Exercise: Principal Component Analysis (PCA)

ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

Exercise : PCA

The following table represents PCA output on wine data (non-normalized) in which the variables represent chemical characteristics of wine, and each case is a different wine.

	Principal Components					Std
	1	2	3	4	5	Dev
Alcohol	0.001	0.013	0.014	-0.030	0.129	0.8
MalicAcid	-0.001	0.009	0.167	-0.427	-0.402	1.2
Ash	0.000	-0.002	0.054	-0.009	0.006	0.3
Ash_Alcalinity	-0.004	-0.045	0.976	0.176	0.060	3.6
Magnesium	0.014	-0.998	-0.040	-0.031	0.006	14.7
Total Phenols	0.001	0.002	-0.015	0.164	0.316	0.7
Flavanoids	0.002	0.000	-0.049	0.214	0.545	1.1
Nonflavanoid_Phenols	0.000	0.002	0.004	-0.025	-0.040	0.1
Proanthocyanins	0.001	-0.007	-0.031	0.082	0.244	0.7
Color Intensity	0.002	0.022	0.097	-0.804	0.536	1.6
Hue	0.000	-0.002	-0.021	0.096	0.064	0.2
OD280/OD315	0.001	-0.002	-0.022	0.220	0.261	0.7
Proline	1.000	0.014	0.004	0.001	-0.004	351.5
Variance	123594.453	194.345	11.424	2.388	1.391	
% Variance	99.830%	0.157%	0.009%	0.002%	0.001%	
Cumulative %	99.830%	99.987%	99.996%	99.998%	99.999%	

a) Consider the row near the bottom labeled “variance”. Explain why column 1’s variance is so much greater that of any other column?

b) Comment on the use of normalization?

c) Use RapidMiner. Compare and comment the results between using normalization+PCA and without using normalization+PCA. How the result different of a)?

d) Use Python or R. Do you obtain same results?