Dani DiTomasso
Professor Eleish
Data Analytics
8 November 2024

Lab 4 Written Responses

Identify the variables that contribute the most to the 1st PC.

```
Loadings:
                          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
Alcohol                    0.131  0.533  0.211  0.190         0.408
Malic acid                -0.279  0.201  0.377 -0.555  0.503  0.237
Magnesium                  0.141  0.304 -0.767 -0.300  0.378 -0.175
Total phenols              0.424         0.255         0.109 -0.476
Flavanoids                 0.456         0.201               -0.271
Nonflavanoid Phenols      -0.322         0.150  0.537  0.634 -0.346
Color Intensity           -0.127  0.552         0.168 -0.278 -0.366
Hue                        0.342 -0.275 -0.165  0.370  0.288  0.331
Od280/od315 of diluted wines  0.418 -0.159  0.256 -0.190  0.129
Proline                    0.293  0.409         0.257  0.107  0.289
                          Comp.7 Comp.8 Comp.9 Comp.10
Alcohol                    0.563  0.300  0.205
Malic acid                       -0.338
Magnesium                  0.131  0.133
Total phenols                    -0.188  0.431 -0.527
Flavanoids                       -0.132         0.806
Nonflavanoid Phenols              0.238
Color Intensity                  -0.314 -0.571
Hue                        0.268 -0.552 -0.249 -0.109
Od280/od315 of diluted wines        0.515 -0.608 -0.214
Proline                   -0.761

               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
SS loadings       1.0    1.0    1.0    1.0    1.0    1.0    1.0    1.0
Proportion Var    0.1    0.1    0.1    0.1    0.1    0.1    0.1    0.1
Cumulative Var    0.1    0.2    0.3    0.4    0.5    0.6    0.7    0.8
               Comp.9 Comp.10
SS loadings       1.0     1.0
Proportion Var    0.1     0.1
Cumulative Var    0.9     1.0
```

Based on the summary run for the principal components loadings, Flavanoids (0.456), Total phenols (0.424), and Od280/od315 of diluted wines (0.418) are the three most significant variables to the 1st PC due to the fact that they are the closest values to 1, meaning that they have the most influence.

Drop the variables least contributing to the 1st PC and rerun PCA.

Based on the summary above, the three variables that are least contributing to the 1st component are Nonflavanoid Phenols (-0.322), Malic acid (-0.279), and Color Intensity (-0.127), due to the fact that they are the farthest away from 1, therefore lessening their impact.

Contingency Tables

Table to predict wine type using 13 attributes.

	actual		
predicted	1	2	3
1	54	3	2
2	2	61	9
3	2	7	37

Table to predict wine type using the data projected into the first 3 PCs.

	actual		
predicted	1	2	3
1	58	1	0
2	0	70	0
3	0	0	48

Table to predict wine type using the data projected into the first 3 PCs after rerunning PCA.

	actual		
predicted	1	2	3
1	58	1	0
2	0	68	1
3	0	2	47

Contingency Tables with prevision/recall/f1 metrics

Wine type using 13 attributes with prevision/recall/f1 metrics.

	Precision	Recall	F1
Class: 1	0.9152542	0.9310345	0.9230769
Class: 2	0.8472222	0.8591549	0.8531469
Class: 3	0.8043478	0.7708333	0.7872340

Wine type using the data projected into the first 3 PCs with prevision/recall/f1 metrics

	Precision	Recall	F1
Class: 1	0.9830508	1.0000000	0.9914530
Class: 2	1.0000000	0.9859155	0.9929078
Class: 3	1.0000000	1.0000000	1.0000000

.

wine type using the data projected into the first 3 PCs after rerunning PCA with prevision/recall/f1 metrics.

	Precision	Recall	F1
Class: 1	0.9830508	1.0000000	0.9914530
Class: 2	0.9855072	0.9577465	0.9714286
Class: 3	0.9591837	0.9791667	0.9690722