Experiment - 8

Student Name: Rajdeep Jaiswal	**UID:** 20BCS2761
Branch: CSE	**Section/Group:** 20BCS-DM-902/(B)
Semester: 6th	**Subject Code:** 20CSP-376
Subject Name: Data Mining Lab	

1. Aim:

To perform the hierarchical clustering using R programming.

2.CODE:

```
# Loading package
library(dplyr)

setwd("C:\\Users\\asg72\\Downloads")

# Importing the dataset dataset
= read.csv('mtcars.csv')
head(dataset)
#find the missing value and give sum
sum(is.na(dataset)) #index value
which(is.na(dataset))

# suppressing warning   oldw  = warning will be saved in the oldW
             # 2 ways globally and at a particuilaar location ,  oldw
<- getOption("warn")
options(warn = -1)   # used to disable the warning

# Finding distance matrix , dist function is used distance_mat
<- dist(dataset, method = 'euclidean') distance_mat
```

Fitting Hierarchical clustering Model to training dataset

```
set.seed(300)  # Setting seed , hculst used for hierachical algomet
Hierar_cl <- hclust(distance_mat, method = "average")
Hierar_cl

# Initate PDF File
pdf("dendrogram.pdf", paper="a4")
# Plotting dendrogram
plot(Hierar_cl)


# Choosing no. of clusters #
Cutting tree by height
abline(h = 110, col = "green")

# Cutting tree by no. of clusters
fit <- cutree(Hierar_cl, k = 3 ) fit

table(fit)
rect.hclust(Hierar_cl, k = 3, border = "green")

#Close PDF file
dev.off()

options(warn = oldw)
```

3.OUTPUT:

```
> head(dataset)
            model  mpg cyl disp  hp drat    wt  qsec vs am gear carb
1         Mazda RX4 21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
2     Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
3        Datsun 710 22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
4    Hornet 4 Drive 21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
5 Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
6          Valiant 18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
> sum(is.na(dataset))
[1] 0
> which(is.na(dataset))
integer(0)
```

```
> options(warn = -1)
> distance_mat <- dist(dataset, method = 'euclidean')
> distance_mat
              1            2            3            4            5            6            7            8
2      0.6426861
3     57.3501684   57.3323195
4    102.4751863  102.4578196  157.7075873
5    219.6902907  219.6886352  276.8703323  126.4114577
6     68.3830346   68.3490386  122.9907748   35.0427400  158.8884772
7    252.1420660  252.1433392  307.5733083  176.9638266   73.2972000  203.2629439
8     52.3833832   52.3429988   51.8665904  126.6665328  252.2457990   93.5748579  293.8043512
9     26.6007840   26.4547042   34.6557811  123.5011141  243.8748593   88.7879236  277.7050380   35.1852423
10    16.0473753   15.9758235   69.9127409   95.4875860  208.1980909   62.9718760  238.0336323   67.6557226
11    16.3693638   16.2767215   70.0065197   95.5282033  208.2045252   62.9453304  238.0142729   67.7752602
12   141.4527547  141.4473034  197.6082488   75.7200542   88.1412773   94.7299546  111.1399766  182.9507499
13   141.4221942  141.4165465  197.5744921   75.6520811   88.1199197   94.7090093  111.1646582  182.9058059
14   141.5036887  141.4962190  197.6490336   75.7988217   88.1876454   94.7427727  111.1322519  183.0027794
15   340.8505856  340.8463207  398.0382643  244.8649988  121.4509378  278.4839619  124.3164162  371.4776268
16   332.1892374  332.1850099  389.3781636  238.1096244  112.8675270  271.1751495  109.1584927  364.5083003
17   318.2700159  318.2664404  375.2781676  227.8552885  101.5272221  259.8331789   85.0506208  353.2341361
18    97.4152437   97.3996856   42.8161851  193.1937919  315.4681111  158.8792766  348.8299049   71.6614244
19   107.4032225  107.3960309   55.1169486  200.0694031  323.8183481  166.0299521  359.3504182   75.2030606
20   105.0774895  105.0615321   49.7724593  201.2387337  323.3229716  166.9373033  356.1856777   79.6725361
21    44.1887669   44.1453162   13.5419963  144.6903595  263.5534306  109.9693564  294.5925384   46.0453827
22   170.3681395  170.3664119  227.4633534   75.6615213   51.1619994  108.0302286  108.5225032  201.4375717
23   156.2570341  156.2536208  213.4041219   64.0886281   64.1588519   95.0928109  115.2134624  188.5762107
24   243.5933492  243.5954371  298.7223974  170.9407043   74.1221240  196.1991421   10.5241644  286.0132080
25   259.7357282  259.7338236  316.8475035  163.1706825   41.7841182  196.9102505   84.3988364  289.7982376
26    96.6181532   96.6068343   41.6548837  192.6487336  314.9764049  158.1717723  348.3130155   70.9363642
27    46.3778031   46.3819737   13.7198025  145.3457410  265.4460619  110.7745706  297.8802436   41.2009722
28    68.6557098   68.6592880   26.2107212  170.4952183  284.4689064  136.6420672  309.6394768   76.1385488
29   256.3377456  256.3426255  310.5135070  188.1229730   93.5772595  212.0451687   22.2111937  300.3088979
30    69.7349202   69.7456925   94.2542259  136.3574309  224.6305578  111.3081918  236.2619895  118.3403933
31   277.4575972  277.4614844  323.5461109  239.5398389  178.3002026  253.2195919  112.5124793  327.8195350
32    40.9319964   40.9040055   21.6141191  143.1297786  259.0341717  108.8191125  287.3695025   56.0693840
              9           10           11           12           13           14           15           16
2
3
4
5
6
7
8
9
10    41.0469013
11    41.1382251    1.5908831
12   166.9244104  127.8052897  127.7863390
13   166.8806832  127.7845343  127.7753139    1.0263439
14   166.9574450  127.8358383  127.7994797    1.4336603    2.2334238
15   364.8144682  329.4146190  329.3782947  206.6833900  206.7159432  206.6503263
16   356.4923417  320.3127104  320.2756808  195.9415162  195.9763611  195.9074551   16.3171112
17   343.0376198  305.7305222  305.7140652  179.2931017  179.3079613  179.2885979   42.6559508   26.4995878
18    72.3948491  111.2411734  111.4267379  238.4774704  238.4089944  238.5614327  436.3452367  428.2526524
19    82.0310189  121.9185233  122.0679325  248.5977057  248.5398799  248.6694788  444.2397324  436.5532937
20    80.1869581  118.6816951  118.8725491  245.9909261  245.9176215  246.0786915  444.2579996  436.1094388
21    22.0342640   56.7815251   56.8459269  184.4548282  184.4242201  184.4845780  384.6972222  376.0356779
22   194.0964336  159.6914533  159.6698511   54.1042378   54.1286659   54.1046521  170.9075153  163.2296608
```

```
23 180.2030192 145.3330455 145.3042185  43.0403569  43.0749886  43.0245861 184.7253835 176.6114040
24 269.2079338 229.3146743 229.2891454 103.1099956 103.1479985 103.0925302 134.1707582 119.1664690
25 283.4546835 248.7631779 248.7715537 129.8656084 129.8484396 129.9029575  82.0308241  75.9272322
26  71.6049147 110.4431765 110.5630250 237.8904522 237.8429706 237.9477807 435.8024174 427.7086297
27  23.1015985  60.2090911  60.4196244 187.4851980 187.4347238 187.5568836 386.5523281 378.1118651
28  52.3375645  77.4412563  77.6899139 201.9030413 201.8334014 201.9964967 405.8115736 396.3452286
29 281.9819925 241.6979096 241.6919517 117.8347495 117.8467546 117.8504967 140.8064858 125.0472695
30  84.2414996  59.3637989  59.4288167 136.8534684 136.8331463 136.8986192 343.1531864 331.8334202
31 301.7206850 261.7300163 261.7195959 164.1517399 164.1659133 164.1570368 224.4940467 208.2059924
32  25.7892790  50.9755203  51.0623330 178.0292898 178.0005131 178.0651207 380.2901443 371.2041799
            17          18          19          20          21          22          23          24
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18 414.8907361
19 423.8601736  15.2064780
20 422.6259476   8.1807574  14.9860626
21 361.9831338  55.2311792  66.7398654  62.5061981
22 152.4079240 265.5416656 273.4932815 273.4772486 214.1518566
23 164.8269138 251.8420046 260.0340520 259.7500160 200.0758606  14.6387118
24  95.3472989 340.1445613 350.8239072 347.4509622 285.7989787 104.7647489 110.3021430
25  71.2357882 354.6857648 362.4981299 362.6038593 303.5469418  89.6230313 103.6983501  90.1025000
26 414.4020459   5.3762229  15.4379450  10.8543883  54.1463411 264.9417408 251.2219263 339.6071500
27 364.3584805  51.2461333  62.1027713  58.8288160   9.0383803 215.8339316 201.9036821 289.2047229
28 380.8116875  52.1306014  66.8975369  56.2806880  32.6434173 236.5720395 222.2173101 300.4070995
29  99.6217234 352.1562330 363.3005255 359.2880266 297.8073096 124.0320824 128.8695540  20.2197141
30 313.5111466 134.1042094 148.0054687 139.4056640  85.8922551 182.7063785 168.2697611 226.3869377
31 182.0438183 365.0762342 378.2659534 371.0571374 312.4901303 194.1724083 193.3884466 107.1565809
32 356.4654302  64.0572261  76.6393634  70.7301283  12.7952587 210.3222800 196.0440042 278.3791805
            25          26          27          28          29          30          31
2
3
4
5
6
7
8
9
10
11
12
13
14
15
```
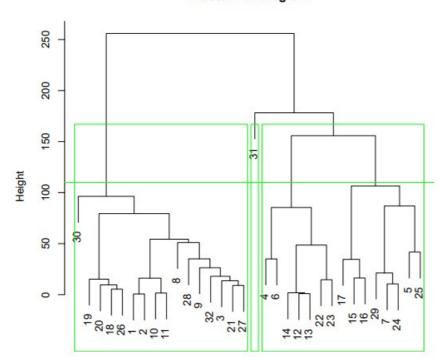
```
24
25
26 354.2197786
27 305.1560233   50.5286726
28 325.2323144   52.0569009   35.2693851
29 106.2628870 351.6736423 301.4175147 310.7679839
30 266.3984024 133.5047653   91.8196236   84.0328673 234.4395173
31 196.6980050 364.6439032 317.4364423 316.7652601   90.8041192 233.4738757
32 299.5003796   63.0983209   19.5895705   29.0470632 289.8187538   73.6088448 301.9715381
> set.seed(300)   # Setting seed
> Hierar_cl <- hclust(distance_mat, method = "average")
> Hierar_cl

Call:
hclust(d = distance_mat, method = "average")

Cluster method   : average
Distance         : euclidean
Number of objects: 32

> pdf("dendrogram.pdf", paper="a4")
> # Plotting dendrogram
> plot(Hierar_cl)
>
>
>
> # Choosing no. of clusters
> # Cutting tree by height
> abline(h = 110, col = "green")
>
> # Cutting tree by no. of clusters
> fit <- cutree(Hierar_cl, k = 3 )
> fit
 [1] 1 1 1 2 2 2 2 1 1 1 1 2 2 2 2 2 2 1 1 1 1 2 2 2 2 1 1 1 2 1 3 1
>
> table(fit)
fit
 1  2  3
16 15  1
> rect.hclust(Hierar_cl, k = 3, border = "green")
>
> #Close PDF file
> dev.off()
null device
          1
>
> options(warn = oldw)
> |
```

Cluster Dendrogram

distance_mat
hclust (*, "average")