NAME-Anmol Sinha
E_MAIL- 1805553@kiit.ac.in

Clustering

- Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters).

- Cluster analysis itself is not one specific algorithm, but the general task to be solved, which can be achieved by various algorithms that differ significantly in their understanding of what constitutes a cluster and how to efficiently find them.

- It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including pattern recognition, image analysis, information retrieval, bioinformatics, data compression, computer graphics and machine learning.

{ R-Project Code }

```
> data("iris")
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
> iris
    Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
1            5.1         3.5          1.4         0.2     setosa
2            4.9         3.0          1.4         0.2     setosa
3            4.7         3.2          1.3         0.2     setosa
4            4.6         3.1          1.5         0.2     setosa
5            5.0         3.6          1.4         0.2     setosa
6            5.4         3.9          1.7         0.4     setosa
7            4.6         3.4          1.4         0.3     setosa
8            5.0         3.4          1.5         0.2     setosa
9            4.4         2.9          1.4         0.2     setosa
10           4.9         3.1          1.5         0.1     setosa
11           5.4         3.7          1.5         0.2     setosa
12           4.8         3.4          1.6         0.2     setosa
13           4.8         3.0          1.4         0.1     setosa
14           4.3         3.0          1.1         0.1     setosa
15           5.8         4.0          1.2         0.2     setosa
16           5.7         4.4          1.5         0.4     setosa
17           5.4         3.9          1.3         0.4     setosa
18           5.1         3.5          1.4         0.3     setosa
19           5.7         3.8          1.7         0.3     setosa
20           5.1         3.8          1.5         0.3     setosa
21           5.4         3.4          1.7         0.2     setosa
22           5.1         3.7          1.5         0.4     setosa
23           4.6         3.6          1.0         0.2     setosa
24           5.1         3.3          1.7         0.5     setosa
25           4.8         3.4          1.9         0.2     setosa
26           5.0         3.0          1.6         0.2     setosa
27           5.0         3.4          1.6         0.4     setosa
28           5.2         3.5          1.5         0.2     setosa
29           5.2         3.4          1.4         0.2     setosa
```

30	4.7	3.2	1.6	0.2	setosa
31	4.8	3.1	1.6	0.2	setosa
32	5.4	3.4	1.5	0.4	setosa
33	5.2	4.1	1.5	0.1	setosa
34	5.5	4.2	1.4	0.2	setosa
35	4.9	3.1	1.5	0.2	setosa
36	5.0	3.2	1.2	0.2	setosa
37	5.5	3.5	1.3	0.2	setosa
38	4.9	3.6	1.4	0.1	setosa
39	4.4	3.0	1.3	0.2	setosa
40	5.1	3.4	1.5	0.2	setosa
41	5.0	3.5	1.3	0.3	setosa
42	4.5	2.3	1.3	0.3	setosa
43	4.4	3.2	1.3	0.2	setosa
44	5.0	3.5	1.6	0.6	setosa
45	5.1	3.8	1.9	0.4	setosa
46	4.8	3.0	1.4	0.3	setosa
47	5.1	3.8	1.6	0.2	setosa
48	4.6	3.2	1.4	0.2	setosa
49	5.3	3.7	1.5	0.2	setosa
50	5.0	3.3	1.4	0.2	setosa
51	7.0	3.2	4.7	1.4	versicolor
52	6.4	3.2	4.5	1.5	versicolor
53	6.9	3.1	4.9	1.5	versicolor
54	5.5	2.3	4.0	1.3	versicolor
55	6.5	2.8	4.6	1.5	versicolor
56	5.7	2.8	4.5	1.3	versicolor
57	6.3	3.3	4.7	1.6	versicolor
58	4.9	2.4	3.3	1.0	versicolor
59	6.6	2.9	4.6	1.3	versicolor
60	5.2	2.7	3.9	1.4	versicolor
61	5.0	2.0	3.5	1.0	versicolor
62	5.9	3.0	4.2	1.5	versicolor
63	6.0	2.2	4.0	1.0	versicolor
64	6.1	2.9	4.7	1.4	versicolor
65	5.6	2.9	3.6	1.3	versicolor
66	6.7	3.1	4.4	1.4	versicolor
67	5.6	3.0	4.5	1.5	versicolor
68	5.8	2.7	4.1	1.0	versicolor
69	6.2	2.2	4.5	1.5	versicolor

70	5.6	2.5	3.9	1.1 versicolor
71	5.9	3.2	4.8	1.8 versicolor
72	6.1	2.8	4.0	1.3 versicolor
73	6.3	2.5	4.9	1.5 versicolor
74	6.1	2.8	4.7	1.2 versicolor
75	6.4	2.9	4.3	1.3 versicolor
76	6.6	3.0	4.4	1.4 versicolor
77	6.8	2.8	4.8	1.4 versicolor
78	6.7	3.0	5.0	1.7 versicolor
79	6.0	2.9	4.5	1.5 versicolor
80	5.7	2.6	3.5	1.0 versicolor
81	5.5	2.4	3.8	1.1 versicolor
82	5.5	2.4	3.7	1.0 versicolor
83	5.8	2.7	3.9	1.2 versicolor
84	6.0	2.7	5.1	1.6 versicolor
85	5.4	3.0	4.5	1.5 versicolor
86	6.0	3.4	4.5	1.6 versicolor
87	6.7	3.1	4.7	1.5 versicolor
88	6.3	2.3	4.4	1.3 versicolor
89	5.6	3.0	4.1	1.3 versicolor
90	5.5	2.5	4.0	1.3 versicolor
91	5.5	2.6	4.4	1.2 versicolor
92	6.1	3.0	4.6	1.4 versicolor
93	5.8	2.6	4.0	1.2 versicolor
94	5.0	2.3	3.3	1.0 versicolor
95	5.6	2.7	4.2	1.3 versicolor
96	5.7	3.0	4.2	1.2 versicolor
97	5.7	2.9	4.2	1.3 versicolor
98	6.2	2.9	4.3	1.3 versicolor
99	5.1	2.5	3.0	1.1 versicolor
100	5.7	2.8	4.1	1.3 versicolor
101	6.3	3.3	6.0	2.5 virginica
102	5.8	2.7	5.1	1.9 virginica
103	7.1	3.0	5.9	2.1 virginica
104	6.3	2.9	5.6	1.8 virginica
105	6.5	3.0	5.8	2.2 virginica
106	7.6	3.0	6.6	2.1 virginica
107	4.9	2.5	4.5	1.7 virginica
108	7.3	2.9	6.3	1.8 virginica
109	6.7	2.5	5.8	1.8 virginica

110	7.2	3.6	6.1	2.5	virginica
111	6.5	3.2	5.1	2.0	virginica
112	6.4	2.7	5.3	1.9	virginica
113	6.8	3.0	5.5	2.1	virginica
114	5.7	2.5	5.0	2.0	virginica
115	5.8	2.8	5.1	2.4	virginica
116	6.4	3.2	5.3	2.3	virginica
117	6.5	3.0	5.5	1.8	virginica
118	7.7	3.8	6.7	2.2	virginica
119	7.7	2.6	6.9	2.3	virginica
120	6.0	2.2	5.0	1.5	virginica
121	6.9	3.2	5.7	2.3	virginica
122	5.6	2.8	4.9	2.0	virginica
123	7.7	2.8	6.7	2.0	virginica
124	6.3	2.7	4.9	1.8	virginica
125	6.7	3.3	5.7	2.1	virginica
126	7.2	3.2	6.0	1.8	virginica
127	6.2	2.8	4.8	1.8	virginica
128	6.1	3.0	4.9	1.8	virginica
129	6.4	2.8	5.6	2.1	virginica
130	7.2	3.0	5.8	1.6	virginica
131	7.4	2.8	6.1	1.9	virginica
132	7.9	3.8	6.4	2.0	virginica
133	6.4	2.8	5.6	2.2	virginica
134	6.3	2.8	5.1	1.5	virginica
135	6.1	2.6	5.6	1.4	virginica
136	7.7	3.0	6.1	2.3	virginica
137	6.3	3.4	5.6	2.4	virginica
138	6.4	3.1	5.5	1.8	virginica
139	6.0	3.0	4.8	1.8	virginica
140	6.9	3.1	5.4	2.1	virginica
141	6.7	3.1	5.6	2.4	virginica
142	6.9	3.1	5.1	2.3	virginica
143	5.8	2.7	5.1	1.9	virginica
144	6.8	3.2	5.9	2.3	virginica
145	6.7	3.3	5.7	2.5	virginica
146	6.7	3.0	5.2	2.3	virginica
147	6.3	2.5	5.0	1.9	virginica
148	6.5	3.0	5.2	2.0	virginica
149	6.2	3.4	5.4	2.3	virginica

```
150          5.9         3.0          5.1         1.8  virginica
> # K-Means Clustering
> library(cluster)
> x=iris[,3:4]
> model=kmeans(x,3)
> clusplot(x,model$cluster)
> clusplot(x,model$cluster,shade=TRUE)
> clusplot(x,model$cluster,shade=TRUE,color=TRUE)
> x
     Petal.Length Petal.Width
1             1.4         0.2
2             1.4         0.2
3             1.3         0.2
4             1.5         0.2
5             1.4         0.2
6             1.7         0.4
7             1.4         0.3
8             1.5         0.2
9             1.4         0.2
10            1.5         0.1
11            1.5         0.2
12            1.6         0.2
13            1.4         0.1
14            1.1         0.1
15            1.2         0.2
16            1.5         0.4
17            1.3         0.4
18            1.4         0.3
19            1.7         0.3
20            1.5         0.3
21            1.7         0.2
22            1.5         0.4
23            1.0         0.2
24            1.7         0.5
25            1.9         0.2
26            1.6         0.2
27            1.6         0.4
28            1.5         0.2
29            1.4         0.2
30            1.6         0.2
```

31	1.6	0.2
32	1.5	0.4
33	1.5	0.1
34	1.4	0.2
35	1.5	0.2
36	1.2	0.2
37	1.3	0.2
38	1.4	0.1
39	1.3	0.2
40	1.5	0.2
41	1.3	0.3
42	1.3	0.3
43	1.3	0.2
44	1.6	0.6
45	1.9	0.4
46	1.4	0.3
47	1.6	0.2
48	1.4	0.2
49	1.5	0.2
50	1.4	0.2
51	4.7	1.4
52	4.5	1.5
53	4.9	1.5
54	4.0	1.3
55	4.6	1.5
56	4.5	1.3
57	4.7	1.6
58	3.3	1.0
59	4.6	1.3
60	3.9	1.4
61	3.5	1.0
62	4.2	1.5
63	4.0	1.0
64	4.7	1.4
65	3.6	1.3
66	4.4	1.4
67	4.5	1.5
68	4.1	1.0
69	4.5	1.5
70	3.9	1.1

71	4.8	1.8
72	4.0	1.3
73	4.9	1.5
74	4.7	1.2
75	4.3	1.3
76	4.4	1.4
77	4.8	1.4
78	5.0	1.7
79	4.5	1.5
80	3.5	1.0
81	3.8	1.1
82	3.7	1.0
83	3.9	1.2
84	5.1	1.6
85	4.5	1.5
86	4.5	1.6
87	4.7	1.5
88	4.4	1.3
89	4.1	1.3
90	4.0	1.3
91	4.4	1.2
92	4.6	1.4
93	4.0	1.2
94	3.3	1.0
95	4.2	1.3
96	4.2	1.2
97	4.2	1.3
98	4.3	1.3
99	3.0	1.1
100	4.1	1.3
101	6.0	2.5
102	5.1	1.9
103	5.9	2.1
104	5.6	1.8
105	5.8	2.2
106	6.6	2.1
107	4.5	1.7
108	6.3	1.8
109	5.8	1.8
110	6.1	2.5

111	5.1	2.0
112	5.3	1.9
113	5.5	2.1
114	5.0	2.0
115	5.1	2.4
116	5.3	2.3
117	5.5	1.8
118	6.7	2.2
119	6.9	2.3
120	5.0	1.5
121	5.7	2.3
122	4.9	2.0
123	6.7	2.0
124	4.9	1.8
125	5.7	2.1
126	6.0	1.8
127	4.8	1.8
128	4.9	1.8
129	5.6	2.1
130	5.8	1.6
131	6.1	1.9
132	6.4	2.0
133	5.6	2.2
134	5.1	1.5
135	5.6	1.4
136	6.1	2.3
137	5.6	2.4
138	5.5	1.8
139	4.8	1.8
140	5.4	2.1
141	5.6	2.4
142	5.1	2.3
143	5.1	1.9
144	5.9	2.3
145	5.7	2.5
146	5.2	2.3
147	5.0	1.9
148	5.2	2.0
149	5.4	2.3
150	5.1	1.8

```
> model=kmeans(x,4)
> clusplot(x,model$cluster,shade=TRUE,color=TRUE)
> model=kmeans(x,5)
> clusplot(x,model$cluster,shade=TRUE,color=TRUE)
> model=kmeans(x,3)
> clusplot(x,model$cluster,shade=TRUE,
color=TRUE,xlab='Petal_Length',ylab='Petal_Width')
```