Scaling

Kenneth C. Enevoldsen

10/14/2019

This is the start of a tutorial explaining why you should scale your variables. Some of it is still missing though.

Scaling formula

Scaling (or Z-score normalization) follows the formula

$$x' = \frac{x - \bar{x}}{\sigma}$$

Where $x - \bar{x}$ centers the variables around zero and where dividing by σ constitutes the normalization of the variables.

The scale function performs a Z-score normalization, which can be seen here:

```
#tinytex::install_tinytex()
set.seed(1) # set seed to ensure random number generation is the same
x <- runif(7) # generate random numbers

# Manually scaling
(x - mean(x)) / sd(x)
```

```
## [1] -1.01951259 -0.68940037 -0.06788275  0.97047346 -1.21713898  0.94007371
## [7]  1.08338753
```

```
#scale using scale
scale(x)
```

```
##              [,1]
## [1,] -1.01951259
## [2,] -0.68940037
## [3,] -0.06788275
## [4,]  0.97047346
## [5,] -1.21713898
## [6,]  0.94007371
## [7,]  1.08338753
## attr(,"scaled:center")
## [1] 0.5947772
## attr(,"scaled:scale")
## [1] 0.3229666
```

```
class(scale(x))
```

```
## [1] "matrix"
```

note that scale is a class matrix, which might cause problems. If you want scale to return a numeric vector (a list of numbers) use `scale(x)[[1]]`.