About the transformation - dt\_pca: Principal Component Analysis (PCA) projects correlated variables onto orthogonal components ordered by explained variance. You can let the tool pick the number of components via an elbow heuristic or set it explicitly.

Method - Learns principal components via an orthogonal transformation (SVD/eigendecomposition of the covariance matrix), applied to centered/scaled numeric predictors. - By default, the number of components is chosen via minimum-curvature (elbow) on the cumulative explained variance curve; alternatively set components manually.

Environment setup.

# installation   
#install.packages("daltoolbox")  
  
# loading DAL  
library(daltoolbox)

Load data and PCA idea.

# Dataset for example  
iris <- datasets::iris  
head(iris)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1 5.1 3.5 1.4 0.2 setosa  
## 2 4.9 3.0 1.4 0.2 setosa  
## 3 4.7 3.2 1.3 0.2 setosa  
## 4 4.6 3.1 1.5 0.2 setosa  
## 5 5.0 3.6 1.4 0.2 setosa  
## 6 5.4 3.9 1.7 0.4 setosa

PCA PCA finds a projection capturing the largest possible variance in the data. Below, we fit PCA and transform the dataset.

# creates and fits PCA using the target column as reference  
mypca <- dt\_pca("Species")  
mypca <- fit(mypca, datasets::iris)  
iris.pca <- transform(mypca, iris)

PCA properties

print(head(iris.pca))

## PC1 PC2 Species  
## 1 2.640270 -5.204041 setosa  
## 2 2.670730 -4.666910 setosa  
## 3 2.454606 -4.773636 setosa  
## 4 2.545517 -4.648463 setosa  
## 5 2.561228 -5.258629 setosa  
## 6 2.975946 -5.707321 setosa

print(head(mypca$pca.transf))

## PC1 PC2  
## Sepal.Length 0.5210659 -0.37741762  
## Sepal.Width -0.2693474 -0.92329566  
## Petal.Length 0.5804131 -0.02449161  
## Petal.Width 0.5648565 -0.06694199

Manually set the number of components and repeat the transformation.

# Manual definition of the number of components  
mypca <- dt\_pca("Species", 3)  
mypca <- fit(mypca, datasets::iris)  
iris.pca <- transform(mypca, iris)  
print(head(iris.pca))

## PC1 PC2 PC3 Species  
## 1 2.640270 -5.204041 2.488621 setosa  
## 2 2.670730 -4.666910 2.466898 setosa  
## 3 2.454606 -4.773636 2.288321 setosa  
## 4 2.545517 -4.648463 2.212378 setosa  
## 5 2.561228 -5.258629 2.392226 setosa  
## 6 2.975946 -5.707321 2.437245 setosa

print(head(mypca$pca.transf))

## PC1 PC2 PC3  
## Sepal.Length 0.5210659 -0.37741762 0.7195664  
## Sepal.Width -0.2693474 -0.92329566 -0.2443818  
## Petal.Length 0.5804131 -0.02449161 -0.1421264  
## Petal.Width 0.5648565 -0.06694199 -0.6342727

References - Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. - Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components.