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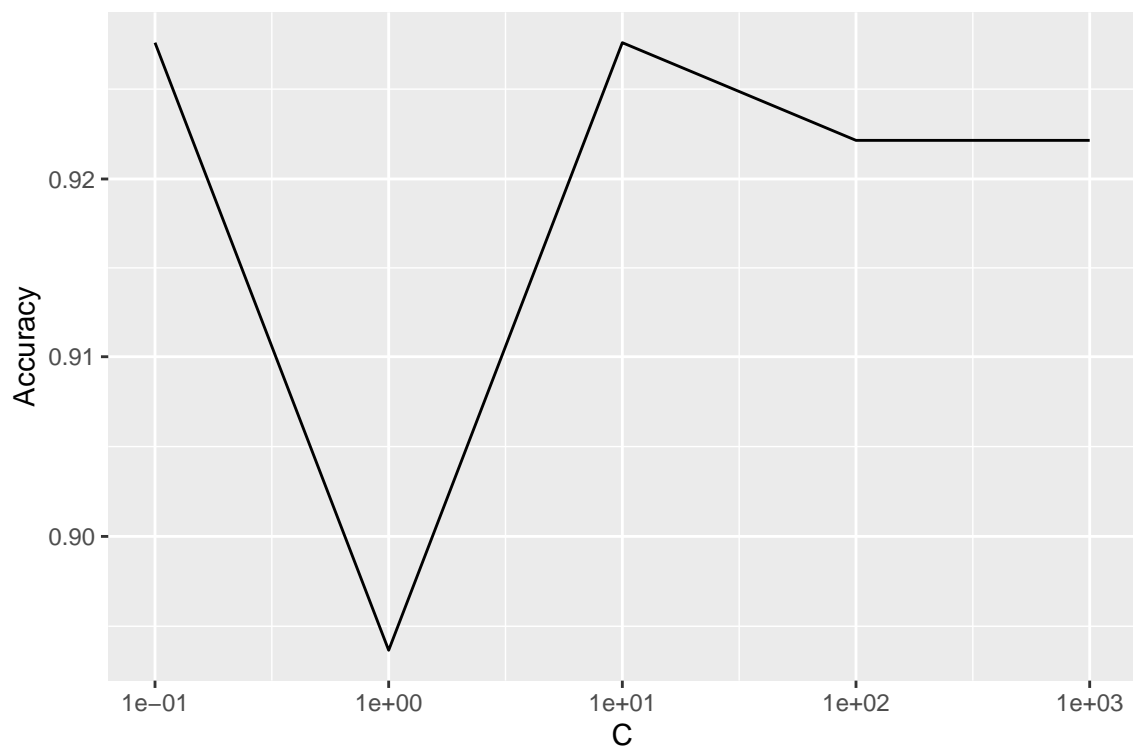
```
## Setting default kernel parameters
```

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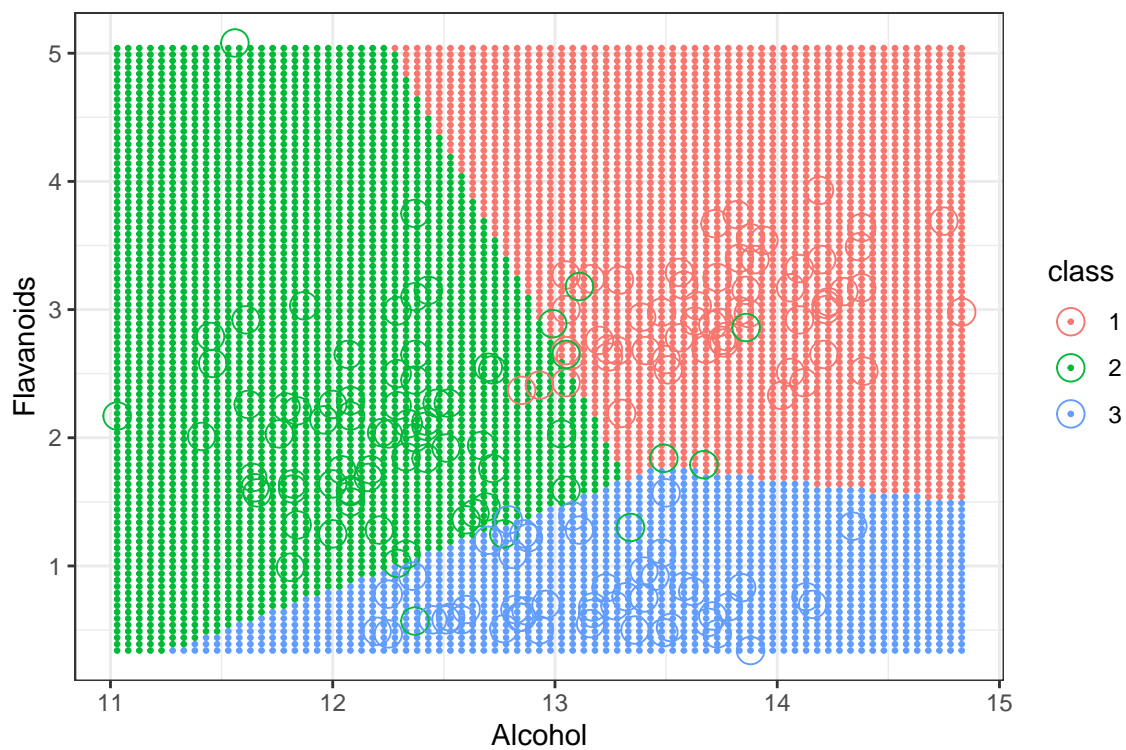
```
## Setting default kernel parameters
```

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

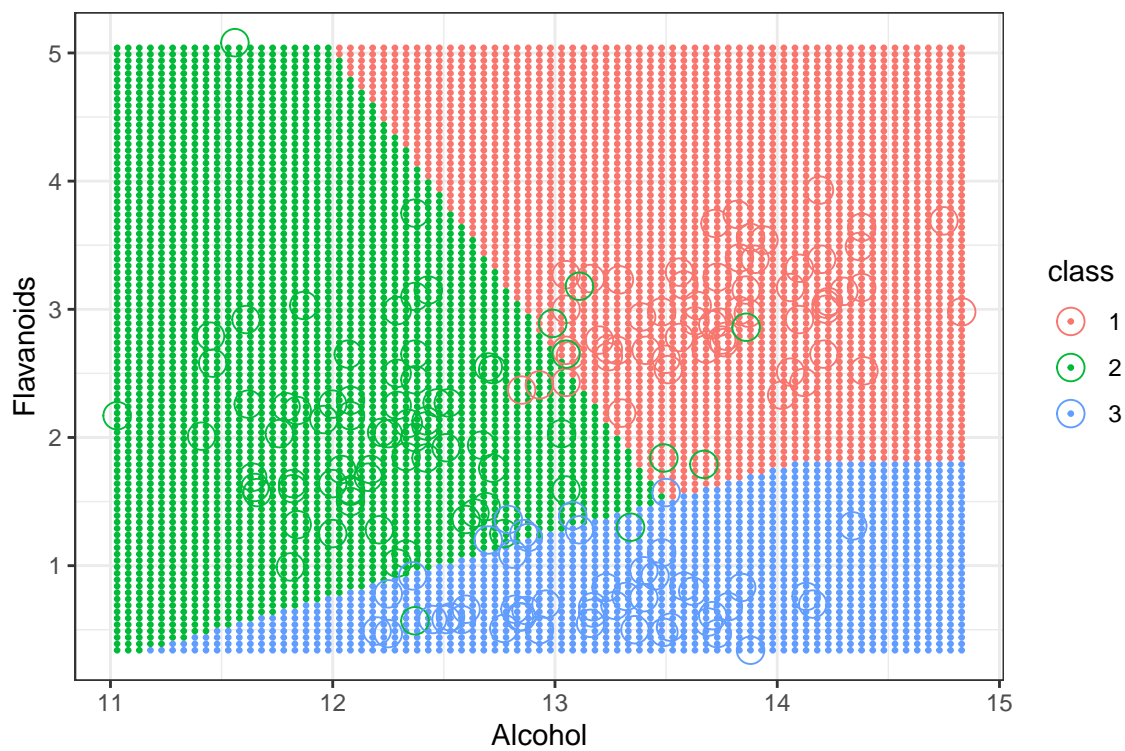% latex table generated in R 3.6.1 by xtable 1.8-4 package % Sat Jun 19 00:11:40 2021

% latex table generated in R 3.6.1 by xtable 1.8-4 package % Sat Jun 19 00:11:45 2021
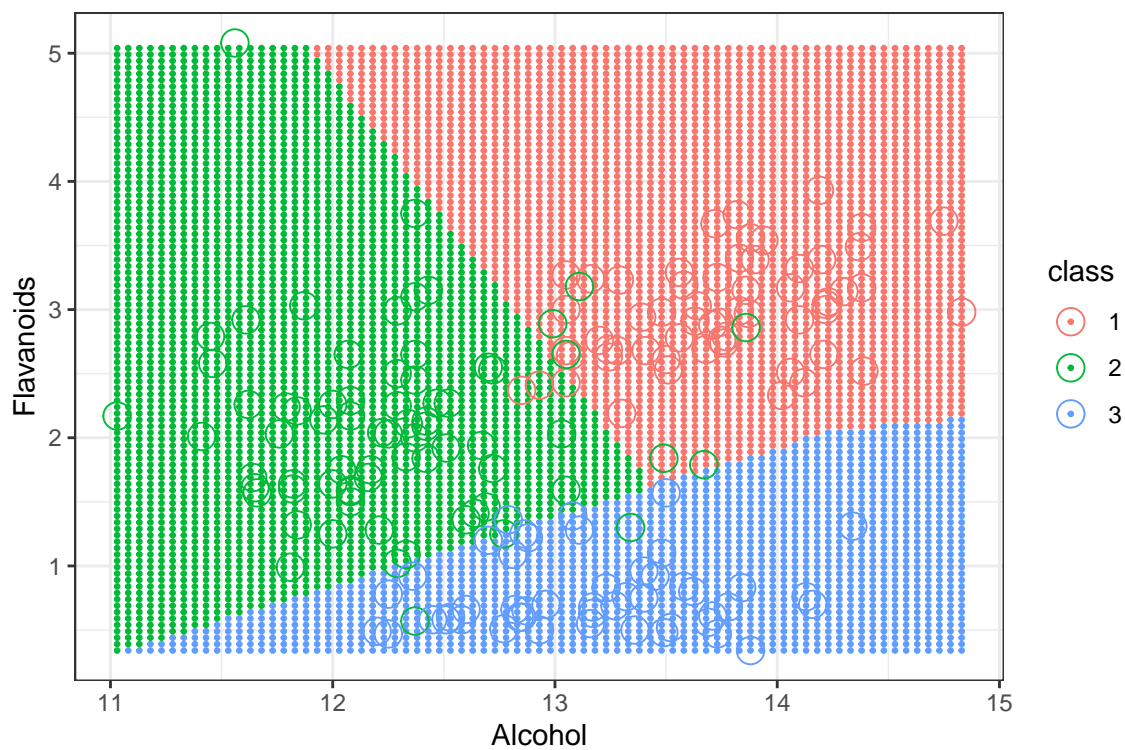
Rysunek 1: Dokladnosc klasyfikatora od parametru kosztu
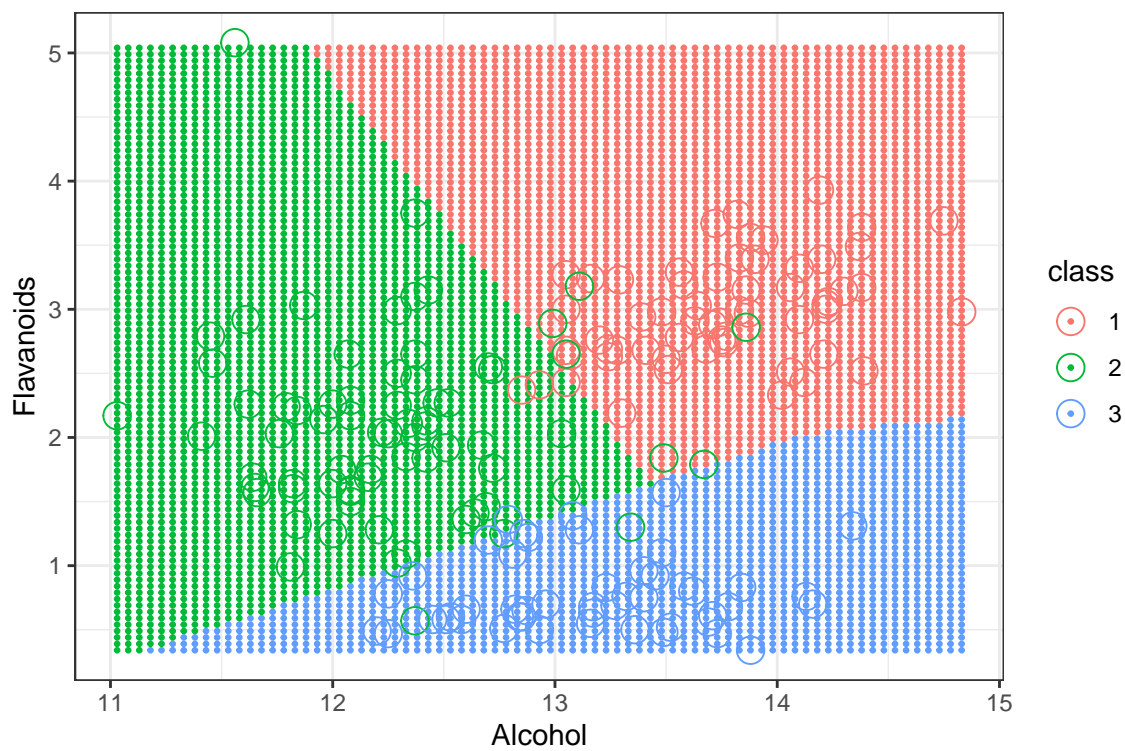

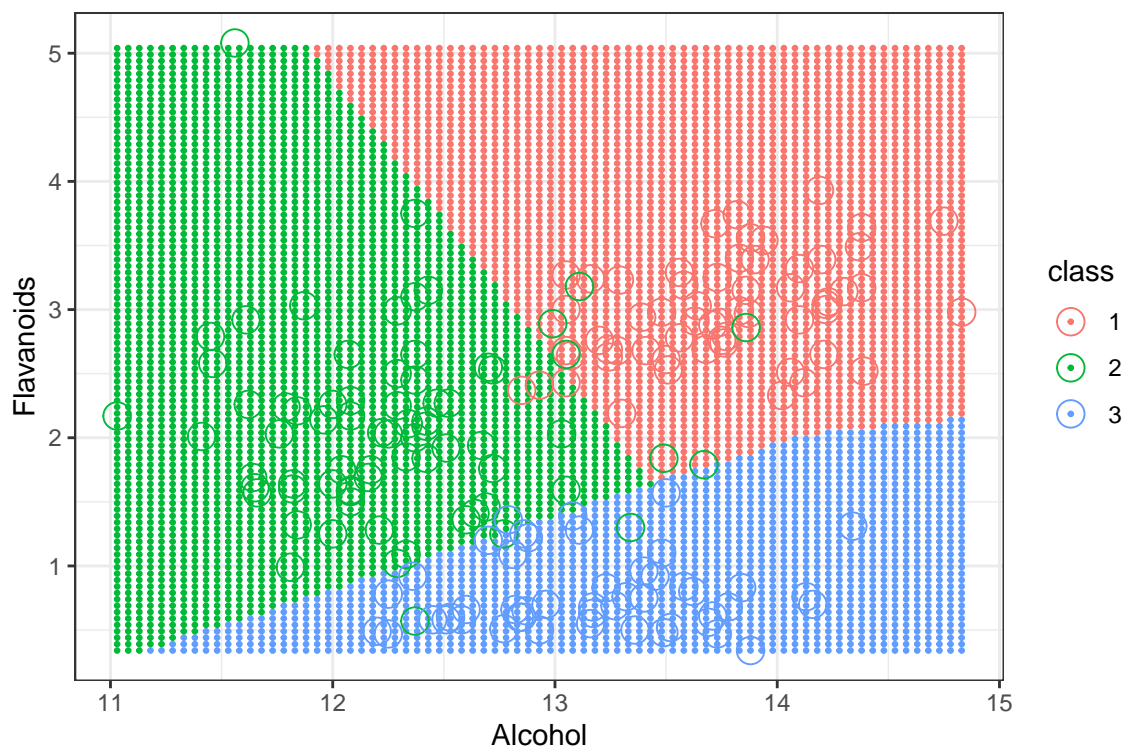Rysunek 2: Obszary decyzyjne dla $C = 0.1$

Rysunek 3: Obszary decyzyjne dla $C = 1$

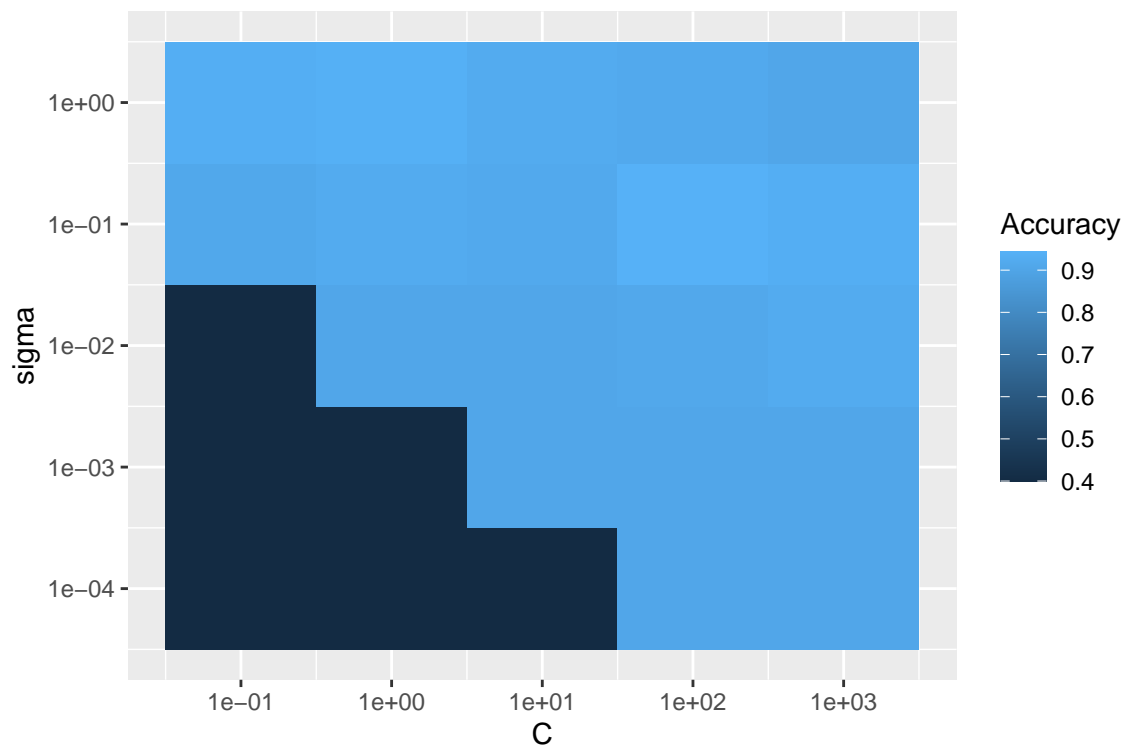Rysunek 4: Obszary decyzyjne dla $C = 10$

Rysunek 5: Obszary decyzyjne dla $C = 100$

Rysunek 6: Obszary decyzyjne dla $C = 1000$

linear	polynomial	radial
0.927	0.932	0.944

Tabela 1: Porównanie klasyfikatorów dla róznych jader

Rysunek 7: Mapa ciepla dokladnosci klasyfikatora

sigma	C
0.10	100.00

Tabela 2: Parametry dla najlepszego klasyfikatora

3 Zadanie 2
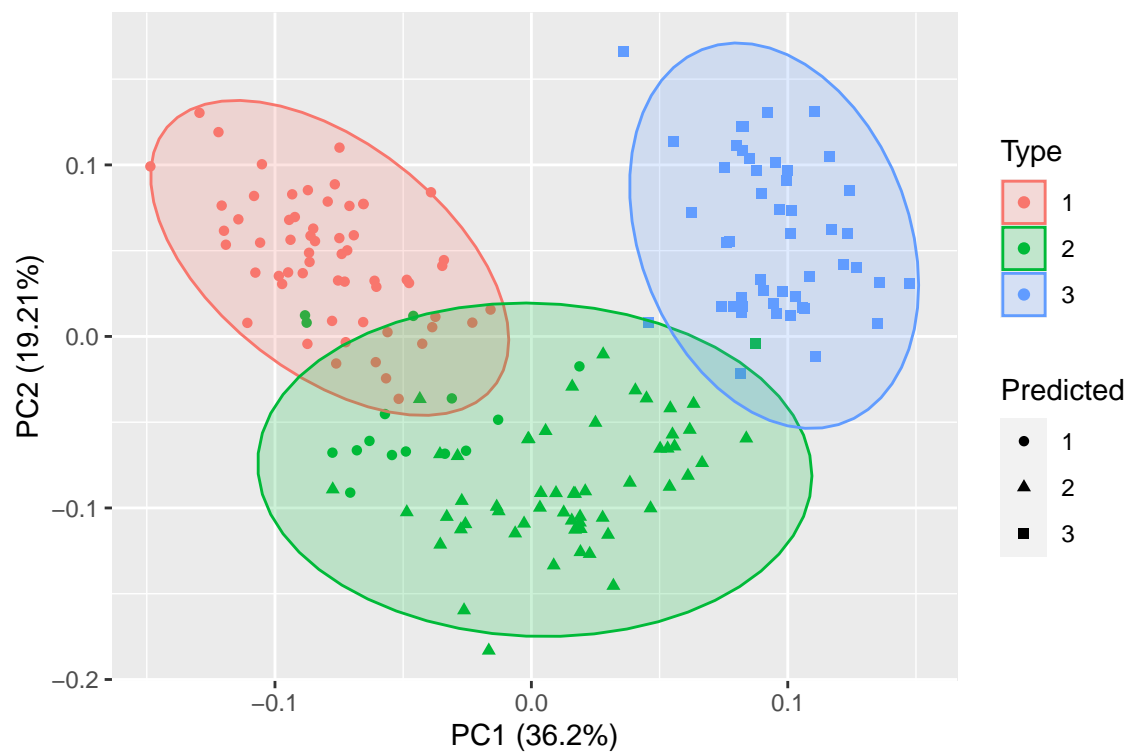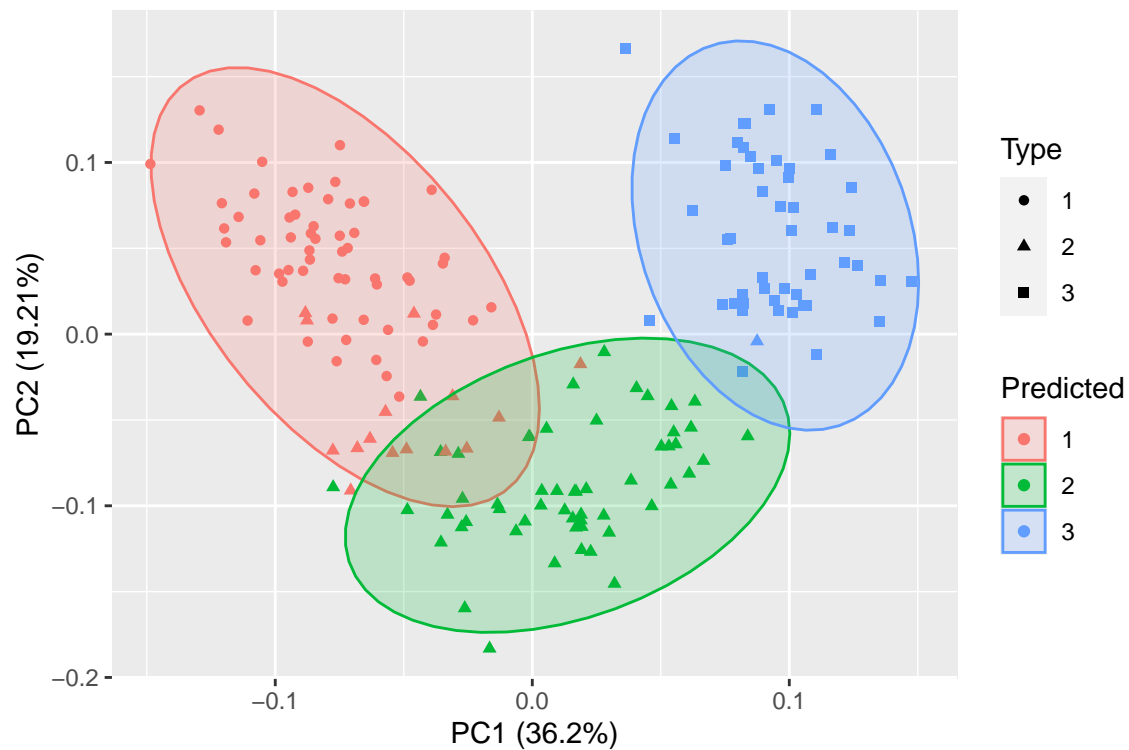
```
## Warning in clValid(wine %>% select(-Type), nClust = cl.range, clMethods =
## cl.methods, : rownames for data not specified, using 1:nrow(data)

##
## 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   2.2329  6.9567 10.5615 15.4302 18.2468 23.0460 24.5746 26.5770 2
##        Dunn          0.0716  0.0830  0.0343  0.0417  0.0417  0.0532  0.0636  0.0725
##        Silhouette    0.6587  0.6101  0.5296  0.5458  0.5409  0.5101  0.5202  0.5051
## 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
##        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         0.0830 agnes    3
## Silhouette   0.6587 agnes    2

##                   Score Method Clusters
## Connectivity 1.52857143    pam        2
## Dunn         0.08304858  agnes        3
## Silhouette   0.65872930  agnes        2
```

Internal validation

Internal validation

Internal validation
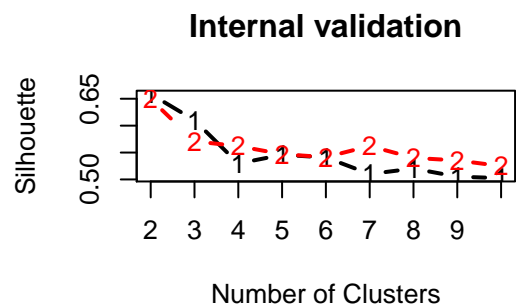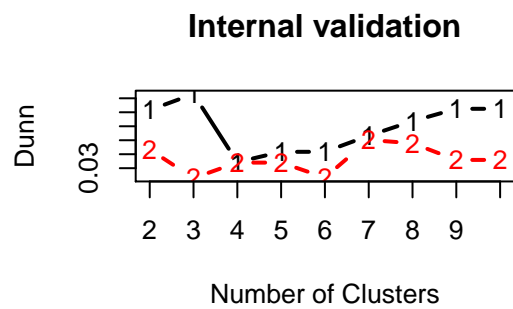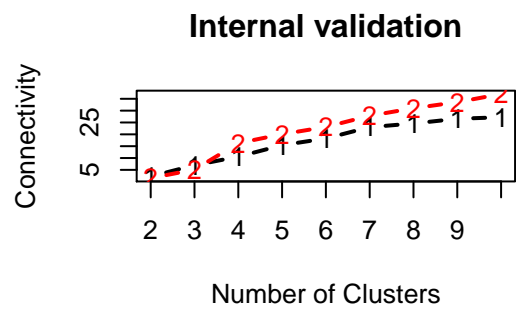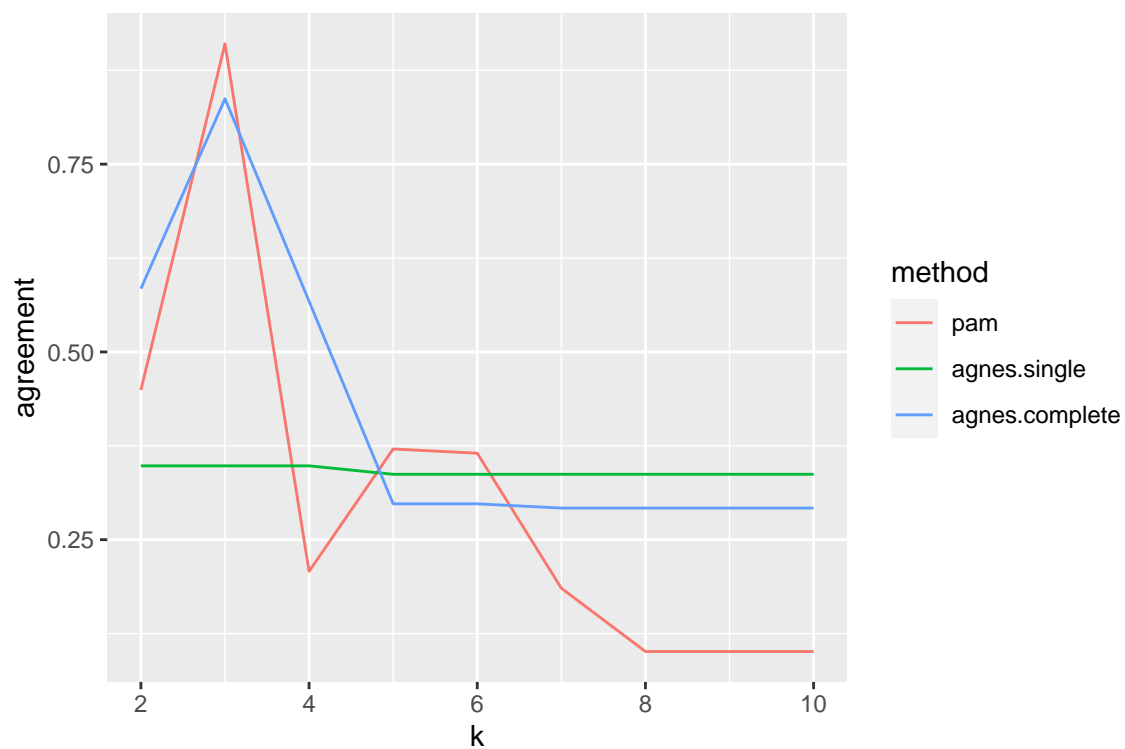

```
wine.pam <- pam(scale(wine %>% select(-Type)), 3)
wine.agnes <- cutree(agnes(scale(wine %>% select(-Type)), method='complete'), 3)
```
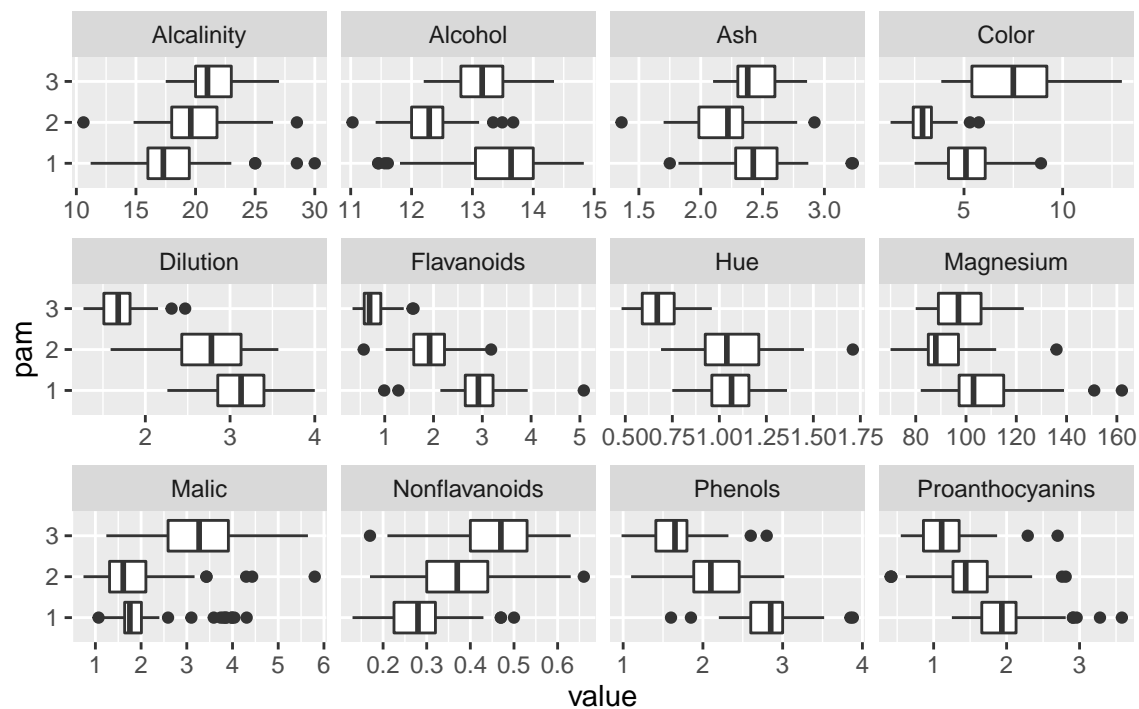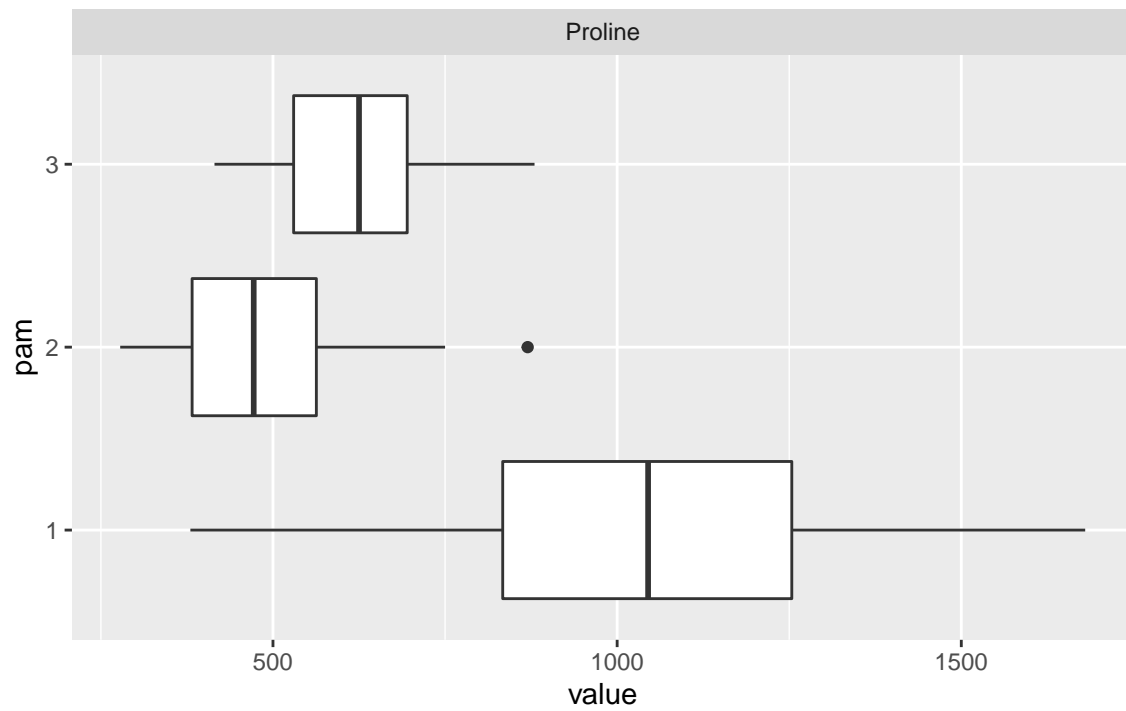
```
wine$pam <- as.factor(wine.pam$clustering)
wine$agnes <- as.factor(wine.agnes)

plot_boxplot(wine, by='pam')
```
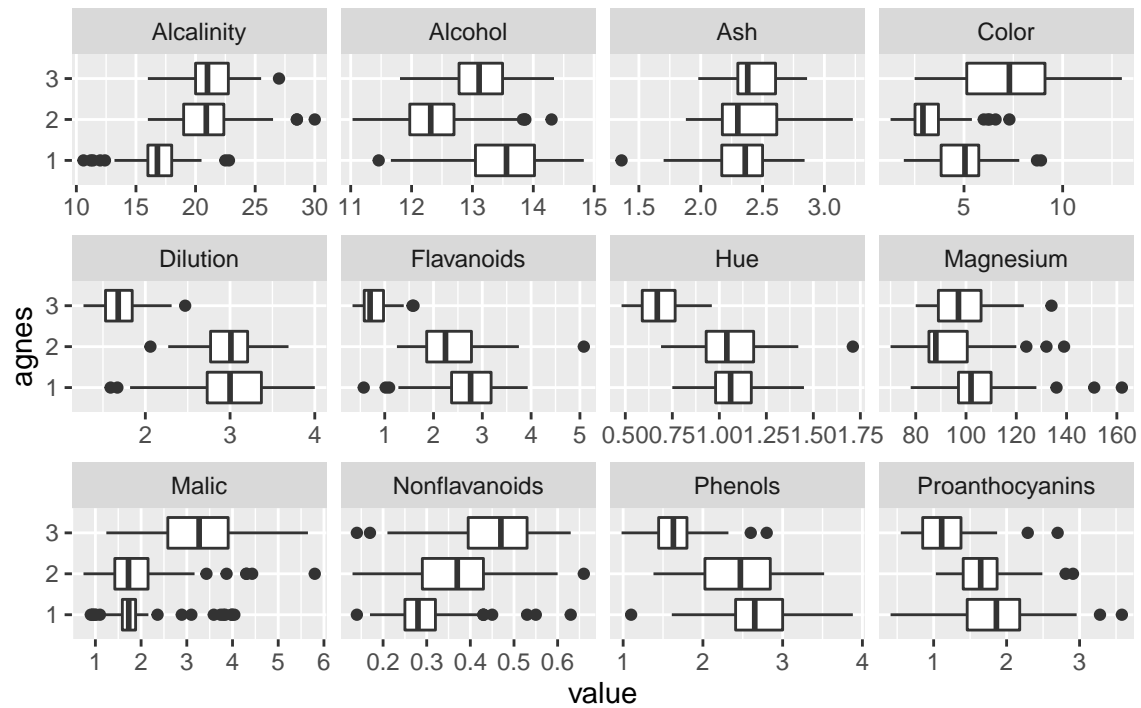
Proline
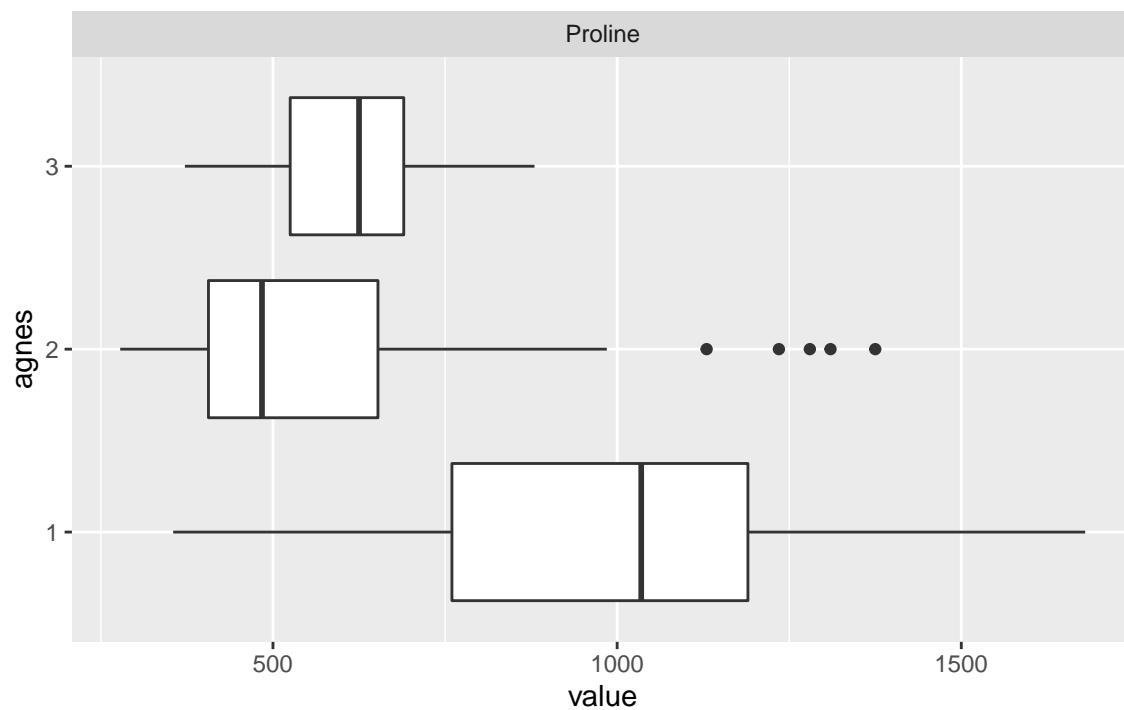
```
plot_boxplot(wine, by='agnes')
```

10

```
wine.pam$medoids
```

```
##          Alcohol      Malic         Ash Alcalinity   Magnesium     Phenols
## [1,]   0.5904981 -0.4711544  0.15849862  0.3009543  0.01809398   0.6469393
## [2,]  -0.9246039 -0.5427655 -0.89856839 -0.1482061 -1.38222271  -1.0307762
## [3,]   0.3934117  0.8088930  0.04914686  0.6003946 -0.54203270  -0.5833854
##         Flavanoids Nonflavanoids Proanthocyanins       Color        Hue
## [1,]   0.9518166597   -0.81841060      0.47016154  0.01807806   0.3611585
## [2,]   0.0007311716    0.06545479      0.06831575 -0.71522236   0.1861586
## [3,]  -1.2707199546    0.70826598     -0.59560339  1.45017064  -1.7825902
##        Dilution    Proline
## [1,]   1.2089101  0.5497067
## [2,]   0.7863692 -0.7522631
## [3,]  -1.3967588 -0.3076880
```

11