
## 4  2.7406  921.0
## 5  2.7575  943.9
## 6  2.6818  953.7
## 7  2.6166  973.8
## 8  2.6123  987.5
## 9  2.7104  963.6
## 10 2.7554  999.7
## 11 2.6989  994.0
## 12 2.7187 1115.3
## 13 2.7309  981.6
## 14 2.8205  977.6
## 15 2.8443 1056.9
## 16 2.7721 1073.7
## 17 2.7353 1071.6
## 18 2.6405 1080.5
## 19 2.5905 1128.5
```

Диаграмма рассеяния

```
plot(dat, type = "p",
      main = "Диаграмма рассеивания",
      xlab = "X", ylab = "Y",
      col = "red", pch = 20, cex = 2)
```

Диаграмма рассеивания



Кластеризация на 2 кластера

```
#Число кластеров = 2  
cl = kmeans(scale(dat), 2)  
#Число наблюдений  
table(cl$cluster)
```

```
##  
## 1 2  
## 13 6
```

```
cl$centers
```

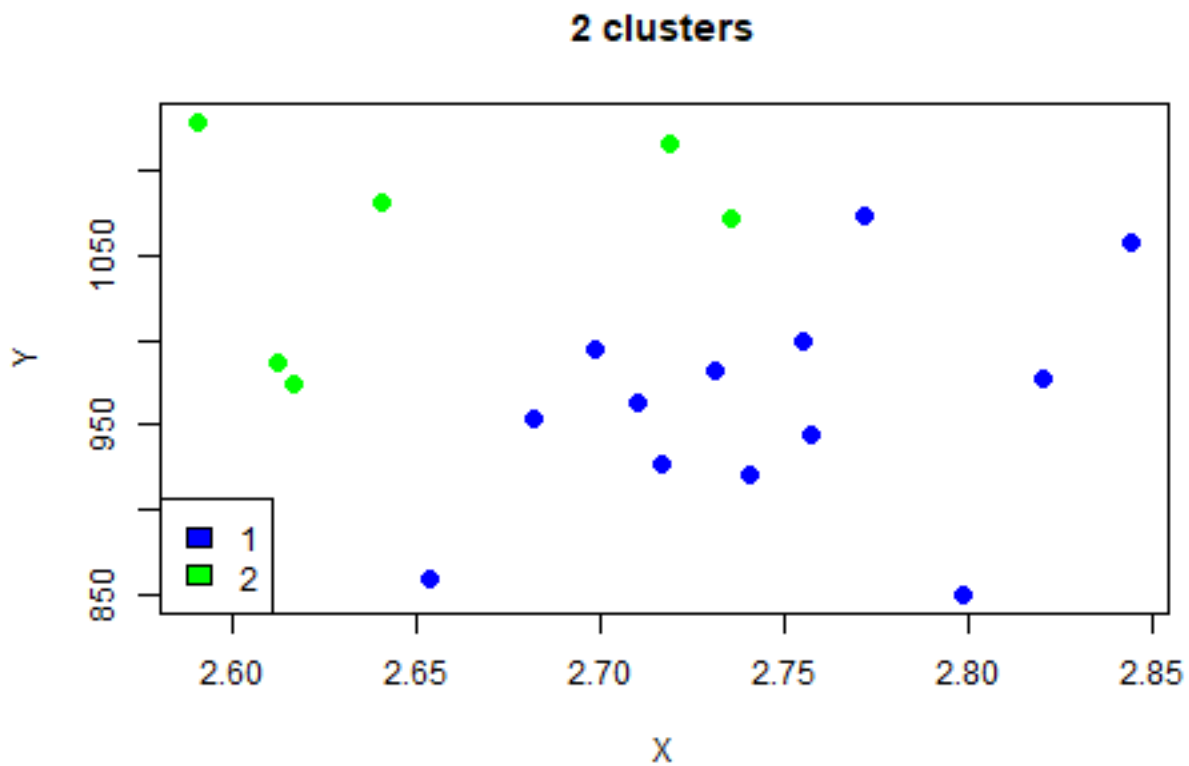
```
##          V1          V2  
## 1  0.4144938 -0.3930722  
## 2 -0.8980699  0.8516565
```

```
sapply(1:2, function(i) cl$centers[,i] * sd(dat[,i]) + mean(dat[,i]))
```

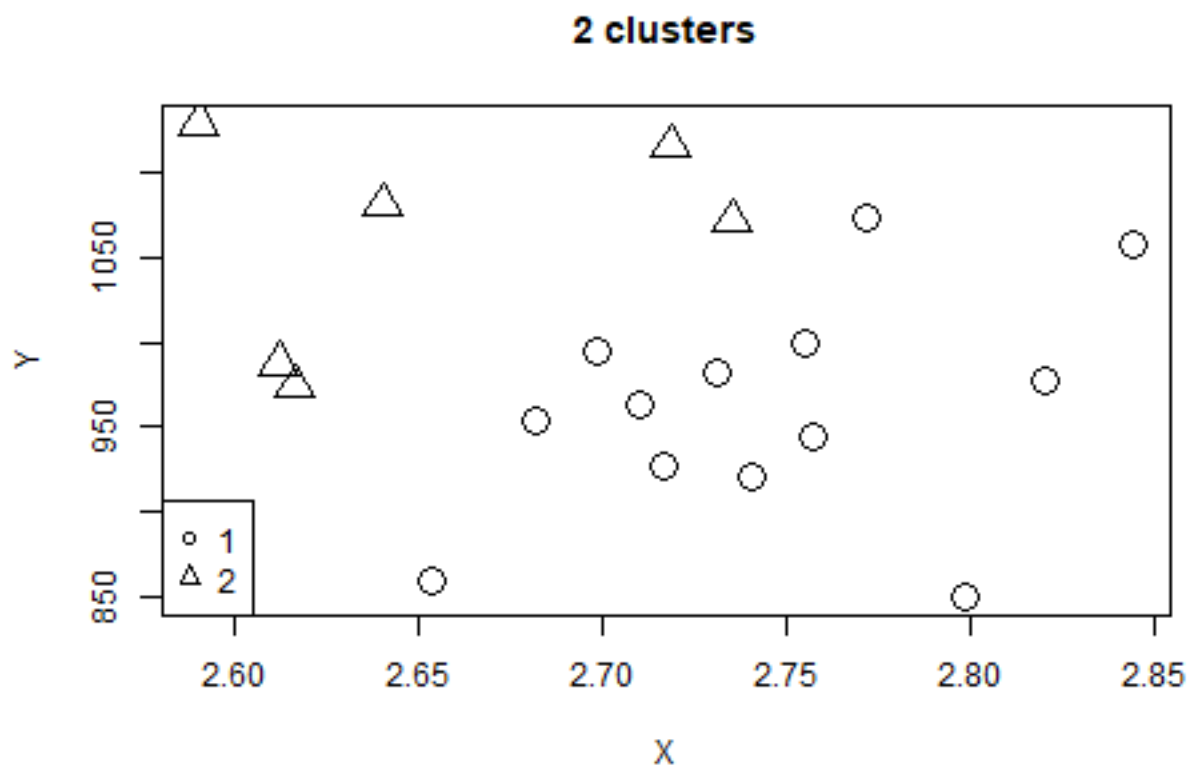
```
##          [,1]      [,2]  
## 1 2.744708  961.6846  
## 2 2.652317 1059.5333
```

Кластеры на графике рассеяния

```
plot(dat, col = c("blue", "green")[cl$cluster], pch = 20, cex = 2, xlab = "X", ylab = "Y", ma
legend("bottomleft", legend = c("1", "2"), fill = c("blue", "green"))
```



```
plot(dat, pch = (1:2)[cl$cluster], cex = 2, xlab = "X", ylab = "Y", main = "2 clusters")
legend("bottomleft", legend = c("1", "2"), pch = 1:2)
```



Кластеризация на 3 кластера

```
#Число кластеров = 3
cl = kmeans(scale(dat), 3)
#Число наблюдений
table(cl$cluster)
```

```
##
## 1 2 3
## 4 6 9
```

```
cl$centers
```

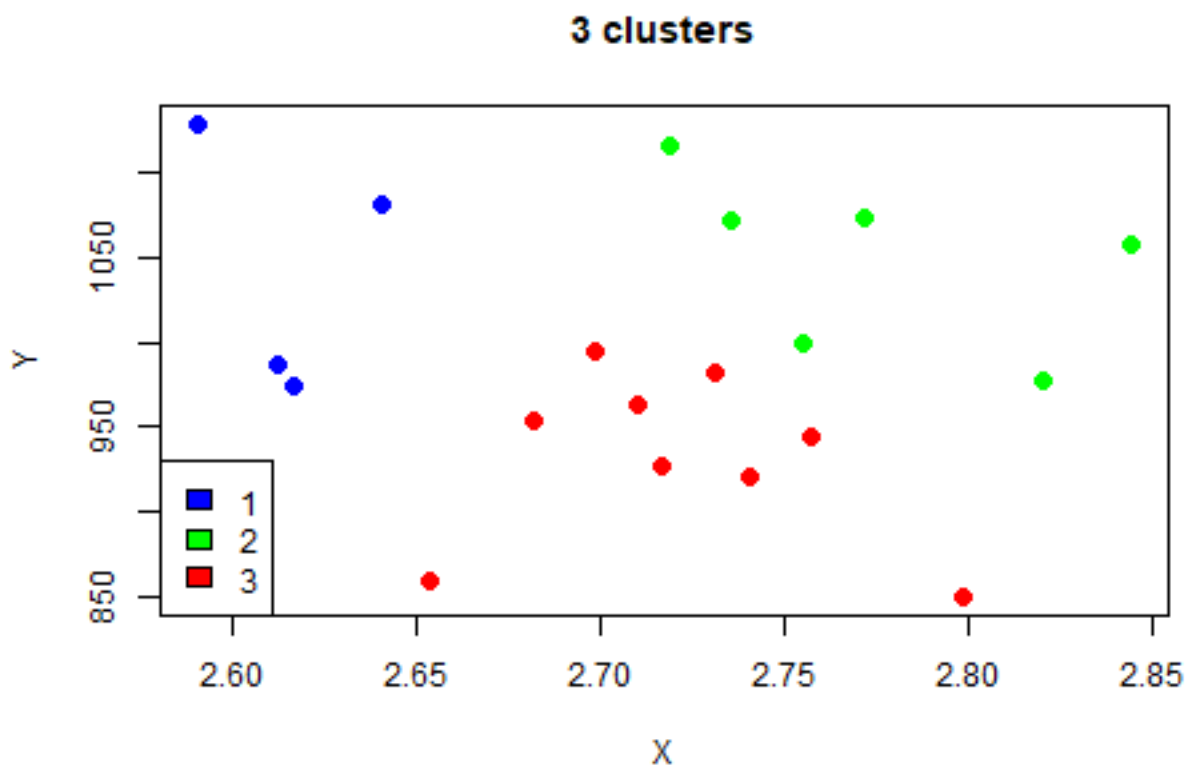
```
##          V1          V2
## 1 -1.4285685  0.6359304
## 2  0.8360842  0.7193586
## 3  0.0775299 -0.7622081
```

```
sapply(1:2, function(i) cl$centers[,i] * sd(dat[,i]) + mean(dat[,i]))
```

```
##          [,1]      [,2]
## 1 2.614975 1042.5750
## 2 2.774383 1049.1333
## 3 2.720989  932.6667
```

Кластеры на графике рассеяния

```
plot(dat, col = c("blue", "green", "red")[cl$cluster], pch = 20, cex = 2, xlab = "X", ylab =  
legend("bottomleft", legend = c("1", "2", "3"), fill = c("blue", "green", "red"))
```



```
plot(dat, pch = (1:3)[cl$cluster], cex = 2, xlab = "X", ylab = "Y", main = "3 clusters")  
legend("bottomleft", legend = c("1", "2", "3"), pch = 1:3)
```

