UNIVERSITÄT BERN INFORMATIK	KURS	TYP	BLATT	AUSGABE	
	ME	UA	1	HS 10	

Einführung in die Mustererkennung

Programmieraufgabe 3

1 Aufgabe

Decomposing of a set of samples via clustering

Let $\mathbf{X}$ be a set of feature vectors. Each feature vector $\mathbf{x}_i = (x_1, \ldots, x_3)$ contains 3 features. The number of classes as well as the class labels of the samples is not known.

a) Implement the non-hierarchical K-means algorithm described in Chapter 9.2 The centers of the clusters should be initialized by a random selection of K elements from the set $\mathbf{X}$. Check as a criterion for the termination of the loop the reduction of the error.

b) Apply the algorithm from part a) to set $\mathbf{X}$. Let $K = 5$ and make 10 runs of the algorithm, each with a different random initialization of the centers of the clusters.

   Plot the elements in the feature space. Indicate the clusters by using different symbols for the elements according to the cluster to which they are assigned.

c) What is the behavior of the decomposition when using a different K (e.g. $K = 4$ or $K = 6$)?

Give the errors E_K from parts b) and c) for the different runs of the algorithm in form of a table and plot the print the plot of the best decomposition each.

Print the source code and the results. Additionally, send the program via e-Mail to Elias Gerber (*elias.gerber@students.unibe.ch*)