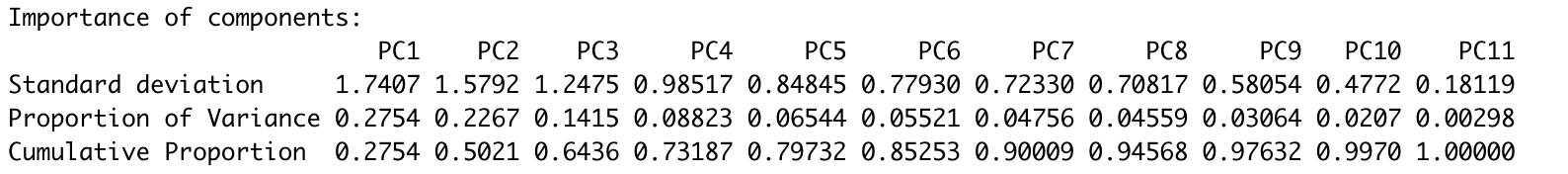
For the wine dataset, there are total 6497 observations and 12 variables. The output is the quality of the wine, scored from 0 to 10. For the principal components analysis, we will do a feature extraction and create 11 new independent variables, from PC1 to PC11.

After calculating the variance of each component, we show the results in a Scree plot. From the Scree plot, the black line indicates the proportion of variance explained by each feature. PC1 can explain 27.54 percent of total variance. PC2 can explain 22.67 percent of total variance. PC3 can explain 14.15 percent of total variance. We can decide how many principal components we want to obtain according to how much variances we want to explain. In this case, if we only want to explain 50 percent of the variance, we can just pick the first two components that can explain 50.21 percent of total variances.



图表, 直方图

描述已自动生成

From the plot for variables with a PCA coordinates, the axes are seen as arrows originating from the center point. The magnitude of the arrows will show by the length and the directions of the arrows will show which principal components they are related to. Here, you see that the variables residual sugar, free sulfur dioxide, total sulfur dioxide all has a positive correlation with PC1, with higher values in those variables moving the samples to the right on this plot. For variables like density, fixed acidity, chloride and citric acid all has a positive correlation with PC2. Alcohol has a negative correlation with PC2. Also, sulphates, pH, volatile acidity has a correlation with PC1. Also, we can calculate the portion of each variables accounts for the two principal components. The results are showing below.

表格

描述已自动生成图表

描述已自动生成

Below is a graph showing all the samples grouped by type of wine (red wine and white wine). From the graph, we can see samples from white wine mostly lie in the positive side of x-axis, which means white wine are characterized by variables like citric acid, residual sugar, total sulfur dioxide, free sulfur dioxide and alcohol content. For red wine, samples are mostly lie in the negative part of x-axis, which means red wine are characterized by variables like pH, volatile acidity, sulphates, chlorides and fixed acidity. Neither of the wines are characterized by density.

图表, 散点图

描述已自动生成