About the transformation - zscore: standardizes numeric attributes to zero mean and unit variance. You can also rescale to a target mean (nmean) and standard deviation (nsd).

Method - For each numeric column j: (x - mean\_j) / sd\_j and optionally \* nsd + nmean. - Constant columns (zero variance) map to nmean.

When to use - Useful to equalize scale across features and as input to scale-sensitive models (e.g., k-NN, SVM with RBF kernel, PCA).

Environment setup.

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

Context and sample data (iris).

# Normalization  
  
# Normalization is a technique used to equal strength among variables.   
  
# It is also important to apply it as an input for some machine learning methods.   
  
# Example  
  
iris <- datasets::iris   
summary(iris)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species   
## Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100 setosa :50   
## 1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300 versicolor:50   
## Median :5.800 Median :3.000 Median :4.350 Median :1.300 virginica :50   
## Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199   
## 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800   
## Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500

Apply standard Z-Score (mean = 0, sd = 1) and inverse-transform.

# Z-Score  
  
# Adjust values to 0 (mean), 1 (variance).  
  
norm <- zscore()  
norm <- fit(norm, iris)  
ndata <- transform(norm, iris)  
summary(ndata)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species   
## Min. :-1.86378 Min. :-2.4258 Min. :-1.5623 Min. :-1.4422 setosa :50   
## 1st Qu.:-0.89767 1st Qu.:-0.5904 1st Qu.:-1.2225 1st Qu.:-1.1799 versicolor:50   
## Median :-0.05233 Median :-0.1315 Median : 0.3354 Median : 0.1321 virginica :50   
## Mean : 0.00000 Mean : 0.0000 Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.: 0.67225 3rd Qu.: 0.5567 3rd Qu.: 0.7602 3rd Qu.: 0.7880   
## Max. : 2.48370 Max. : 3.0805 Max. : 1.7799 Max. : 1.7064

ddata <- inverse\_transform(norm, ndata)  
summary(ddata)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species   
## Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100 setosa :50   
## 1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300 versicolor:50   
## Median :5.800 Median :3.000 Median :4.350 Median :1.300 virginica :50   
## Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199   
## 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800   
## Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500

Standardization to custom target mean and standard deviation.

norm <- zscore(nmean=0.5, nsd=0.5/2.698)  
norm <- fit(norm, iris)  
ndata <- transform(norm, iris)  
summary(ndata)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species   
## Min. :0.1546 Min. :0.05044 Min. :0.2105 Min. :0.2327 setosa :50   
## 1st Qu.:0.3336 1st Qu.:0.39059 1st Qu.:0.2735 1st Qu.:0.2813 versicolor:50   
## Median :0.4903 Median :0.47562 Median :0.5621 Median :0.5245 virginica :50   
## Mean :0.5000 Mean :0.50000 Mean :0.5000 Mean :0.5000   
## 3rd Qu.:0.6246 3rd Qu.:0.60318 3rd Qu.:0.6409 3rd Qu.:0.6460   
## Max. :0.9603 Max. :1.07088 Max. :0.8298 Max. :0.8162

Inverse transform for checking.

ddata <- inverse\_transform(norm, ndata)  
summary(ddata)

## Sepal.Length Sepal.Width Petal.Length Petal.Width Species   
## Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100 setosa :50   
## 1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300 versicolor:50   
## Median :5.800 Median :3.000 Median :4.350 Median :1.300 virginica :50   
## Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199   
## 3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800   
## Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500

References - Han, J., Kamber, M., Pei, J. (2011). Data Mining: Concepts and Techniques. (Standardization)