Cluster Analysis: Identifying Parkinson's Disease Subtypes

Jesse Mu

Wednesday, June 10

1 Preprocessing

1.1 Dataset Description

951 subjects, 145 metrics, collected 15-4-2012. From Pablo Martinez Martín. 170 subjects with missing values (brought down to 781); these were removed automatically, even if the missing values were not included in the selected features below. This will need to be changed later on, by keeping those removed that still have all selceted features and perhaps with some compensation for missing values.

1.2 Selected Features

Combination of non-motor scale (NMS) symptoms and standard motor symptoms.

Name	Type	Format	Description
nms_d1	byte	%8.0g	cardiovascular
nms_d2	byte	%8.0g	sleep/fatigue
nms_d3	byte	%8.0g	mood/cognition
nms_d4	byte	%8.0g	percep/hallucinations
nms_d5	byte	%8.0g	attention/memory
nms_d6	byte	%8.0g	gastrointestinal
nms_d7	byte	%8.0g	urinary
nms_d8	byte	%8.0g	sexual function
nms_d9	byte	%8.0g	miscellaneous
tremor	float	%9.0g	tremor
bradykin	float	%9.0g	bradykinesia[1]
rigidity	float	%9.0g	rigidity
axial	float	%9.0g	axial[2]
pigd	float	%9.0g	postural instability and gait difficulty

Table 1: Selected Features and Details

Name	μ	σ	min-max
nms_d1	1.76	3.32	0-24
nms_d2	8.71	8.76	0-48
nms_d3	8.70	11.83	0-60
nms_d4	1.65	3.94	0-33
nms_d5	5.22	7.44	0-36
nms_d6	5.67	6.92	0-36
nms_d7	8.02	9.09	0-36
nms_d8	3.57	5.97	0-24
nms_d9	6.99	7.74	0-48
tremor	2.59	2.63	0-12
bradykin	2.49	1.39	0-6
rigidity	2.34	1.36	0-6
axial	3.28	2.75	0-12
pigd	3.36	2.77	0-12

Table 2: Descriptive Statistics

1.3   Dimensionality Reduction: PCA

May not be useful? If we're trying to identify *clinically* relevant features, merging them may not be a good idea.

Figure 1 shows scree test elbow occurs around 2 or 3. Also, eigenvalues 1 and 2 > 1, while 3 is around .9

2 k-means

2.1   Identifying optimal number of clusters

Figure 2 shows no optimal elbow in scree test! Maybe 2-3?

2.2   Decision Tree classifier based on clusters

3   Biclustering

4   Subspace clustering

5   Bayesian Networks

---

[1]Impaired ability to adjust the body's position.
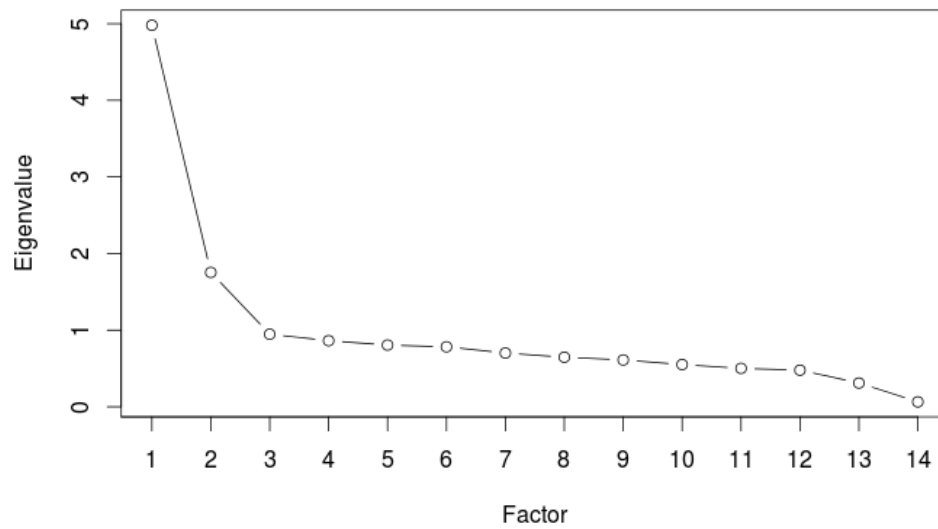
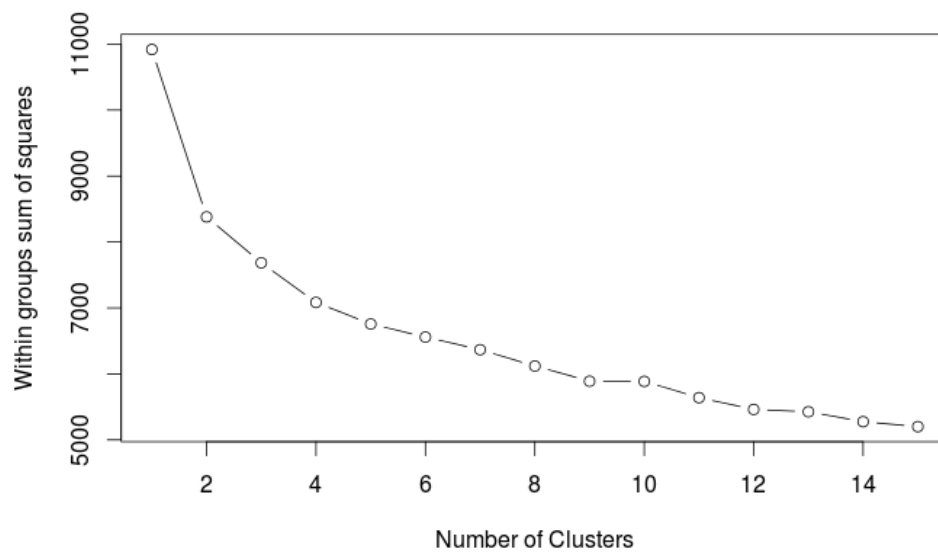[2]Issues affecting the middle of the body.

Figure 1: Scree test: eigenvalues by factor

Figure 2: Scree test: WSS error by cluster size