

Assignment 6

Task 1 - Multivariate Data

- (a) Yes, there is a correlation between x and y, since the correlation coefficient is very high, there is a positive linear correlation. Surprisingly, the mean, variance and persons correlation coefficient are the same for all four groups.

	I		II		III		IV	
	x	y	x	y	x	y	x	y
mean	9	7.50	9	7.50	9	7.5	9	7.50
variance	11.0	4.127	11.0	4.127	11.0	4.127	11.0	4.127
pearson	0.8164		0.8164		0.8164		0.8164	

- (b) The plots are shown in Figure 1

It is not visible in the plots, that each of these has the same correlation and variance, as well as mean values for x and y.

- (c)

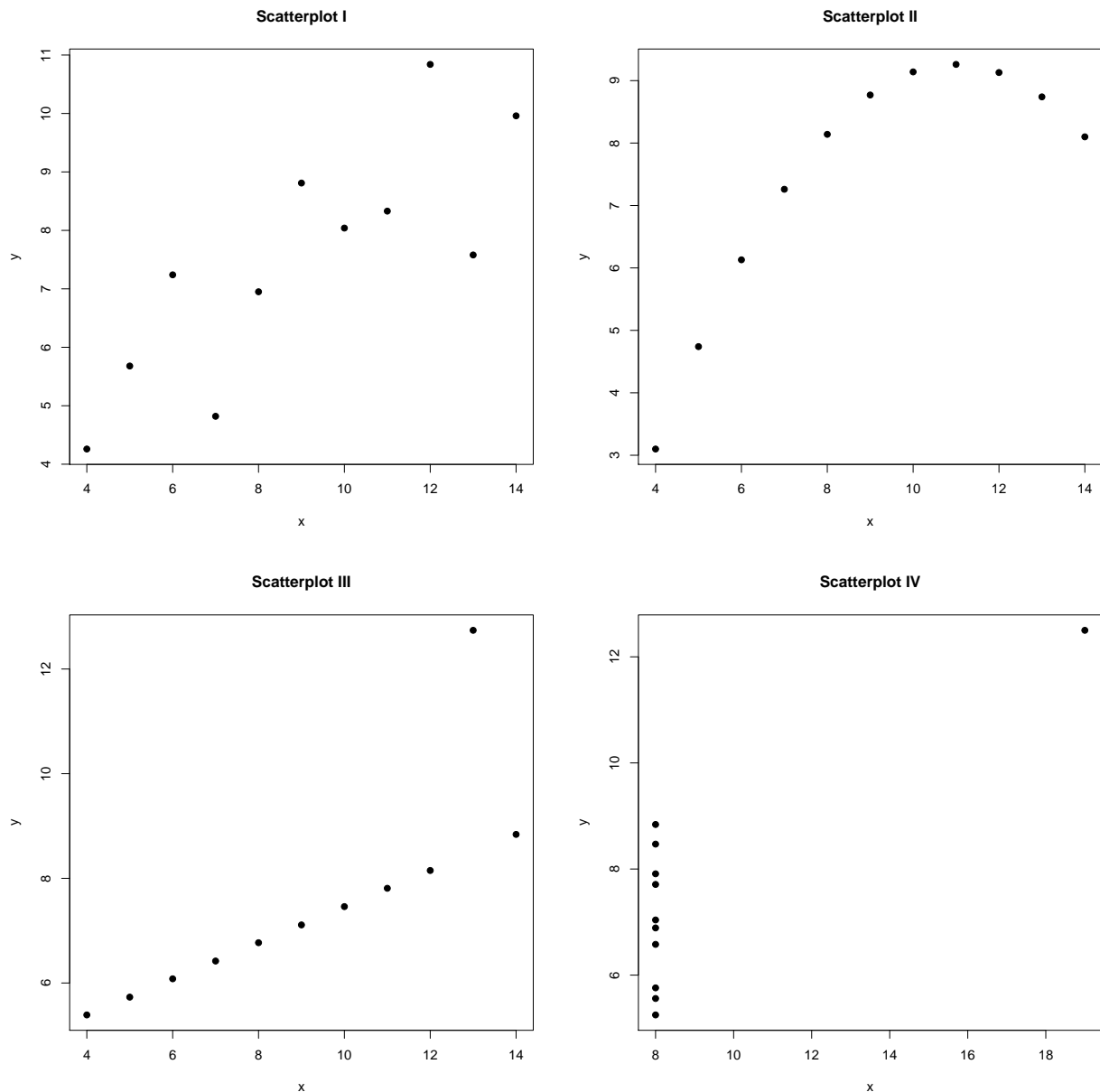


Figure 1: Task 1 (b) - Scatter plots

Task 2 - Dimensionality Reduction (DR)

(a)

dimensionality reduction is specially useful for reducing computation time and space (storage and memory). Sometime redundant dimensions are causing an overhead, or dimensions can be combined, without losing information. Also better for visualization, because the user can understand the data.

(b)

(c)

•

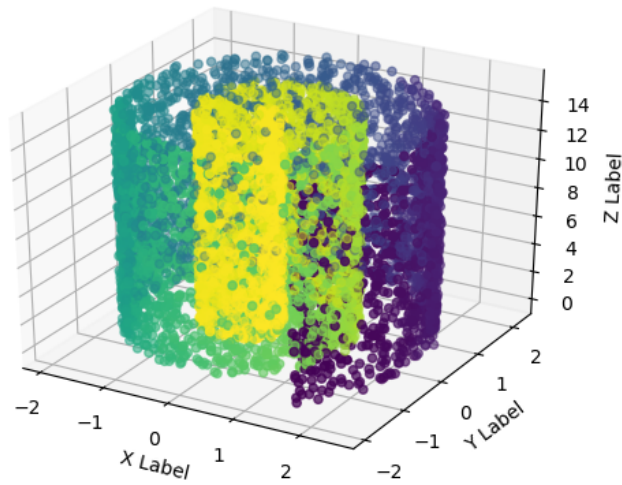


Figure 2: raw data

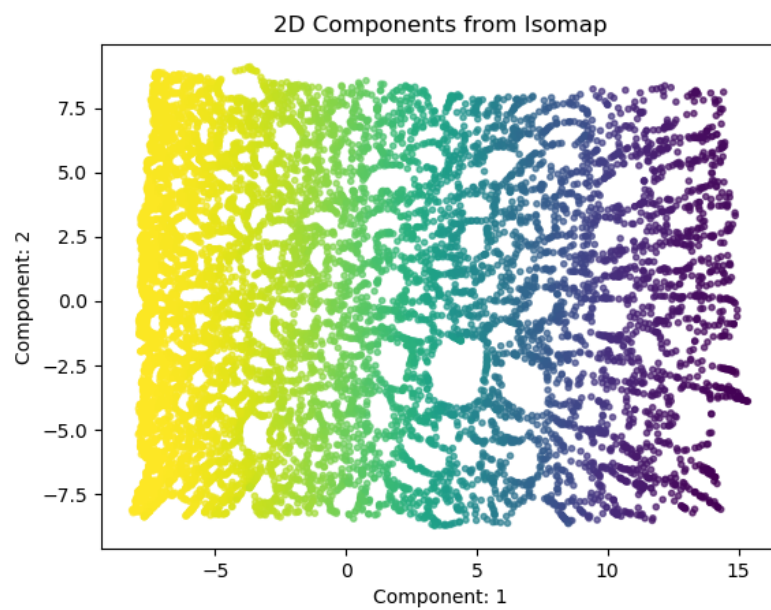


Figure 3: after pca

- TODO
- TODO

(d)