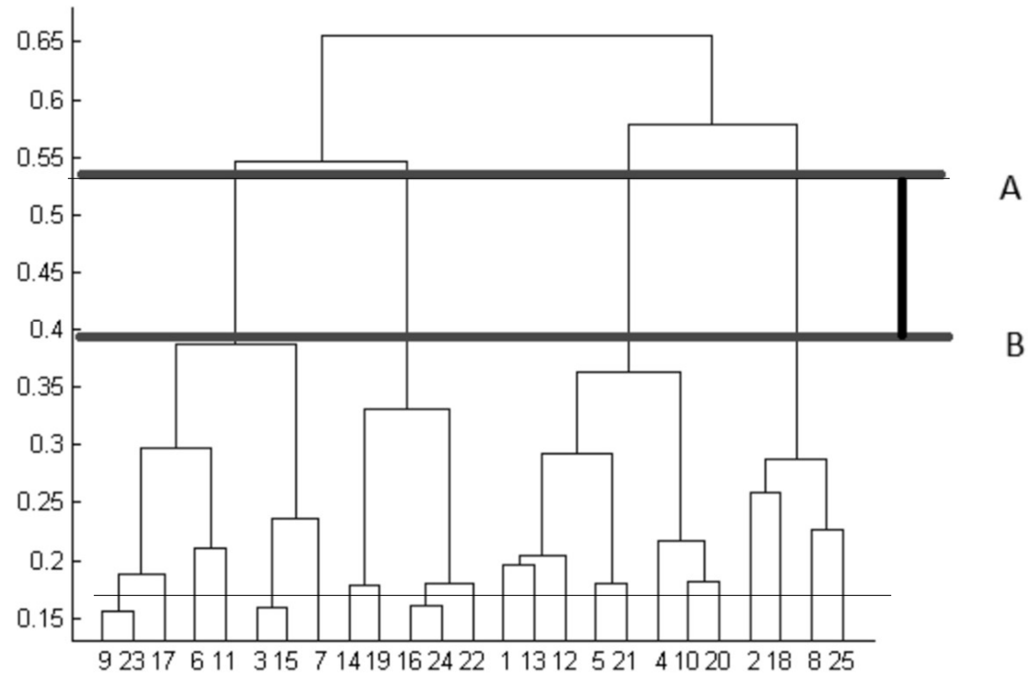
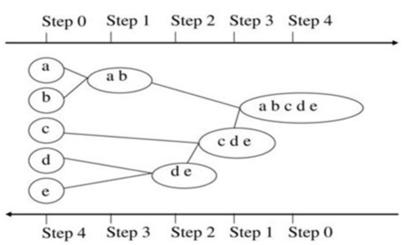
Agglomerative



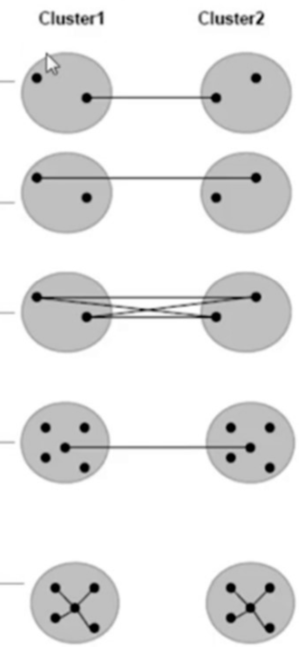
- Use distance matrix as clustering criteria. This method does not require the number of clusters ***k*** as an input, but needs a termination condition



Linkage

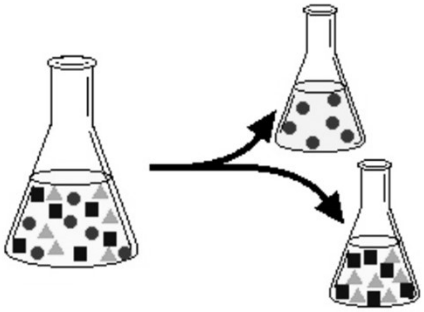
Agglomerative Clustering Linkage Algorithms

- Single linkage – Minimum distance or Nearest neighbour rule
- Complete linkage – Maximum distance or Farthest distance
- Average linkage – Average of the distances between all pairs
- Centroid method – combine cluster with minimum distance between the centroids of the two clusters
- Ward's method – Combine clusters with which the increase in within cluster variance is to the smallest degree

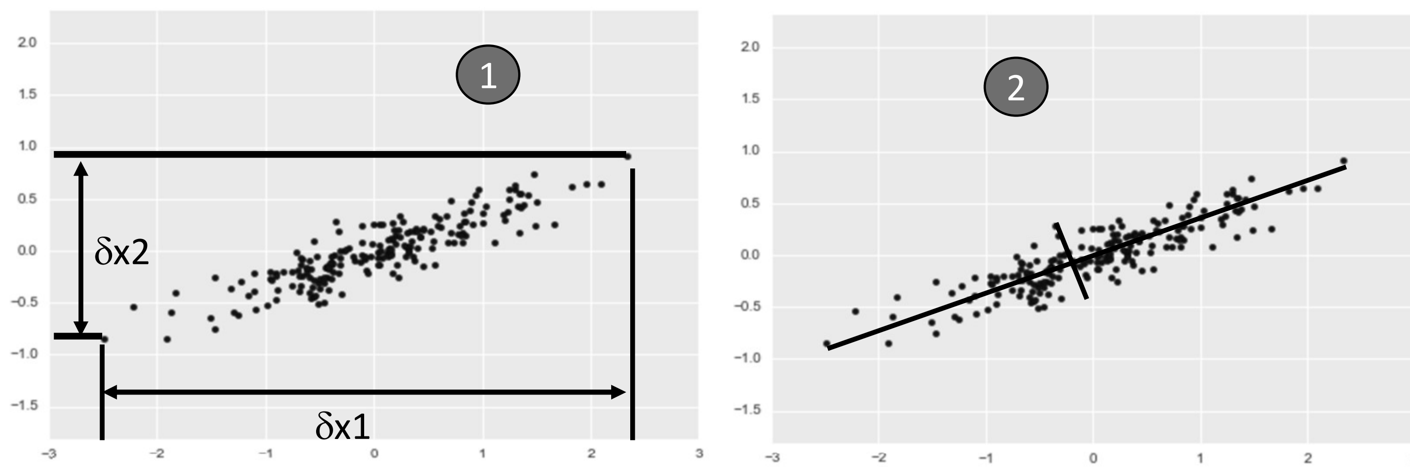


DR/PCA

Dimensionality Reduction



Elimination	Extraction
1 Missing Value Ratio 2 Low Variance Filter 3 High Correlation Filter 4 Feature ranking: (Random Forest) 5 VIF	7 PCA/Principal Component Analysis 8 FA/Factor Analysis 9 Independent Component Analysis

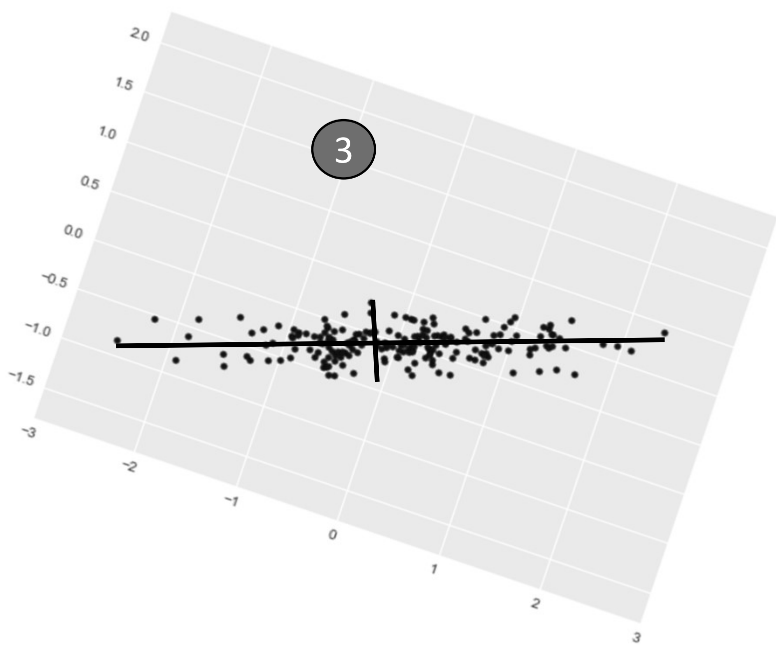


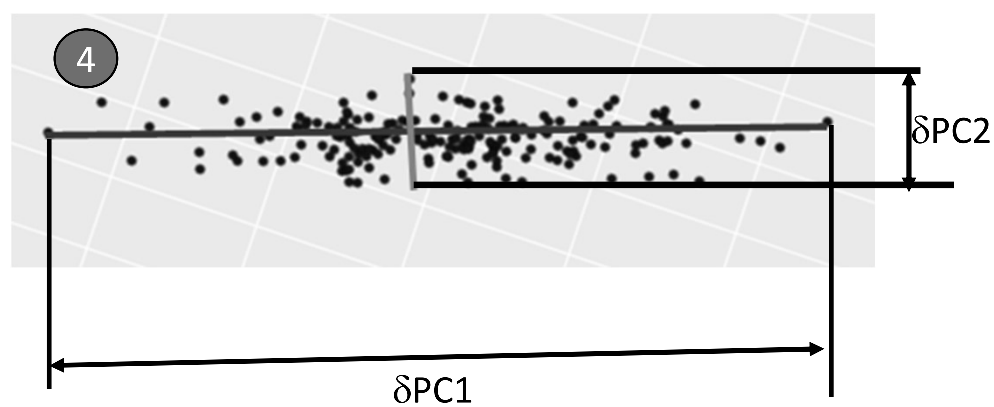
y= beta0+beta1*x1

PC1=w11x1+w12X2

y= beta0'+beta1'*PC1

Correlated variables to uncorrelated Components

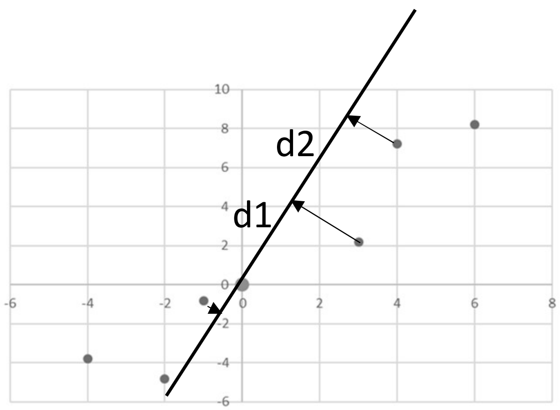




Note: no longer the data points are in terms of x_1,x_2. The plot is in terms of PC_1 and PC_2

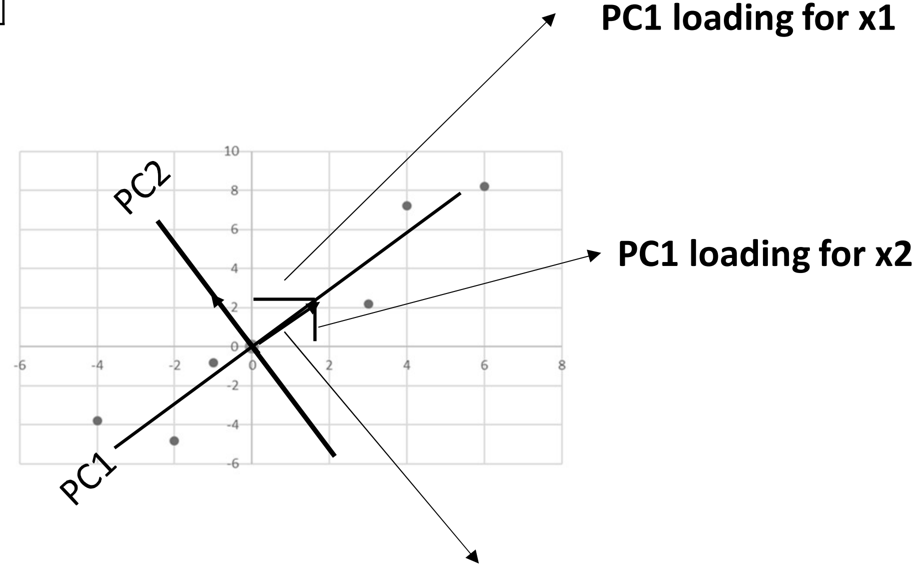
How PCA works

x1	1	4	3	8	9	11
x2	3	6	2	9	14	16



SSD = $d_1^2 + d_2^2 + \ldots . d_6^2$

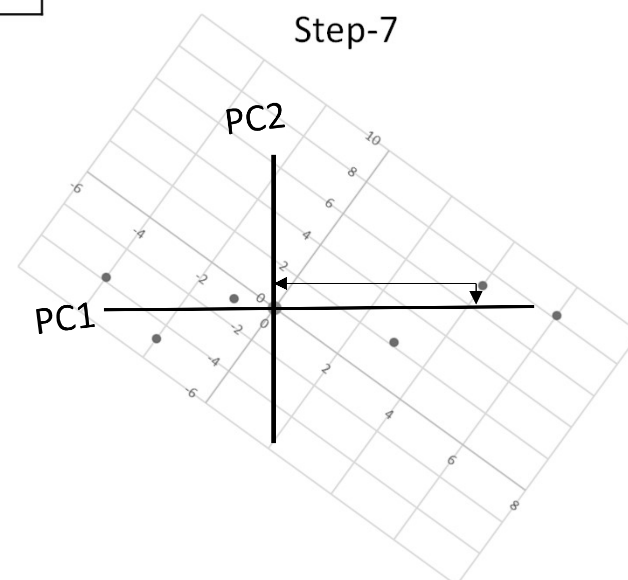
Maximize (SSD)



Eigen Value → SSD

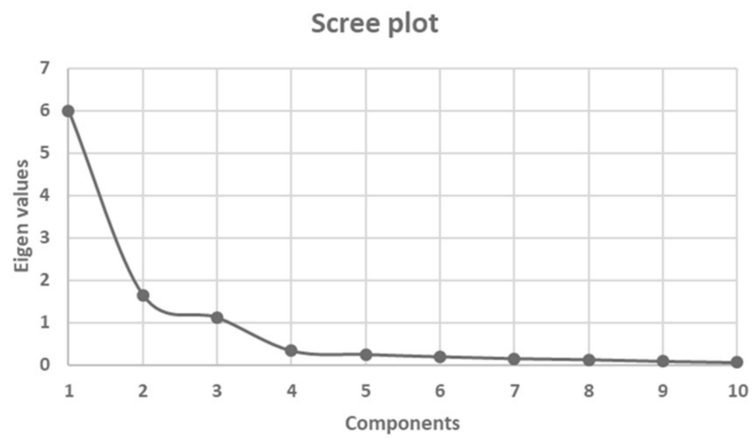
How PCA works

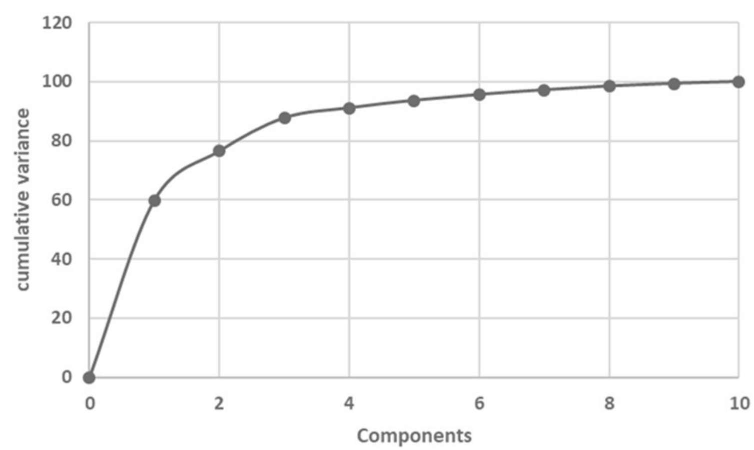
x1	1	4	3	8	9	11
x2	3	6	2	9	14	16



Determining the number of PCs/Fs

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	5.994	59.938	59.938
2	1.654	16.545	76.482
3	1.123	11.227	87.709
4	.339	3.389	91.098
5	.254	2.541	93.640
6	.199	1.994	95.633
7	.155	1.547	97.181
8	.130	1.299	98.480
9	.091	.905	99.385
10	.061	.615	100.000





- PCA for DR
- PCA for Noise reduction