Euclidean Cloud

Cory Brunson

November 6, 2015

This vignette (loosely or closely?) follows the treatment in Chapter 3 of Le Roux and Rouanet's *Geometric Data Analysis*, with drastically simplified exposition but employing the same example data.

```
## Warning: package 'devtools' was built under R version 3.1.3
```

```
## Loading cloud
```

The central object of GDA is a point cloud in a Euclidean space, which encodes a set of statistical observations arranged in a table, or two-dimensional array (for instance, a contingency table in Correspondence Analysis or an Individuals–Variables table in Principal Components Analysis). By convention, the columns of the table correspond to the coordinate axes of the Euclidean space while the rows correspond to the points.

3.1 Basic Statistics

Rigorously, a *point cloud* in a Euclidean space $\mathcal{U}$ consists of a set J of labels and a mapping $M^J : J \to \mathcal{U}$ that takes each label $j \in J$ to its corresponding point $M^j \in \mathcal{U}$.

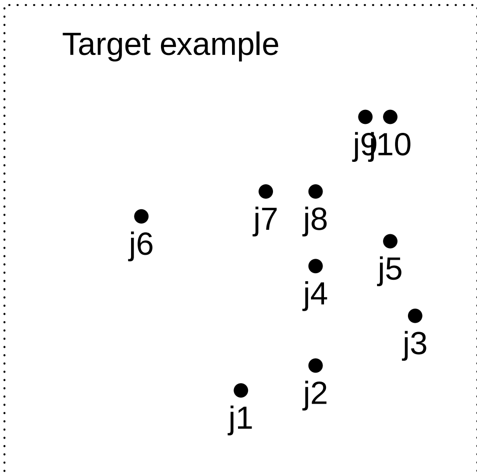
As a running example, we load the *Target example* dataset, used throughout the chapter, whose coordinates are provided in Exercise 3.4:

```r
data(Target)
print(Target)
```

```
##     x1  x2
## j1   0 -12
## j2   6 -10
## j3  14  -6
## j4   6  -2
## j5  12   0
## j6  -8   2
## j7   2   4
## j8   6   4
## j9  10  10
## j10 12  10
```

```r
class(Target)
```

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



The labels consist of j_1 through j_{10}. Because these data occupy two dimensions, i.e. the labels are mapped into $\mathcal{U} = \mathbb{R}^2$, they constitute a *plane cloud*.

3.1.1 Mean Point

The *mean point* of the cloud M^J in $\mathcal{U}$ is meant to be a proxy point location for the entire cloud. Given an arbitrary point $P \in \mathcal{U}$, the mean point is defined as the point arrived to from P via the sum of the vectors $\overrightarrow{PM^j}$ from P to each point M^j—that is, as the (unique!) point $G \in \mathcal{U}$ for which

$$\overrightarrow{PG} = \sum_{j \in J} \overrightarrow{PM^j}$$

doesn't depend on the choice of point $P \in \mathcal{U}$. The mean point satisfies the *Barycentric property* that

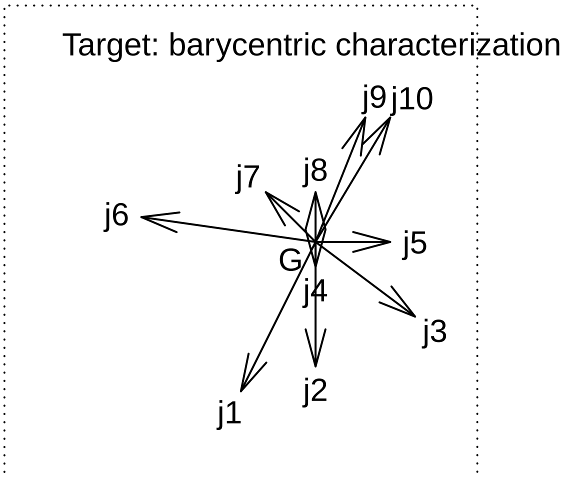
$$\sum_{j \in J} \overrightarrow{GM^j} = \overrightarrow{0},$$

which can be checked by substituting $P = G$ in the definition.[1]

```
Target_barycenter <- barycenter(Target)
print(Target_barycenter)
```

```
##      [,1] [,2]
## [1,]    6    0
```

[1] These and other formulas are given in more general terms in *GDA*, in which the points M^j are weighted by masses ω_j. I may be forced to adopt this generality, but i haven't yet.


Target: barycentric characterization
j9
j10
j8
j7
j6
j5
G
j4
j3
j2
j1