Raport 4

Eksploracja danych

Mikołaj Langner, Marcin Kostrzewa
nr albumów: 255716, 255749

2021-05-28

Spis treści

1 Wstęp

2 Zadanie 1

2.1 a)

2.2 b)

```
##  Setting default kernel parameters
```

```
##  Setting default kernel parameters
```

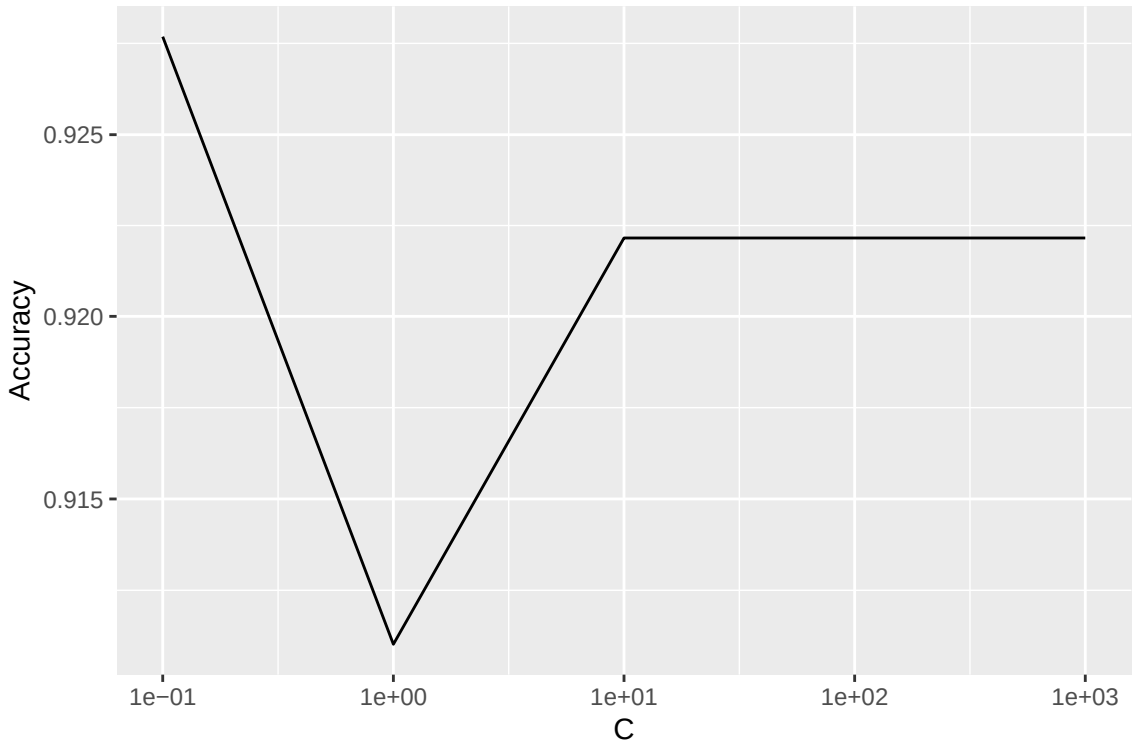
```
##  Setting default kernel parameters
```

```
##  Setting default kernel parameters
```

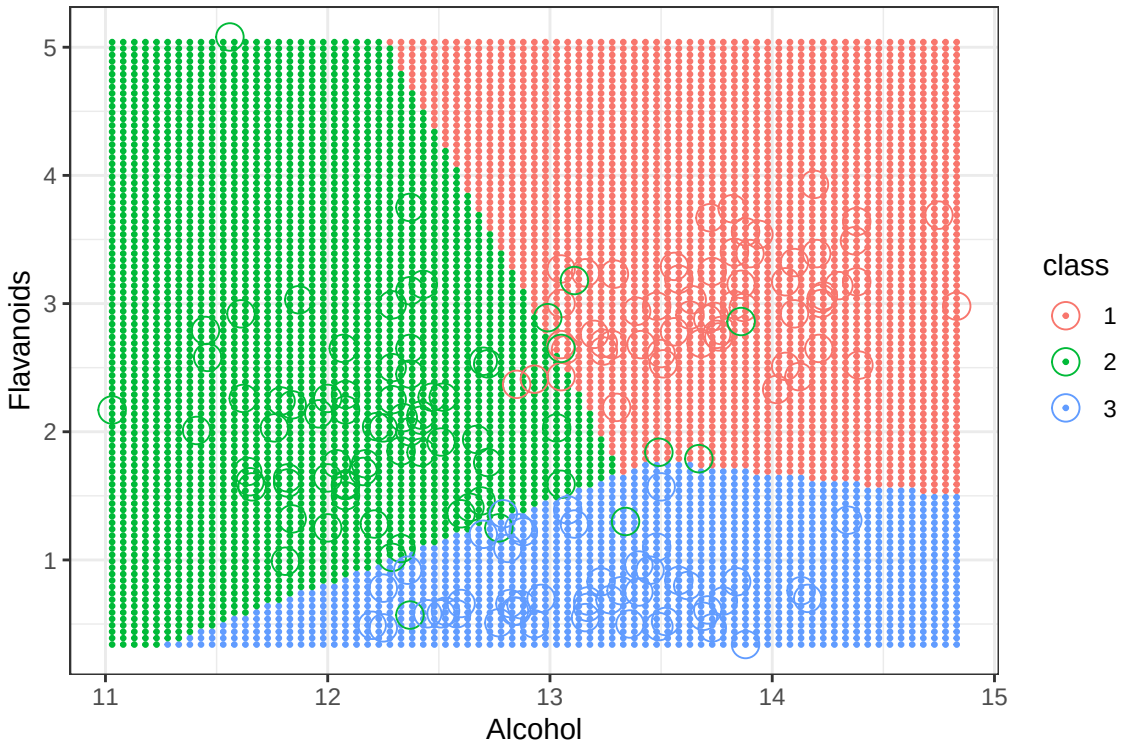
```
##  Setting default kernel parameters
```

% latex table generated in R 3.6.1 by xtable 1.8-4 package % Sat Jun 19 02:11:06 2021

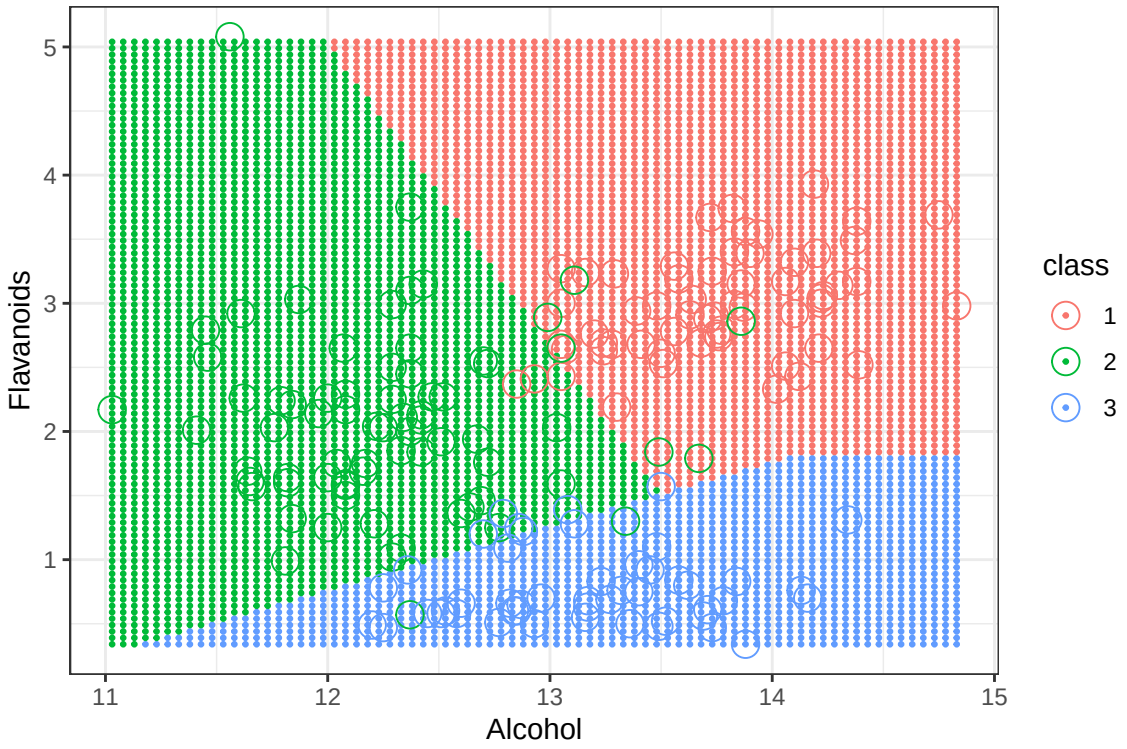
% latex table generated in R 3.6.1 by xtable 1.8-4 package % Sat Jun 19 02:11:11 2021



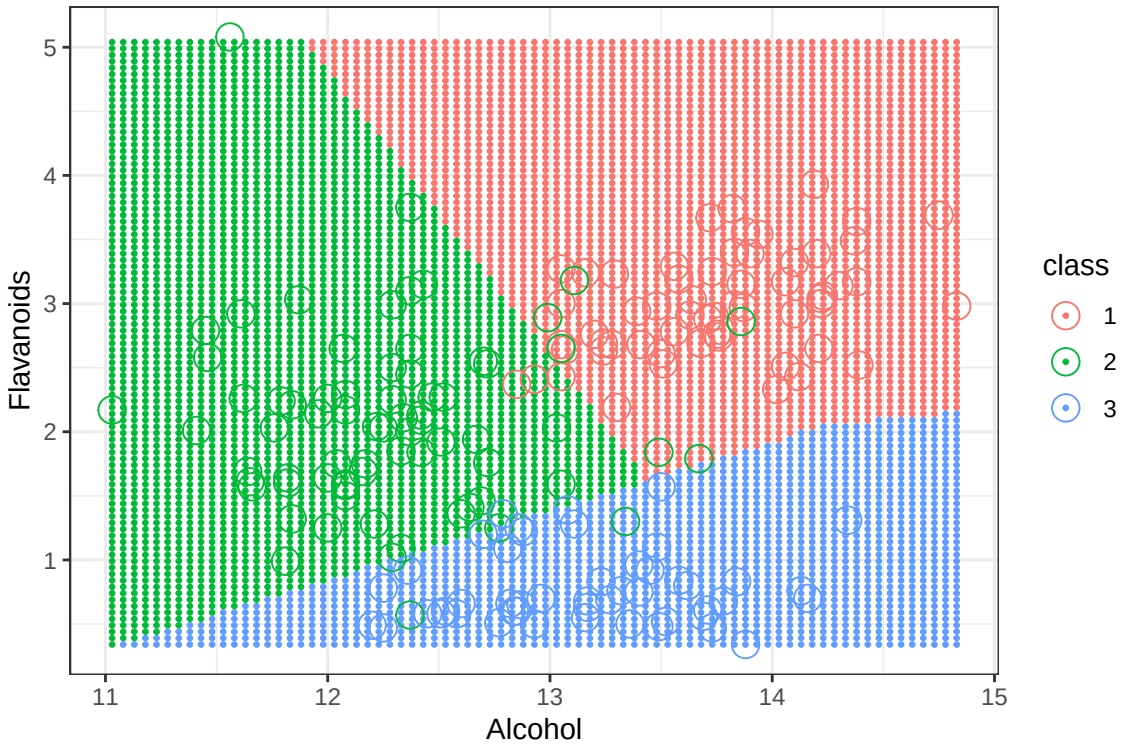
Rysunek 1: Dokladnosc klasyfikatora od parametru kosztu



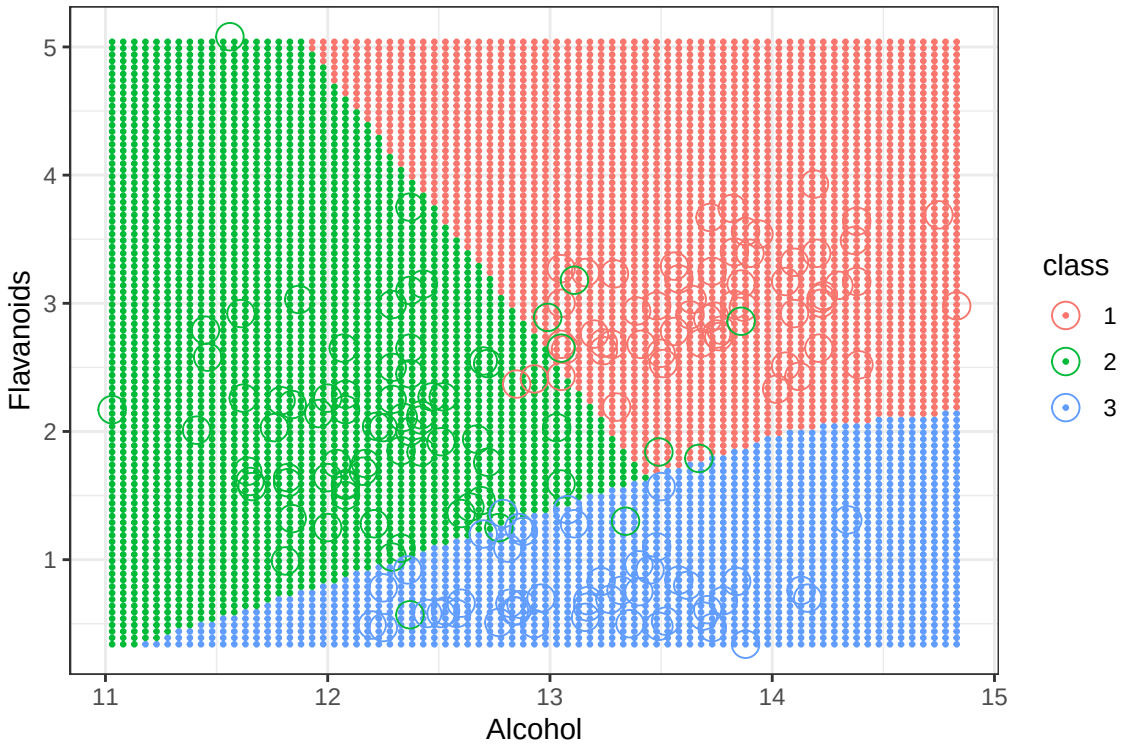
Rysunek 2: Obszary decyzyjne dla $C = 0.1$



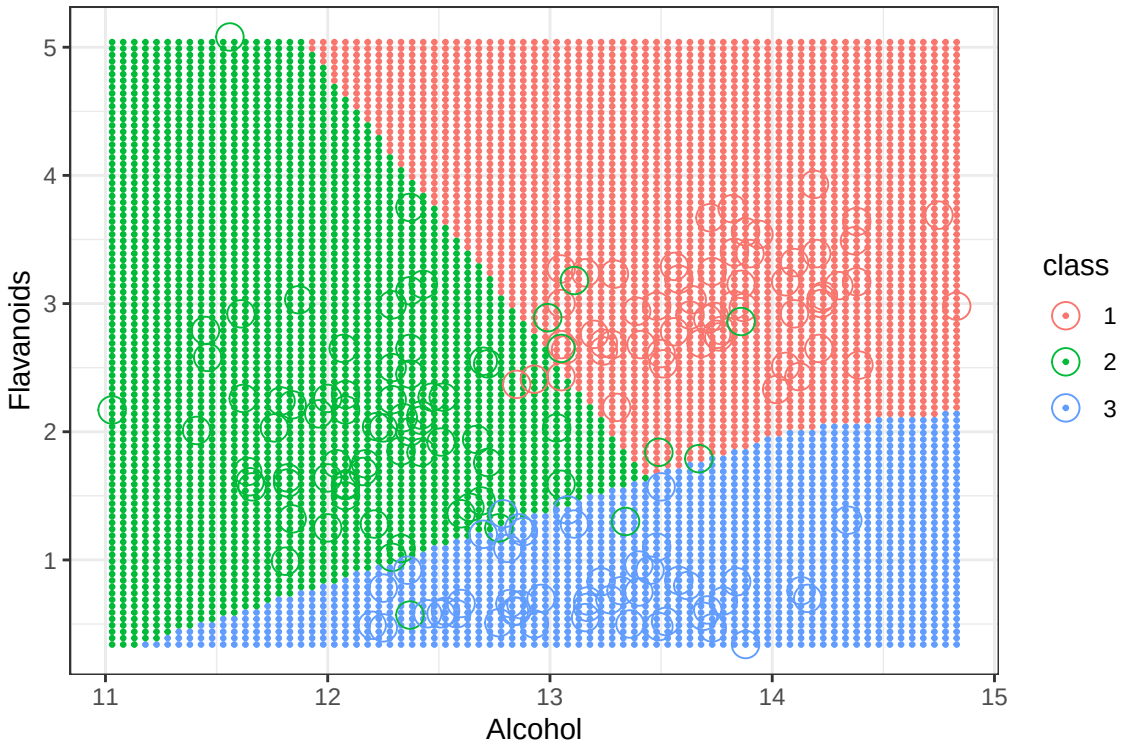
Rysunek 3: Obszary decyzyjne dla $C = 1$



Rysunek 4: Obszary decyzyjne dla $C = 10$



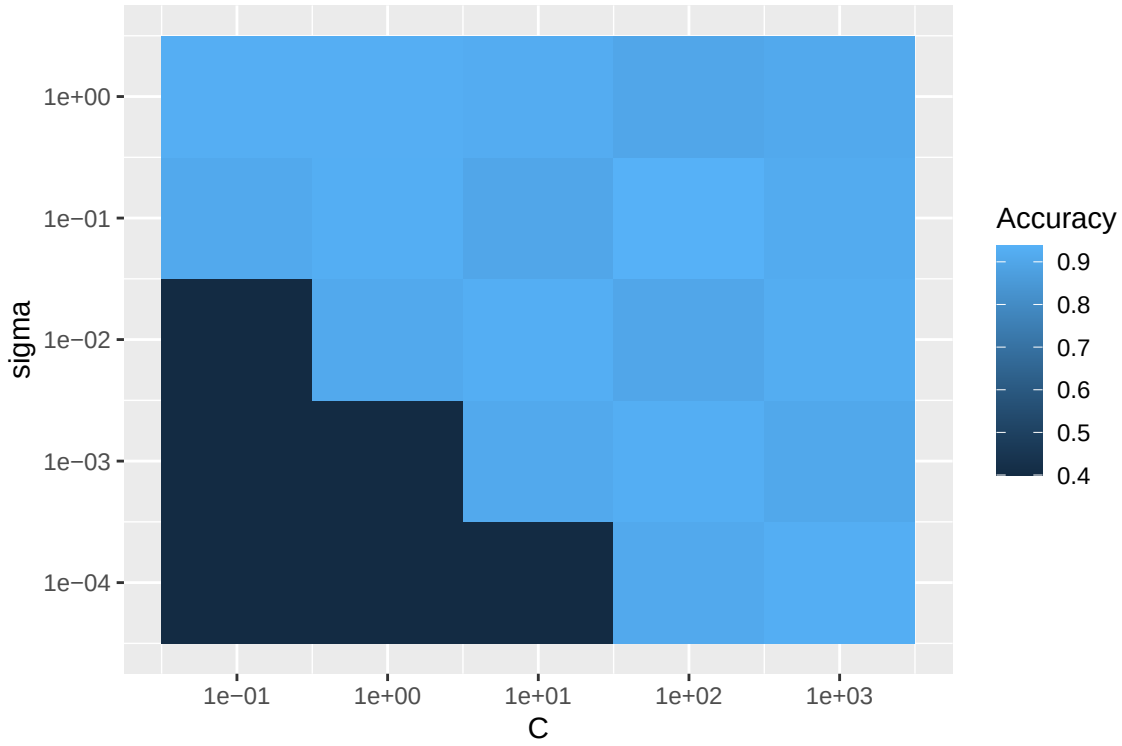
Rysunek 5: Obszary decyzyjne dla $C = 100$



Rysunek 6: Obszary decyzyjne dla $C = 1000$

linear	polynomial	radial
0.922	0.926	0.944

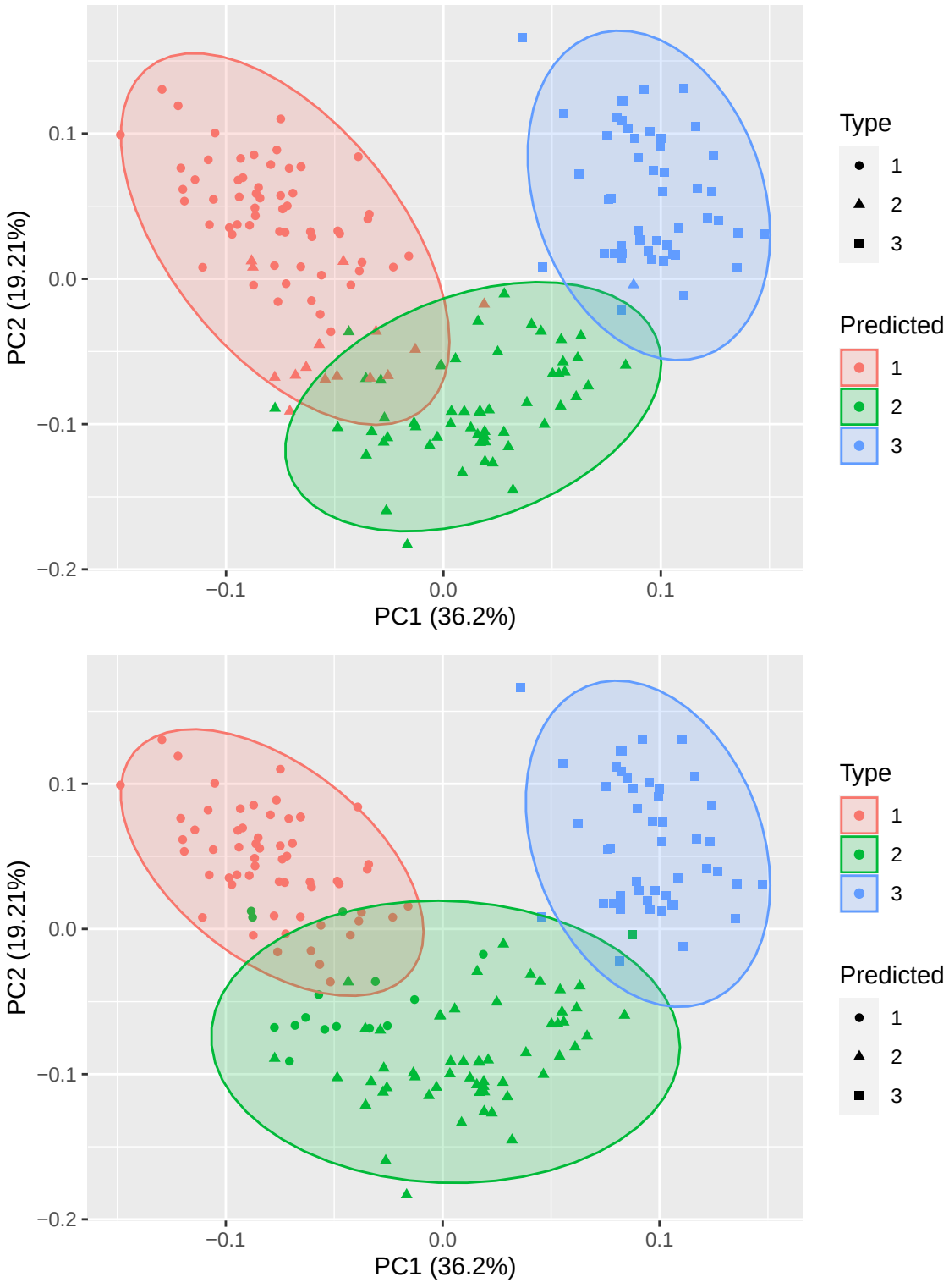
Tabela 1: Porównanie klasyfikatorów dla róznych jader



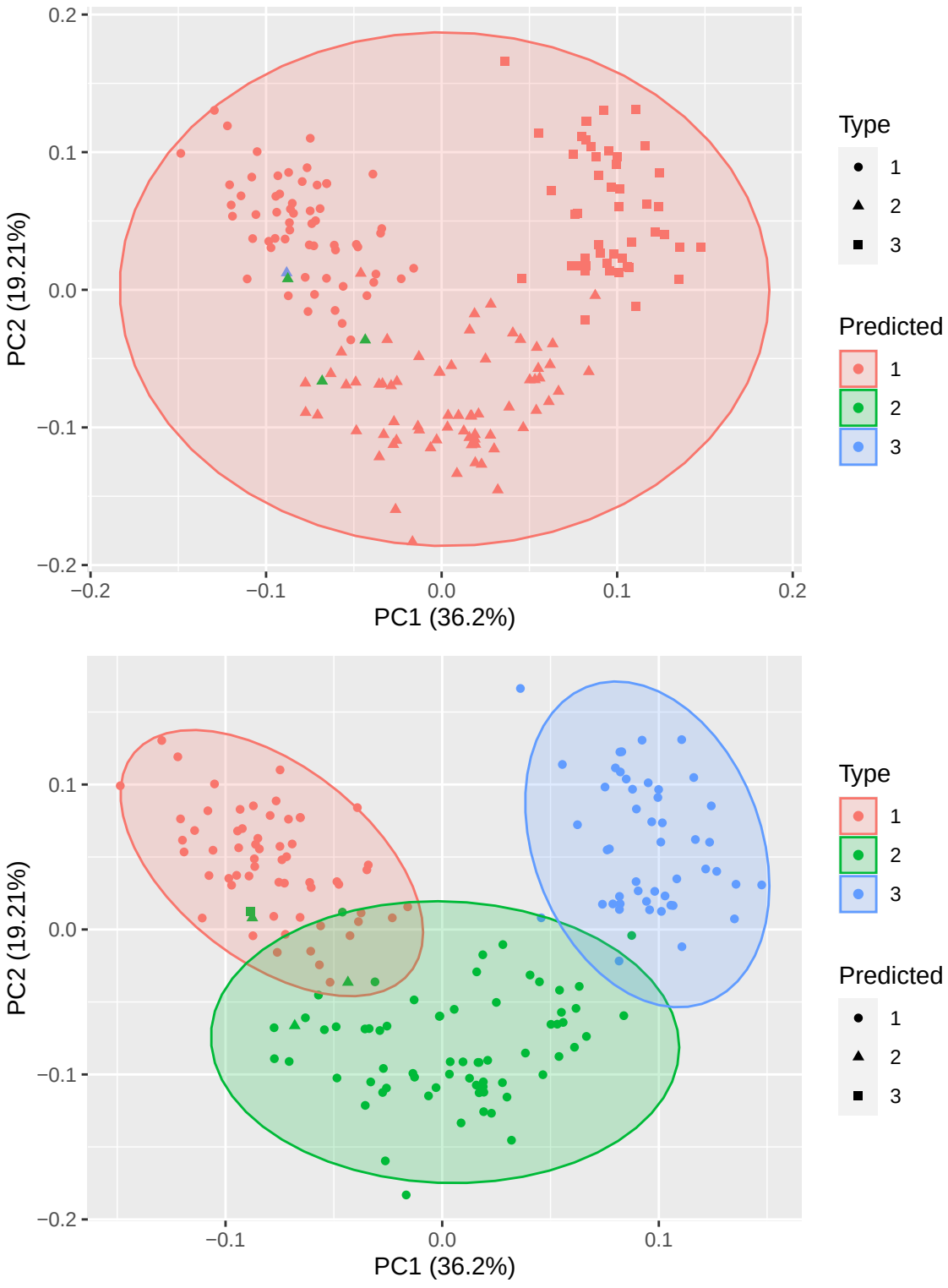
Rysunek 7: Mapa ciepla dokladnosci klasyfikatora

3 Zadanie 2

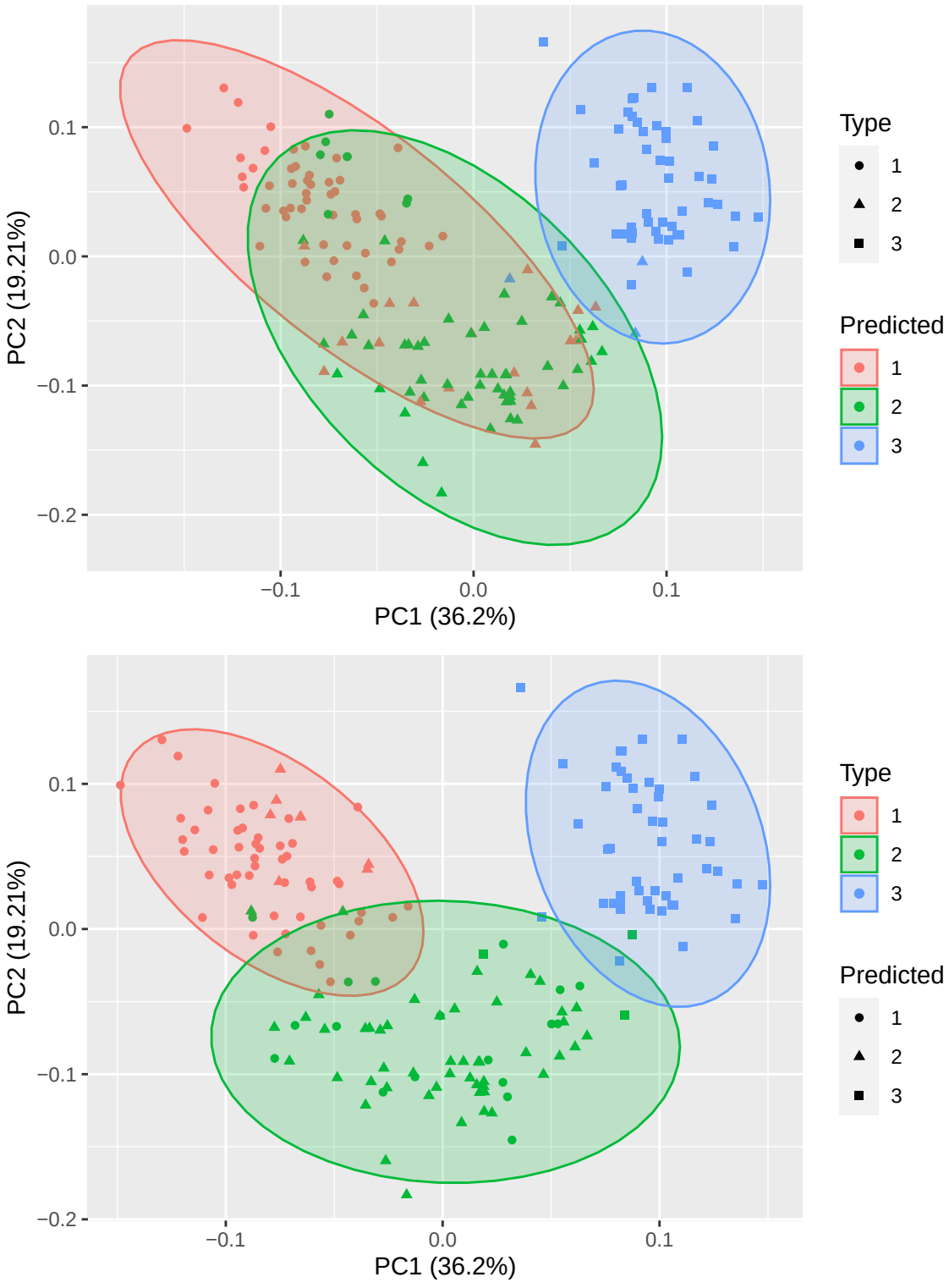
```
##
## Clustering Methods:
##  agnes pam
##
## Cluster sizes:
##  2 3 4 5 6 7 8 9 10
##
## Validation Measures:
##                            2       3       4       5       6       7       8       9
##
## agnes Connectivity   4.4925  8.0972 12.8210 17.7913 21.4591 22.9877 25.8044 30.6730 3
##       Dunn           0.0374  0.0227  0.0417  0.0347  0.0368  0.0544  0.0561  0.0656
##       Silhouette     0.6413  0.5419  0.5336  0.4806  0.4824  0.5075  0.5055  0.5024
## pam   Connectivity   1.5286  5.1048 16.2798 20.0643 23.1155 27.8393 31.0163 33.5841 3
##       Dunn           0.0434  0.0229  0.0340  0.0340  0.0233  0.0502  0.0478  0.0359
```



Rysunek 8: Skupienia dla metody PAM



Rysunek 9: Skupienia dla metody AGNES z single-linkage

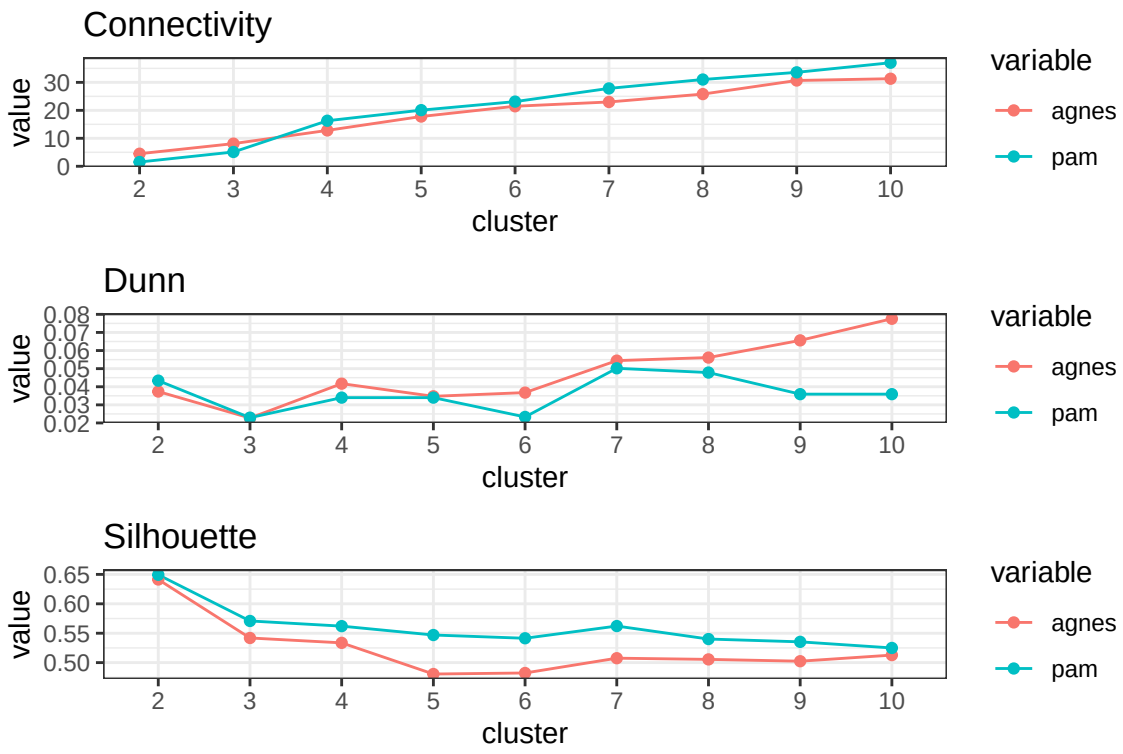


Rysunek 10: Skupienia dla metody AGNES z complete-linkage

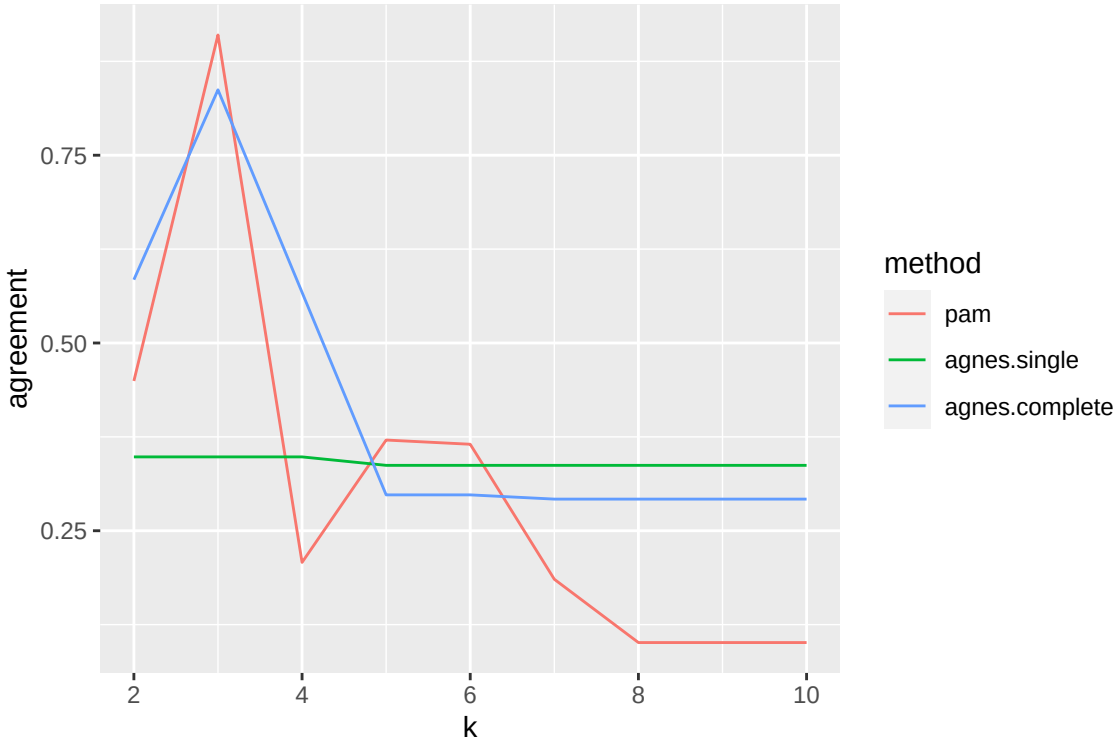
sigma	C
0.10	100.00

Tabela 2: Parametry dla najlepszego klasyfikatora

```
##       Silhouette     0.6494  0.5708  0.5620  0.5469  0.5414  0.5622  0.5401  0.5353
##
## Optimal Scores:
##
##              Score  Method Clusters
## Connectivity 1.5286 pam    2
## Dunn         0.0776 agnes  10
## Silhouette   0.6494 pam    2

##                   Score Method Clusters
## Connectivity 1.52857143    pam        2
## Dunn         0.07755693  agnes       10
## Silhouette   0.64936476    pam        2

## Using cluster as id variables
## Using cluster as id variables
## Using cluster as id variables
```

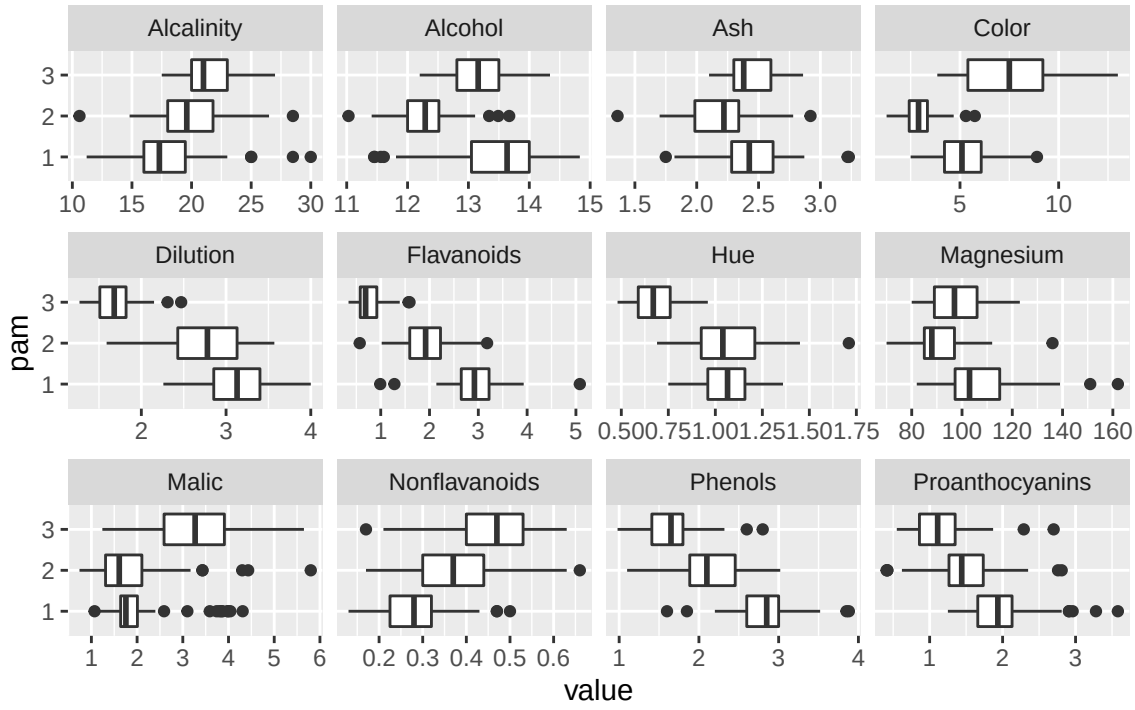


Rysunek 11: Wskazniki wewnetrzne dla PAM i AGNES z complete-linkage

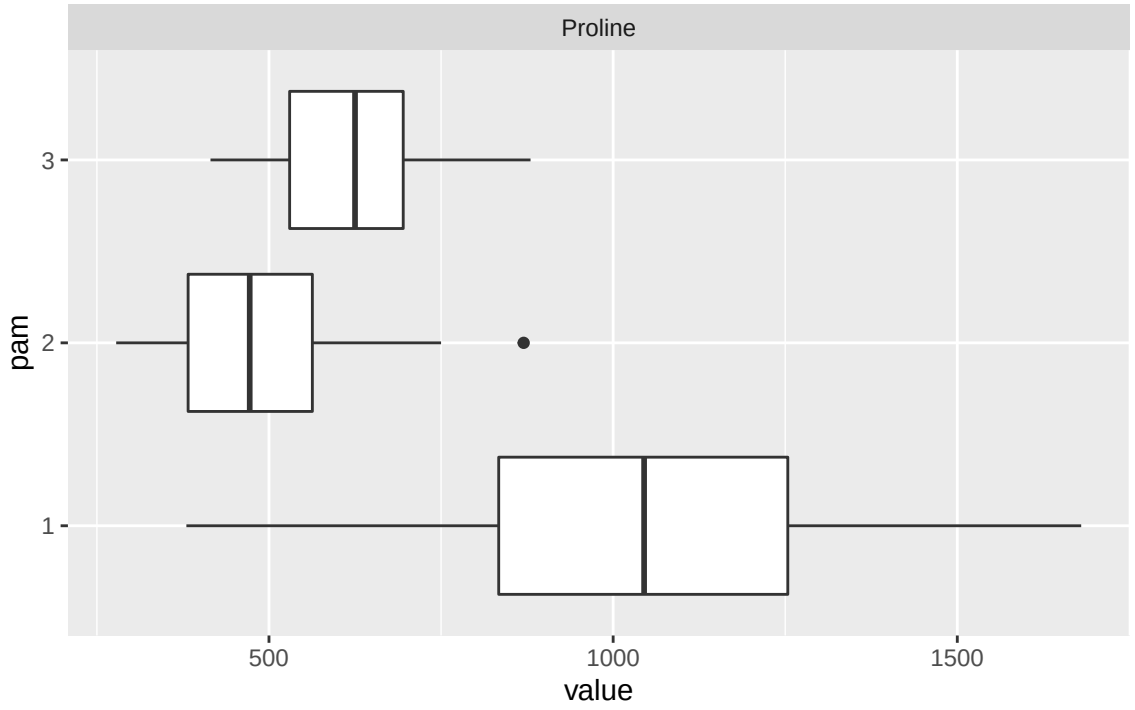


Rysunek 12: Porównanie wskazników zewnetrznych

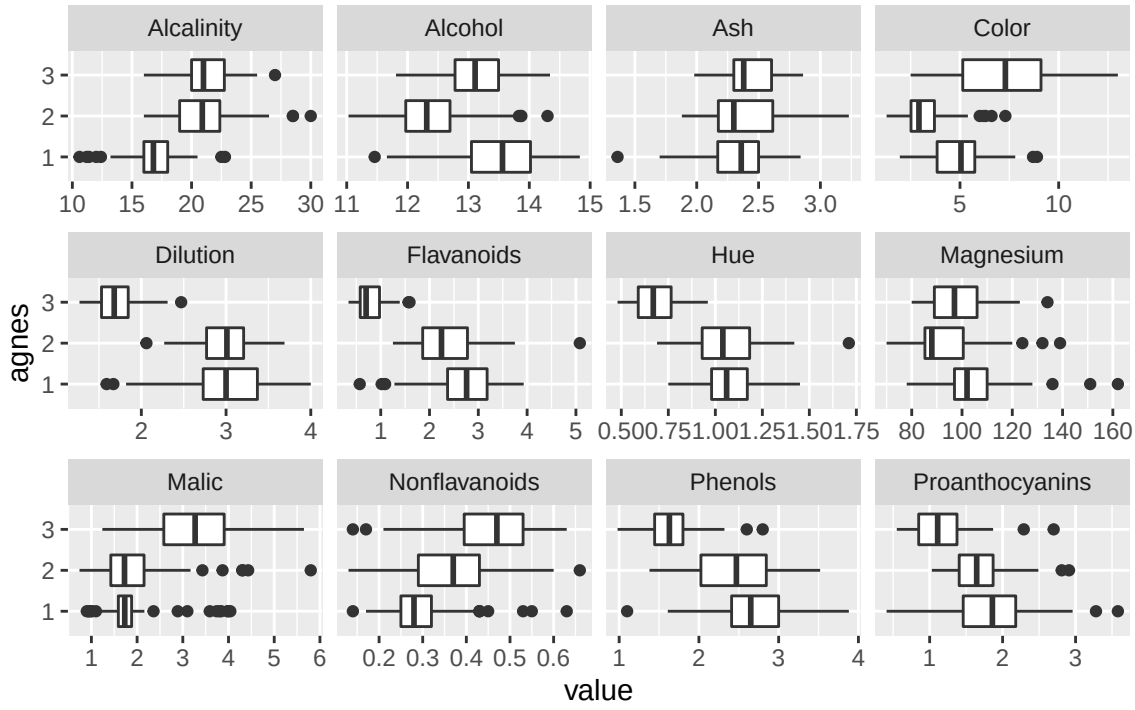
```
## Cases in matched pairs: 80.9 %

## 1 2 3
## 1 2 3
```


Alcalinity
Alcohol
Ash
Color
Dilution
Flavanoids
Hue
Magnesium
Malic
Nonflavanoids
Phenols
Proanthocyanins
pam
value

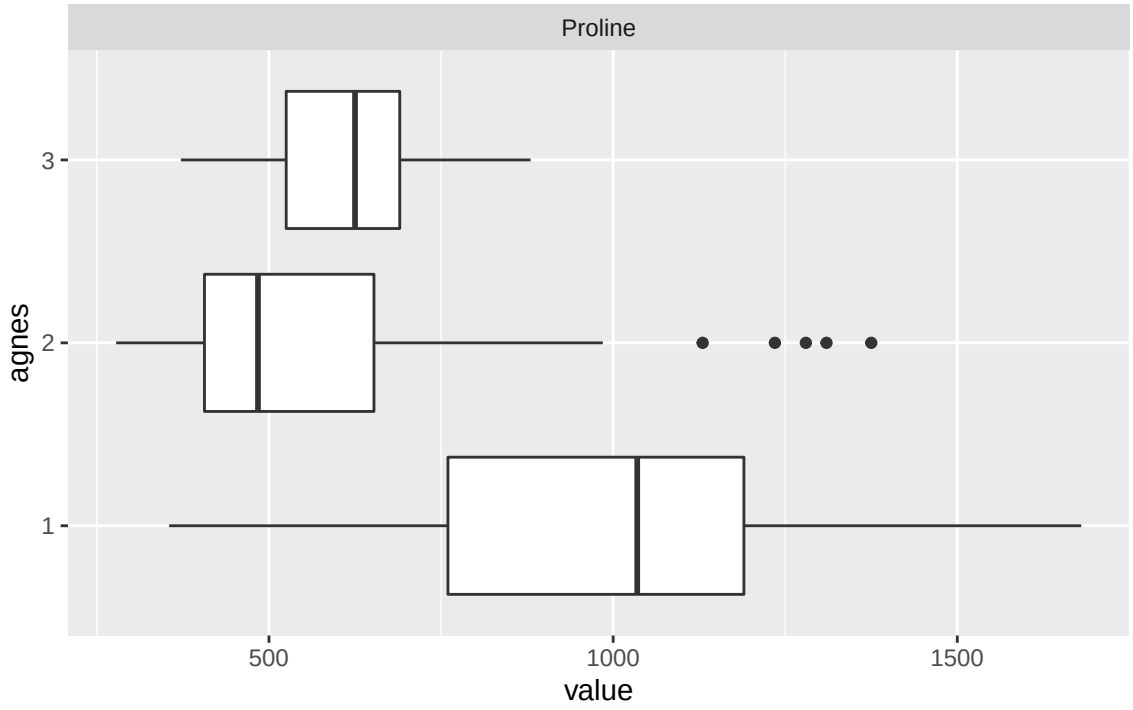
Page 1


Proline
pam
value

Page 2



Page 1



Page 2

% latex table generated in R 3.6.1 by xtable 1.8-4 package % Sat Jun 19 02:11:23 2021

	1	2	3
Alcohol	0.59	-0.92	0.39
Malic	-0.47	-0.54	0.81
Ash	0.16	-0.90	0.05
Alcalinity	0.30	-0.15	0.60
Magnesium	0.02	-1.38	-0.54
Phenols	0.65	-1.03	-0.58
Flavanoids	0.95	0.00	-1.27
Nonflavanoids	-0.82	0.07	0.71
Proanthocyanins	0.47	0.07	-0.60
Color	0.02	-0.72	1.45
Hue	0.36	0.19	-1.78
Dilution	1.21	0.79	-1.40
Proline	0.55	-0.75	-0.31

Tabela 3: Medoidy dla metody PAM przy K=3