

Multidimensionale Skalierung in R

Achim Zeileis

2009-02-20

Um die Analyse aus der Vorlesung zu reproduzieren, wird zunächst der `Languages` Datensatz geladen

```
R> load("Languages.rda")
```

der die Distanzen zwischen verschiedenen Sprachen auf Basis der Zahlwörter für die Zahlen 1 bis 10 enthält:

```
R> Languages
```

	English	Danish	Dutch	German	French	Polish	Hungarian	Finnish	Russian
English	0	2	7	6	6	7	9	9	7
Danish	2	0	6	5	6	6	8	9	7
Dutch	7	6	0	5	9	10	8	9	10
German	6	5	5	0	7	8	9	9	8
French	6	6	9	7	0	5	10	9	5
Polish	7	6	10	8	5	0	10	9	2
Hungarian	9	8	8	9	10	10	0	8	10
Finnish	9	9	9	9	9	9	8	0	9
Russian	7	7	10	8	5	2	10	9	0
Arabic	6	7	10	8	7	7	10	10	8
	Arabic								
English		6							
Danish		7							
Dutch		10							
German		8							
French		7							
Polish		7							
Hungarian		10							
Finnish		10							
Russian		8							
Arabic		0							

Die entsprechende klassische multidimensionale Skalierung wird mit dem Befehl `cmdscale` durchgeführt. Dabei kann man einstellen, in wie vielen Dimensionen eine Konfiguration berechnet werden soll. Die Voreinstellung ist eine 2-dimensionale Konfiguration:

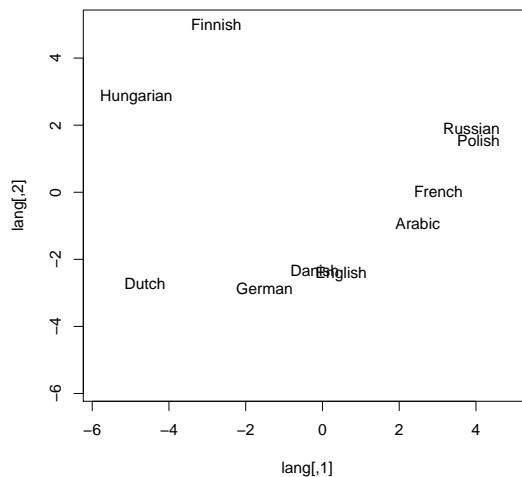
```
R> lang <- cmdscale(Languages)
R> lang
```

```
[,1]      [,2]
English  0.4919195 -2.43310384
```

Danish	-0.2012374	-2.33022168
Dutch	-4.6260932	-2.73000296
German	-1.5026666	-2.86519887
French	3.0208410	0.01959413
Polish	4.0589176	1.54632515
Hungarian	-4.8433885	2.83226035
Finnish	-2.7665327	5.00460390
Russian	3.8856574	1.88735807
Arabic	2.4825831	-0.93161426

Diese visualisieren wir wieder indem wir zunächst den Plot aufsetzen, aber keine Punkte plotten (`type = "n"`) und dann den Text mit den Namen der Beobachtungen an die entsprechenden Koordinaten hinzufügen.

```
R> plot(lang, type = "n", xlim = c(-5.8, 5), ylim = c(-5.8, 5))
R> text(lang, rownames(lang))
```



Trotz des extrem simplen Abstandsmasses, das der MDS zugrunde liegt, können hier einige interessante Beobachtungen gemacht werden. Zum einen deckt die Grafik auf, daß die indogermanischen Sprachen wie Englisch, Deutsch und Dänisch sehr ähnlich sind (d.h. die entsprechenden skalierten Werte nahe beieinander liegen). Zum anderen wird die Ähnlichkeit der slawischen Sprachen Russisch und Polnisch visualisiert. Zuletzt isoliert die Grafik auch noch die beiden Sprachen Ungarisch und Finnisch, die in der Tat miteinander verwandt sind.