hw2_4.2.R

guestuser

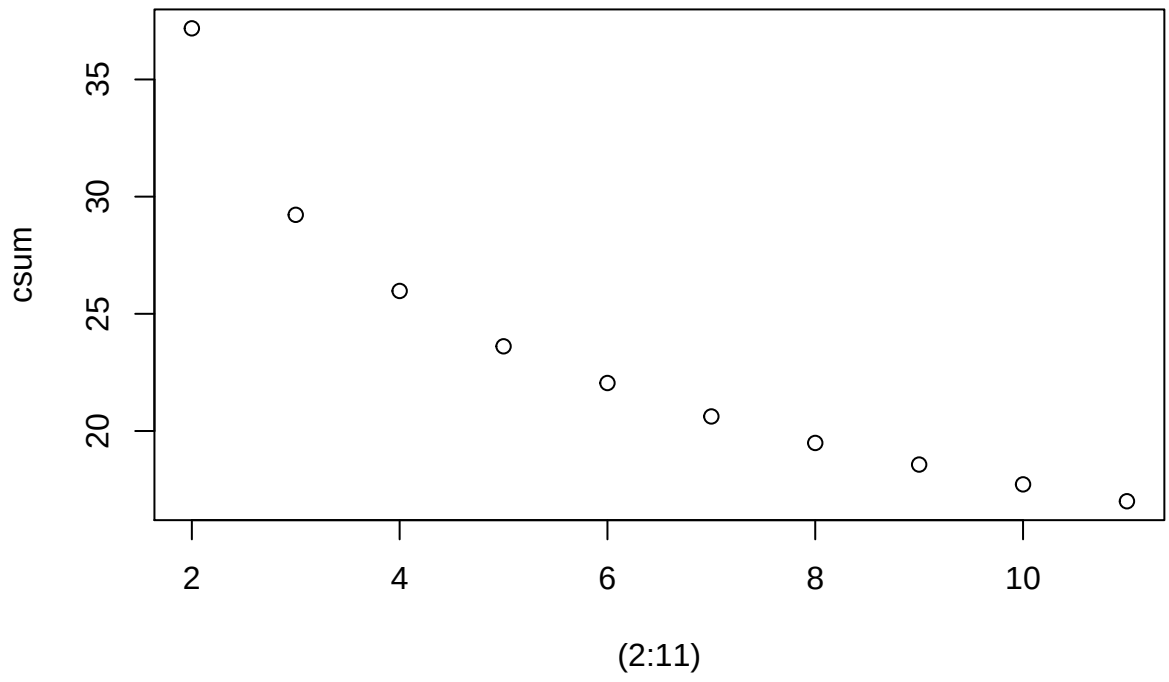
2020-05-25

```
#HW2 Question 4.2
# Clear environment
rm(list = ls())
data <- read.table("iris.txt",  header = TRUE)


set.seed(1)
#type_num<-length(levels(data[,5]))



################## MODEL 1: USE 4 PREDICTORS#############################
#creat a vector to store cluster number
cluster_num<-rep(0,5)


scdata <- data # initialize value/size of sdata
for (i in 1:4) {scdata[,i] <- (data[,i]-min(data[,i]))/(max(data[,i])-min(data[,i])) }

#x: numeric matrix, numeric data frame or a numeric vector
#centers: Possible values are the number of clusters (k) or a set of initial (distinct) cluster centers
#iter.max: The maximum number of iterations allowed. Default value is 10.
#nstart: The number of random starting partitions when centers is a number. Trying nstart > 1 is often

csum<- rep(0,10)
result <- frame()

for (cluster_num in 2:11 ){
# first use all four predictors to feed the kmeans function
nmodel<-kmeans(scdata[,1:4], cluster_num, iter.max = 10, nstart = 20)
#print(cluster_num)


for (i in 1:nrow(scdata)) {

  #calculate the total distance between cluster center and each data point
  csum[(cluster_num-1)] = csum[(cluster_num-1)] + dist(rbind(scdata[i,1:4],nmodel$centers[nmodel$cluste
}
cat("4 predictors results:")
print(table(nmodel$cluster, data$Species))

}
```

```
## 4 predictors results:
##     setosa versicolor virginica
##   1     50          0         0
##   2      0         50        50
## 4 predictors results:
##     setosa versicolor virginica
##   1      0         47        14
##   2     50          0         0
##   3      0          3        36
## 4 predictors results:
##     setosa versicolor virginica
##   1     50          0         0
##   2      0          0        29
##   3      0         23        19
##   4      0         27         2
## 4 predictors results:
##     setosa versicolor virginica
##   1      0         23        19
##   2      0         27         2
##   3     28          0         0
##   4      0          0        29
##   5     22          0         0
## 4 predictors results:
##     setosa versicolor virginica
##   1      0         23        17
##   2     28          0         0
##   3     22          0         0
##   4      0          0        19
##   5      0          0        12
##   6      0         27         2
## 4 predictors results:
##     setosa versicolor virginica
##   1      0          0        18
##   2     28          0         0
##   3      0          3        19
##   4     22          0         0
##   5      0         25         1
##   6      0         22         1
##   7      0          0        11
## 4 predictors results:
##     setosa versicolor virginica
##   1     22          0         0
##   2      0          0        18
##   3      0         13         0
##   4      0          4        20
##   5      0          0        11
##   6      0         12         1
##   7     28          0         0
##   8      0         21         0
## 4 predictors results:
##     setosa versicolor virginica
##   1      0          2        19
##   2      0          0        16
##   3     21          0         0
```

```
##   4      0         12         1
##   5     17          0         0
##   6     12          0         0
##   7      0          0        11
##   8      0         14         1
##   9      0         22         2
## 4 predictors results:
##      setosa versicolor virginica
##   1      21          0         0
##   2      12          0         0
##   3       0          4        17
##   4       0          0        20
##   5       0          0         3
##   6       0          0         9
##   7       0         12         1
##   8       0         21         0
##   9       0         13         0
##   10     17          0         0
## 4 predictors results:
##      setosa versicolor virginica
##   1      17          0         0
##   2      21          0         0
##   3       0          4         4
##   4       0          1        17
##   5       0          0         3
##   6       0         16         1
##   7       0         18         1
##   8      12          0         0
##   9       0          0        15
##   10      0          0         9
##   11      0         11         0
```

```
# Compare the model results with raw data result
plot((2:11),csum)
```



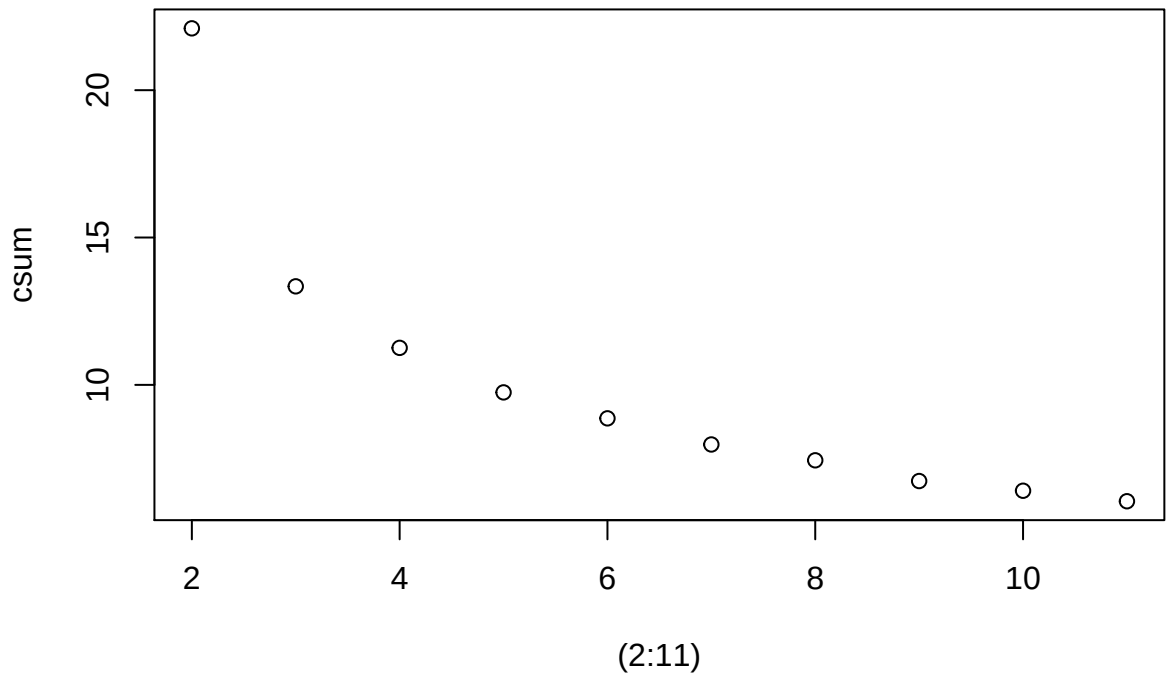

```
################### MODEL 2: Use Petal Length and Petal Width only, repeat the same testing ###########

csum<- rep(0,10)
result <- frame()

for (cluster_num in 2:11 ){

  nmodel<-kmeans(scdata[,3:4], cluster_num, iter.max = 10, nstart = 20)
  print(cluster_num)
  for (i in 1:nrow(scdata)) {

    csum[(cluster_num-1)] = csum[(cluster_num-1)] + dist(rbind(scdata[i,3:4],nmodel$centers[nmodel$clus
  }
  cat("Petal Length and Petal Width as predictors only results:")
  print(table(nmodel$cluster, data$Species))
}
```

```
## [1] 2
## Petal Length and Petal Width as predictors only results:
##     setosa versicolor virginica
##   1      0         50        50
##   2     50          0         0
## [1] 3
## Petal Length and Petal Width as predictors only results:
##     setosa versicolor virginica
##   1     50          0         0
```

```
##   2      0         48         4
##   3      0          2        46
## [1] 4
## Petal Length and Petal Width as predictors only results:
##     setosa versicolor virginica
##   1      0          0        27
##   2      0         42         0
##   3      0          8        23
##   4     50          0         0
## [1] 5
## Petal Length and Petal Width as predictors only results:
##     setosa versicolor virginica
##   1      0          0        25
##   2      0         25         4
##   3      0         23         0
##   4     50          0         0
##   5      0          2        21
## [1] 6
## Petal Length and Petal Width as predictors only results:
##     setosa versicolor virginica
##   1      0         23         0
##   2      0         24         4
##   3      0          3        18
##   4      0          0        20
##   5     50          0         0
##   6      0          0         8
## [1] 7
## Petal Length and Petal Width as predictors only results:
##     setosa versicolor virginica
##   1     50          0         0
##   2      0         16         4
##   3      0         22         0
##   4      0          0         8
##   5      0         10         0
##   6      0          0        20
##   7      0          2        18
## [1] 8
## Petal Length and Petal Width as predictors only results:
##     setosa versicolor virginica
##   1      0          0         8
##   2     50          0         0
##   3      0          2        13
##   4      0         22         0
##   5      0          0        20
##   6      0         10         0
##   7      0          0         5
##   8      0         16         4
## [1] 9
## Petal Length and Petal Width as predictors only results:
##     setosa versicolor virginica
##   1      0         22         0
##   2      0          0         5
##   3      0          2        13
##   4      0         10         0
```

```
##   5      0         16         4
##   6     16          0         0
##   7      0          0        20
##   8      0          0         8
##   9     34          0         0
## [1] 10
## Petal Length and Petal Width as predictors only results:
##      setosa versicolor virginica
##   1      16          0         0
##   2       0          0         8
##   3       0         22         0
##   4       0          2        13
##   5       0          0         5
##   6      34          0         0
##   7       0         16         4
##   8       0          0        12
##   9       0         10         0
##   10      0          0         8
## [1] 11
## Petal Length and Petal Width as predictors only results:
##      setosa versicolor virginica
##   1      16          0         0
##   2       0          0         5
##   3       0          0         7
##   4       0          0         5
##   5       0         10         0
##   6      34          0         0
##   7       0          0         8
##   8       0          0         8
##   9       0         22         0
##   10      0          2        13
##   11      0         16         4
```

```
plot((2:11),csum)
```



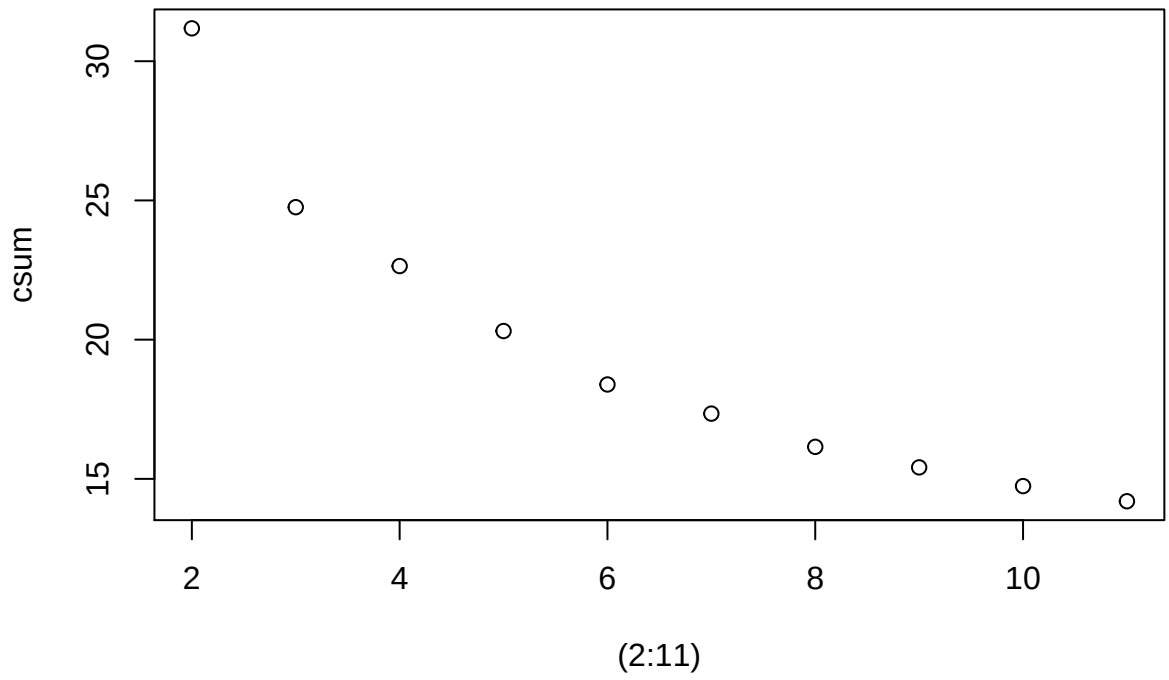

```
#################### MODEL 3: Use Sepal.Length, Sepal.Width and Petal.Length as predictors, repeat the

csum<- rep(0,10)
result <- frame()

for (cluster_num in 2:11 ){

  nmodel<-kmeans(scdata[,1:3], cluster_num, iter.max = 10, nstart = 20)
  print(cluster_num)
  for (i in 1:nrow(scdata)) {

    csum[(cluster_num-1)] = csum[(cluster_num-1)] + dist(rbind(scdata[i,1:3],nmodel$centers[nmodel$clus
  }
  cat("SepalLength,Sepal Width and PetalLength as predictors results:")
  print(table(nmodel$cluster, data$Species))
}
```

```
## [1] 2
## SepalLength,Sepal Width and PetalLength as predictors results:
##     setosa versicolor virginica
##   1     50          1         0
##   2      0         49        50
## [1] 3
## SepalLength,Sepal Width and PetalLength as predictors results:
##     setosa versicolor virginica
##   1     50          0         0
```

```
##   2      0         10        35
##   3      0         40        15
## [1] 4
## SepalLength,Sepal Width and PetalLength as predictors results:
##     setosa versicolor virginica
##   1      0          0        14
##   2      0         21        30
##   3      0         29         6
##   4     50          0         0
## [1] 5
## SepalLength,Sepal Width and PetalLength as predictors results:
##     setosa versicolor virginica
##   1      0          0        14
##   2      0         29         6
##   3      0         21        30
##   4     22          0         0
##   5     28          0         0
## [1] 6
## SepalLength,Sepal Width and PetalLength as predictors results:
##     setosa versicolor virginica
##   1     24          0         0
##   2      0          0        12
##   3     26          0         0
##   4      0         19         2
##   5      0         21        13
##   6      0         10        23
## [1] 7
## SepalLength,Sepal Width and PetalLength as predictors results:
##     setosa versicolor virginica
##   1     28          0         0
##   2      0         14         1
##   3      0         21         5
##   4      0          0        12
##   5      0          5        13
##   6      0         10        19
##   7     22          0         0
## [1] 8
## SepalLength,Sepal Width and PetalLength as predictors results:
##     setosa versicolor virginica
##   1      0          0        12
##   2     12          0         0
##   3      0         21         5
##   4     22          0         0
##   5     16          0         0
##   6      0          5        13
##   7      0         10        19
##   8      0         14         1
## [1] 9
## SepalLength,Sepal Width and PetalLength as predictors results:
##     setosa versicolor virginica
##   1      0         12         3
##   2     12          0         0
##   3     16          0         0
##   4      0          0        12
```

```
##   5      0         20         5
##   6      0         14         1
##   7      0          4        13
##   8      0          0        16
##   9     22          0         0
## [1] 10
## SepalLength,Sepal Width and PetalLength as predictors results:
##      setosa versicolor virginica
##   1       0          0        19
##   2       0          0         3
##   3       0          4        11
##   4      26          0         0
##   5       0         20         5
##   6       0          0         9
##   7       0         12         2
##   8       7          0         0
##   9      17          0         0
##   10      0         14         1
## [1] 11
## SepalLength,Sepal Width and PetalLength as predictors results:
##      setosa versicolor virginica
##   1      20          0         0
##   2      16          0         0
##   3       0          0         3
##   4       0          4        13
##   5       0         15         1
##   6       0          0         9
##   7       4          0         0
##   8      10          0         0
##   9       0          5        19
##   10      0         14         4
##   11      0         12         1
```

```
plot((2:11),csum)
```




```
################### CONCLUSTION ###################
cat("\n Conclusion: From the results from 3 models with different predictors we can conclude: \n 1. Fro
```

```
##
##  Conclusion: From the results from 3 models with different predictors we can conclude:
##  1. From the plot K VS Total Distance, The best K valve is 3.
##  2. The best model is the one using Petal Length and Petal Width.
```