

Result:

RStudio interface showing the execution of R code to load and inspect the 'breast-cancer-wisconsin.data.csv' dataset. The console displays the first 41 rows of the dataset, and the Environment pane shows the loaded objects: cancerdataset, test, and trainingdata.

```
rm(list=ls()) # Clears the environment , removes all the objects

# Question 3.1
# Assignment 3.1
# Load the "breast-cancer-wisconsin.data.csv" from file location
cancerdataset<-read.csv("C:/Users/Vaah/Desktop/breast-cancer-wisconsin.data.csv", header=TRUE)
cancerdataset
```

Sample	P1	P2	P3	P4	P5	P6	P7	P8	P9	Class
1	1000025	5	1	1	1	2	3	5	1	2
2	1002945	5	4	4	5	7	10	2	1	2
3	1015425	3	1	1	1	2	3	1	1	2
4	1016277	6	8	8	1	3	4	3	7	1
5	1017023	4	1	3	2	2	3	1	1	2
6	1017122	8	10	10	8	7	10	9	7	1
7	1018099	1	1	1	1	2	10	1	1	2
8	1018581	2	1	2	1	2	1	1	1	2
9	1033078	2	1	1	1	2	1	1	1	2
10	1033078	4	2	1	1	2	1	1	1	2
11	1035283	1	1	1	1	1	1	1	1	2
12	1036372	2	5	3	2	3	2	4	4	1
13	1041801	5	3	3	3	2	2	4	4	1
14	1043999	1	1	1	1	2	3	1	1	2
15	1044572	8	7	3	10	7	9	5	4	4
16	1047810	7	4	6	4	6	4	3	1	4
17	1048652	4	1	1	1	2	2	3	1	2
18	1049815	4	1	1	1	2	3	1	1	2
19	1050670	10	7	7	6	4	10	4	1	2
20	1050718	6	1	1	1	2	1	1	1	2
21	1054590	7	3	2	10	5	10	5	4	4
22	1054591	10	5	3	3	7	10	1	1	2
23	1056784	3	1	1	1	2	2	2	1	2
24	1057013	8	4	5	1	2	7	7	1	1
25	1059332	1	1	1	1	2	1	1	1	2
26	1065726	5	2	3	4	2	7	3	6	1
27	1066373	3	2	1	1	1	2	1	1	2
28	1066979	5	1	1	1	2	1	2	1	2
29	1067444	2	1	1	1	1	1	1	1	2
30	1070995	1	1	3	1	2	1	1	1	2
31	1070995	3	1	1	1	1	1	1	1	2
32	1073760	2	1	1	1	2	1	1	1	2
33	1072179	10	7	7	3	8	5	7	4	3
34	1074610	2	1	1	1	2	1	1	1	2
35	1075123	3	1	2	1	2	1	2	1	2
36	1079104	2	1	1	1	2	1	1	1	2
37	1080185	10	10	10	8	1	8	9	1	4
38	1081791	6	2	1	1	1	1	1	1	2
39	1084584	5	4	6	2	10	6	1	1	4
40	1091262	2	5	3	3	6	7	7	5	1
41	1096800	6	6	6	9	6	7	8	1	2

RStudio interface showing the execution of R code to load and inspect the 'breast-cancer-wisconsin.data.csv' dataset. The console displays the first 41 rows of the dataset, and the Environment pane shows the loaded objects: cancerdataset, test, and trainingdata.

```
rm(list=ls()) # Clears the environment , removes all the objects

# Question 3.1
# Assignment 3.1
# Load the "breast-cancer-wisconsin.data.csv" from file location
cancerdataset<-read.csv("C:/Users/Vaah/Desktop/breast-cancer-wisconsin.data.csv", header=TRUE)
cancerdataset
```

Sample	P1	P2	P3	P4	P5	P6	P7	P8	P9	Class
42	1099510	10	4	3	1	3	3	6	5	2
43	1100224	6	10	10	2	8	10	7	3	4
44	1102573	5	6	5	6	10	1	3	1	4
45	1103608	10	10	10	4	8	1	8	10	1
46	1103722	1	1	1	1	1	1	1	1	2
47	1105237	3	7	7	4	4	9	4	8	1
48	1105324	1	1	1	1	2	1	1	1	2
49	1106095	4	1	1	3	2	1	3	1	1
50	1106829	7	8	7	2	4	2	8	2	4
51	1108370	9	5	8	1	2	2	2	1	4
52	1108449	5	3	3	4	2	4	3	1	4
53	1110102	10	3	6	2	3	4	10	2	4
54	1110301	5	5	5	8	10	8	7	3	7
55	1110324	10	5	9	8	8	9	1	1	4
56	1111249	10	6	6	3	4	5	3	6	1
57	1112009	8	10	10	1	1	6	3	9	1
58	1113018	8	2	4	1	5	1	5	4	4
59	1113483	5	2	3	1	6	10	5	1	4
60	1113966	9	5	3	1	2	1	1	1	4
61	1115282	5	3	5	5	3	1	4	10	1
62	1115299	1	1	1	1	2	2	1	1	2
63	1116116	9	10	10	1	10	3	5	1	4
64	1116132	6	3	4	1	5	2	3	9	1
65	1116182	1	1	1	1	2	1	1	1	2
66	1116998	10	4	2	1	3	2	4	3	10
67	1117152	4	1	1	2	1	1	1	1	2
68	1118039	5	3	4	1	8	10	4	9	1
69	1120559	8	3	8	3	4	9	8	9	8
70	1121712	1	1	1	1	2	1	1	1	2
71	1121919	5	1	3	1	2	1	2	1	2
72	1123061	6	10	2	8	10	7	8	10	4
73	1124631	1	3	3	2	2	1	7	1	2
74	1125035	9	4	5	10	10	4	8	1	4
75	1126417	10	6	4	1	4	2	3	3	4
76	1131294	1	1	2	1	2	2	4	2	1
77	1132347	1	1	4	1	2	1	1	1	2
78	1133041	5	3	1	2	2	1	2	1	2
79	1133186	5	1	1	1	1	1	1	1	2
80	1136142	2	1	1	1	3	2	1	1	2
81	1137136	2	2	2	1	1	1	7	1	2
82	1143878	4	1	2	1	2	1	1	1	2
83	1143978	5	2	1	1	2	1	3	1	1
84	1147044	3	1	1	1	2	1	1	1	2
85	1147699	3	5	7	8	8	9	7	10	7
86	1147748	10	6	1	10	4	4	10	10	4
87	1148278	3	3	6	5	8	4	4	1	4
88	1148873	3	6	6	5	10	6	8	3	4
89	1152311	4	1	1	1	2	1	1	1	2
90	1155546	2	1	1	2	3	1	2	1	2

RStudio interface showing a script for data sampling and a console output of a k-Nearest Neighbor Classifier model.

```
rm(list=ls()) # clears the environment , removes all the objects
# Question 1.1
# Import data
# View(Cancerdataset)
# Index
# index=sample(nrow(Cancerdataset),size =0.25*(nrow(Cancerdataset)),replace=TRUE, prob=NULL )
# test=Cancerdataset[index,]
# test
```

Sample	F1	F2	F3	F4	F5	F6	F7	F8	F9	Class	
469	1105324	4	1	1	1	2	1	1	1	2	
32	1071760	2	1	1	1	2	1	1	1	2	
677	1463128	1	1	1	1	2	1	1	1	2	
281	486193	3	1	1	1	2	1	1	1	2	
645	1299196	2	1	1	1	2	1	1	1	2	
549	1240603	3	1	1	1	2	1	1	1	2	
184	1203138	5	8	8	8	5	10	7	8	1	4
339	1135090	4	1	1	1	2	1	1	1	2	
629	1213273	2	1	1	1	2	1	1	1	2	
677.1	1163128	1	1	1	1	2	1	1	1	2	
235	1236043	3	1	1	1	2	1	1	1	2	
598	1133063	5	1	1	1	2	1	1	1	2	
132	1189566	7	2	4	1	6	10	5	4	1	2
425	1257938	3	1	1	1	2	1	1	1	2	
47	1070322	2	1	1	1	2	1	1	1	2	
409	1236837	2	3	2	2	2	3	1	1	2	
129	1177399	8	3	5	4	5	10	1	6	2	4
676	1180075	4	1	1	1	2	1	1	1	2	
129.1	1177399	8	3	5	4	5	10	1	6	2	4
353	1468632	3	1	1	1	2	1	1	1	2	
25	1059552	1	1	1	1	2	1	1	1	2	
304	646804	1	1	1	1	2	1	1	1	2	
446	627627	2	1	1	1	2	1	1	1	2	
389	1114570	2	1	1	1	2	1	1	2	2	1
199	1214092	1	1	1	1	2	1	1	1	1	2
420	1253505	2	3	1	1	5	1	1	1	1	2
581	671549	1	1	1	1	2	1	1	1	1	2
358	632930	8	10	10	7	10	10	7	1	1	2
143	1183983	9	5	5	4	4	5	4	3	3	4
62	1116316	9	10	10	1	10	8	2	9	1	2
304.1	646804	1	1	1	1	2	1	1	1	1	2
341	614265	2	1	1	1	2	1	1	1	1	2
505	1268766	1	1	1	1	2	1	1	1	1	2
582	178358	5	7	10	6	5	10	7	5	1	4
478	1296063	4	1	1	1	2	1	1	1	1	2
649	1315807	5	10	10	10	2	10	10	10	4	4
365	896404	2	1	1	1	2	1	1	1	1	2

RStudio interface showing a script for data sampling and a console output of a k-Nearest Neighbor Classifier model.

```
rm(list=ls()) # clears the environment , removes all the objects
# Question 1.1
# Import data
# View(Cancerdataset)
# Index
# index=sample(nrow(Cancerdataset),size =0.25*(nrow(Cancerdataset)),replace=TRUE, prob=NULL )
# test=Cancerdataset[index,]
# test
```

Sample	F1	F2	F3	F4	F5	F6	F7	F8	F9	Class	
469	1105324	4	1	1	1	2	1	1	1	2	
32	1071760	2	1	1	1	2	1	1	1	2	
677	1463128	1	1	1	1	2	1	1	1	2	
281	486193	3	1	1	1	2	1	1	1	2	
645	1299196	2	1	1	1	2	1	1	1	2	
549	1240603	3	1	1	1	2	1	1	1	2	
184	1203138	5	8	8	8	5	10	7	8	1	4
339	1135090	4	1	1	1	2	1	1	1	2	
629	1213273	2	1	1	1	2	1	1	1	2	
677.1	1163128	1	1	1	1	2	1	1	1	2	
235	1236043	3	1	1	1	2	1	1	1	2	
598	1133063	5	1	1	1	2	1	1	1	2	
132	1189566	7	2	4	1	6	10	5	4	1	2
425	1257938	3	1	1	1	2	1	1	1	2	
47	1070322	2	1	1	1	2	1	1	1	2	
409	1236837	2	3	2	2	2	3	1	1	2	
129	1177399	8	3	5	4	5	10	1	6	2	4
676	1180075	4	1	1	1	2	1	1	1	2	
129.1	1177399	8	3	5	4	5	10	1	6	2	4
353	1468632	3	1	1	1	2	1	1	1	2	
25	1059552	1	1	1	1	2	1	1	1	2	
304	646804	1	1	1	1	2	1	1	1	2	
446	627627	2	1	1	1	2	1	1	1	2	
389	1114570	2	1	1	1	2	1	1	2	2	1
199	1214092	1	1	1	1	2	1	1	1	1	2
420	1253505	2	3	1	1	5	1	1	1	1	2
581	671549	1	1	1	1	2	1	1	1	1	2
358	632930	8	10	10	7	10	10	7	1	1	2
143	1183983	9	5	5	4	4	5	4	3	3	4
62	1116316	9	10	10	1	10	8	2	9	1	2
304.1	646804	1	1	1	1	2	1	1	1	1	2
341	614265	2	1	1	1	2	1	1	1	1	2
505	1268766	1	1	1	1	2	1	1	1	1	2
582	178358	5	7	10	6	5	10	7	5	1	4
478	1296063	4	1	1	1	2	1	1	1	1	2
649	1315807	5	10	10	10	2	10	10	10	4	4
365	896404	2	1	1	1	2	1	1	1	1	2

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source Editor: `rm(list=ls()) # clears the environment, removes all the objects`

Console: `## reached getOption("max.print") -- omitted 84 rows`

Environment: `cancerdataset` (699 obs. of 11 variables), `test` (174 obs. of 11 variables), `trainingdata` (548 obs. of 11 variables)

Global Environment: `Index` (1:174) 469 32 677 281 645 549 184 539 629 677 ...

Weighted k-Nearest Neighbor Classifier

Description: Performs k-nearest neighbor classification of a test set using a training set. For each row of the test set, the k nearest training set vectors (according to Minkowski distance) are found, and the classification is done via the maximum of summed kernel densities. In addition even ordinal and continuous variables can be predicted.

Usage: `knkn(formula = formula(train), train, test, na.action = na.omit(), k = 7, distance = 2, kernel = "optimal", ykernel = NULL, scale=TRUE, contrasts = c("unordered" = "contr.dummy", ordered = "contr.ordinal"))`

Arguments: `formula` A formula object, `train` Matrix or data frame of training set cases, `test` Matrix or data frame of test set cases, `learn` Matrix or data frame of training set cases, `valid` Matrix or data frame of test set cases, `na.action` A function which indicates what should happen when the data contain NA's, `k` Number of neighbors considered.

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Source Editor: `rm(list=ls()) # clears the environment, removes all the objects`

Console: `## reached getOption("max.print") -- omitted 84 rows`

Environment: `cancerdataset` (699 obs. of 11 variables), `test` (174 obs. of 11 variables), `trainingdata` (548 obs. of 11 variables)

Global Environment: `Index` (1:174) 469 32 677 281 645 549 184 539 629 677 ...

Weighted k-Nearest Neighbor Classifier

Description: Performs k-nearest neighbor classification of a test set using a training set. For each row of the test set, the k nearest training set vectors (according to Minkowski distance) are found, and the classification is done via the maximum of summed kernel densities. In addition even ordinal and continuous variables can be predicted.

Usage: `knkn(formula = formula(train), train, test, na.action = na.omit(), k = 7, distance = 2, kernel = "optimal", ykernel = NULL, scale=TRUE, contrasts = c("unordered" = "contr.dummy", ordered = "contr.ordinal"))`

Arguments: `formula` A formula object, `train` Matrix or data frame of training set cases, `test` Matrix or data frame of test set cases, `learn` Matrix or data frame of training set cases, `valid` Matrix or data frame of test set cases, `na.action` A function which indicates what should happen when the data contain NA's, `k` Number of neighbors considered.

RStudio interface showing a script editor, console, and environment pane.

Script Editor:

```
rm(list=ls()) # clears the environment, removes all the objects
# Random Forest
# Training data
```

Console:

```
49 1106095 4 1 1 3 2 1 3 1 1 2
50 1106839 7 8 7 2 4 3 2 5 2 4
51 1108370 9 5 8 1 2 3 2 1 5 4
52 1108649 3 3 4 2 2 2 4 1 4
53 1110192 10 3 6 2 3 5 4 10 2 4
54 1110593 5 5 5 8 10 8 7 3 7 4
55 1113038 8 2 4 1 5 1 5 4 4 4
59 1113483 5 2 3 1 6 10 5 1 1 4
60 1113966 9 5 5 2 2 2 1 1 4
61 1115282 5 3 5 5 3 4 10 1 4
62 1115293 1 1 1 1 2 2 1 1 2
64 1116112 6 3 4 1 5 2 3 5 1 4
65 1116192 1 1 1 1 2 1 2 1 2
67 1117152 4 1 1 1 2 1 1 1 2
68 1118039 5 3 4 1 8 10 4 9 1 4
69 1120519 8 3 8 3 4 9 8 9 8
71 1121949 5 1 1 1 2 1 1 1 2
73 1124631 1 3 3 2 2 1 7 2 1 2
74 1125035 9 4 3 10 8 10 8 1 4
75 1126417 10 6 4 1 3 4 3 2 3 4
77 1127287 1 4 4 2 3 2 1 2
78 1133081 5 3 1 2 2 2 2 1 1 2
79 1133136 3 1 1 1 2 3 3 1 1 2
80 1136142 2 1 1 1 3 1 2 1 2
81 1137136 2 2 2 1 1 1 7 1 1 2
82 1143978 4 1 1 2 2 1 2 1 2
83 1143978 5 2 1 1 2 1 3 1 1 2
84 1147044 5 1 1 1 2 2 7 1 1 2
86 1147748 5 10 6 1 10 4 4 10 10 4
88 1148873 3 6 6 6 5 10 6 8 3 4
89 1152331 4 1 1 1 2 1 1 1 2
91 1156272 1 1 1 1 2 1 3 1 1 2
92 1159948 9 1 1 2 2 1 1 1 1 2
93 1157734 4 1 1 1 2 1 3 1 1 2
95 1160476 2 1 1 1 2 1 3 1 1 2
96 1164666 1 1 1 1 2 1 3 1 1 2
98 1165790 5 1 1 1 2 1 3 1 1 2
99 1165926 9 6 9 2 10 6 9 10 4
100 1166630 7 5 6 10 5 10 7 9 4 4
101 1166654 10 3 5 1 10 5 3 10 2 4
102 1167439 2 3 4 4 2 5 5 1 4
105 1168736 10 10 10 10 1 8 8 8 4
106 1169049 7 3 4 4 3 2 7 4
107 1170419 10 10 10 8 2 10 4 1 1 4
108 1170420 1 6 8 10 6 10 5 7 1 4
109 1171710 1 1 1 1 2 1 2 3 1 2
110 1171710 6 5 4 4 3 9 7 8 3 4
111 1171795 1 3 1 2 1 2 1 2 2 2
[ reached getOption("max.print") -- omitted 458 rows ]
> view(trainingdata)
```

Environment:

- cancerdataset: 699 obs. of 11 variables
- test: 174 obs. of 11 variables
- trainingdata: 548 obs. of 11 variables

Global Environment:

index: fnc [1:174] 469 32 677 281 645 549 184 539 629 677 ...

Files: Plots Packages Help Viewer

R Weighted k-Nearest Neighbor Classifier

knn (knn) R Documentation

Weighted k-Nearest Neighbor Classifier

Description

Performs k-nearest neighbor classification of a test set using a training set. For each row of the test set, the k nearest training set vectors (according to Minkowski distance) are found, and the classification is done via the maximum of summed kernel densities. In addition even ordinal and continuous variables can be predicted.

Usage

```
knn(formula = formula(train), train, test, na.action = na.omit(),
    k = 7, distance = 2, kernel = "optimal", ykernel = NULL, scale=TRUE,
    contrasts = c("unordered" = "contr.dummy", ordered = "contr.ordinal"))
knn.dist(learn, valid, k = 10, distance = 2)
```

Arguments

Argument	Description
formula	A formula object
train	Matrix or data frame of training set cases.
test	Matrix or data frame of test set cases.
learn	Matrix or data frame of training set cases.
valid	Matrix or data frame of test set cases.
na.action	A function which indicates what should happen when the data contain NA's.
k	Number of neighbors considered.