

R for Multidimensional Scaling

Dolores Romero Morales

The aim of this workshop is to work on building Multidimensional Scaling using a cars dataset that is familiar to you (mtcars), which can be found in R, and an email dataset that is also familiar to you (spam.txt), which can be found on CANVAS.

Note: Luckily, 'cmdscale()' and related commands to do multidimensional scaling are included in the 'stats' package, which is available in the basic download of R.

Step 1. Load the 'mtcars' dataset and get the dimension of the dataset.

Answer:

```
data(mtcars)
dim(mtcars)
```

Step 2. Normalize the data, get summary statistics and plot the data.

Answer:

```
minimum <- apply(mtcars,2,min)
maximum <- apply(mtcars,2,max)
mtcarsNORM <- scale(mtcars,center=minimum,scale=(maximum-minimum))
mymtcarsNORM <- as.data.frame(mtcarsNORM)
summary(mymtcarsNORM)
with(mymtcarsNORM, pairs(mymtcarsNORM))
```

Step 3. Using the normalized dataset, derive a Multidimensional Scaling with $l=2$ and the Euclidean distance.

Answer:

```
myMDS <- cmdscale(dist(mymtcarsNORM), 2, eig=TRUE)
```

Note

As for other data science tools, you may want to run your MDS several times for different seeds to make sure that you do not get trapped in local optima, as opposed to the global optimum. Recall we did the same a similar thing when building k-means clusterings.

Step 4. Plot the representative points obtained in Step 3.

Answer:

```
x <- myMDS$points[,1]
y <- myMDS$points[,2]
plot(x, y, xlab="Representative's Coordinate 1", ylab="Representative's Coordinate 2",
main="MDS")
text(x, y, labels=row.names(mymtcarsNORM), cex = 0.7)
```

Note

The parameter 'cex' defines the size of the label. If you want to set range of the axes to be, e.g., [-2,2], you would need to add the option 'xlim=c(-2,2), ylim=c(-2,2)' to the plot function.

Step 5. Experiment with other distances, repeating Step 3 and 4. Compare the outputs.

Answer:

```
myMDSManhattan <- cmdscale(dist(mymtcarsNORM,method="manhattan"), 2, eig=TRUE)
x <- myMDSManhattan$points[,1]
y <- myMDSManhattan$points[,2]
plot(x, y, xlab="Representative's Coordinate 1", ylab="Representative's Coordinate 2",
main="MDSManhattan")
```

```
text(x, y, labels=row.names(mymtcarsNORM), cex = 0.7)
myMDSMaximum <- cmdscale(dist(mymtcarsNORM,method="maximum"), 2, eig=TRUE)
x <- myMDSMaximum$points[,1]
y <- myMDSMaximum$points[,2]
plot(x, y, xlab="Representative's Coordinate 1", ylab="Representative's Coordinate 2",
main="MDSMaximum")
text(x, y, labels=row.names(mymtcarsNORM), cex = 0.7)
```