Imaduddin Sheikh

CS 422 - Homework 8

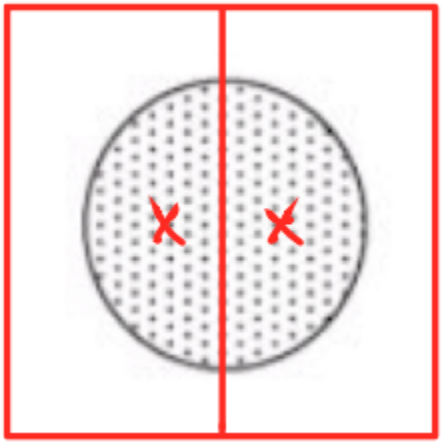
1. Exercises

1.1

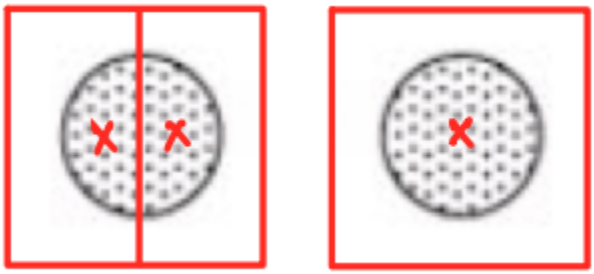
Q.2)

Q.6)

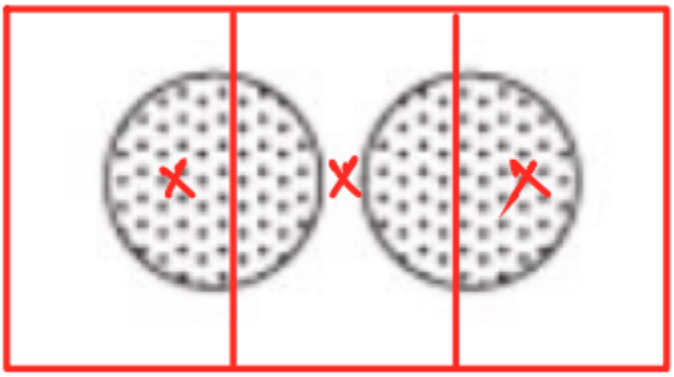
a)



b)

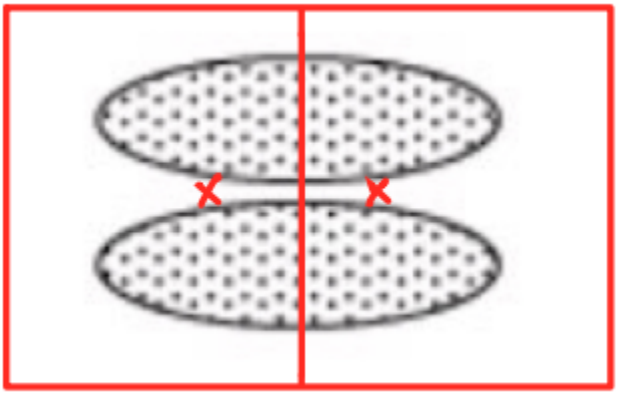


c)

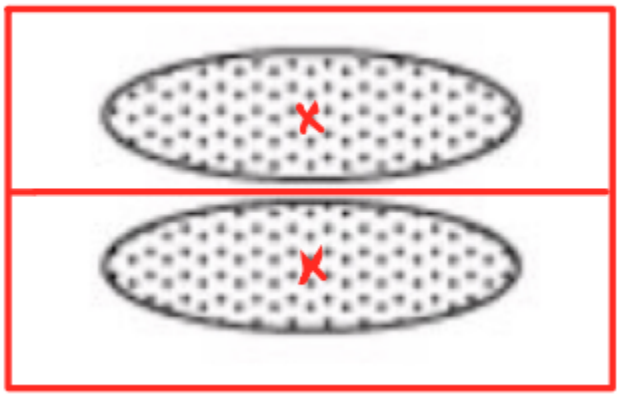


d)

Global Minimum

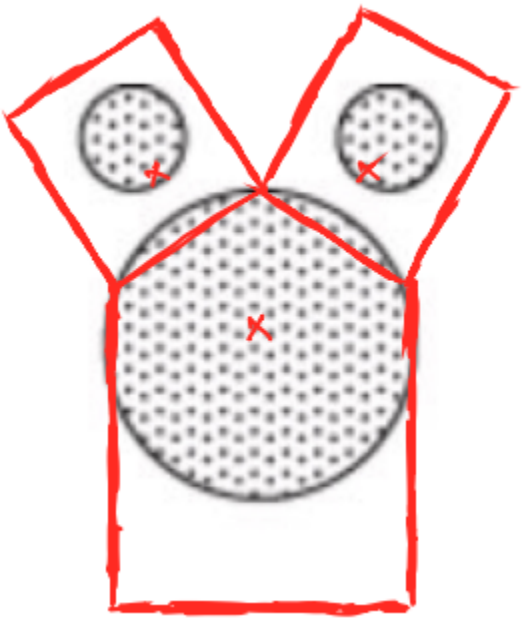


Local Minimum

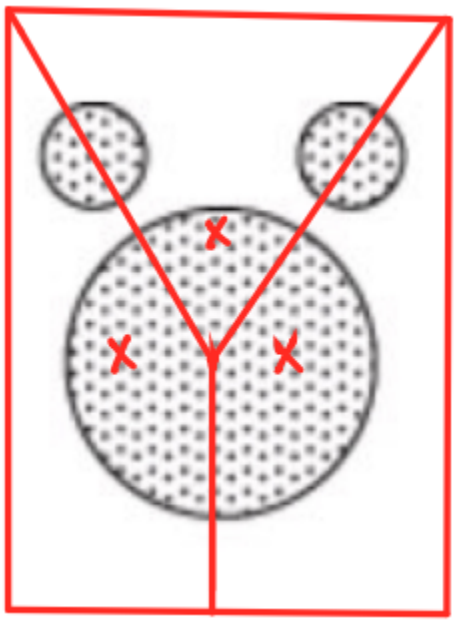


e)

Global Minimum



Local Minimum



Q.11)

a) If the SSE of one attribute is low for all clusters, then the variable is Pretty much a constant and dividing the data into groups is not useful.

b) The attribute with the relatively low SSE in a cluster helps define the cluster.

c) An attribute with a relatively high SSE for all clusters indicates that the attribute is a noise.

d) If an attribute's SSE is relatively high for one cluster, then it contradicts the information provided by the attributes with low SSE that define the cluster. It is possible that the clusters defined by this attribute are different from those

defined by the other attributes, but in any case, it means that this attribute does not help define the cluster.

e) Attributes having poor distinguishing power between clusters are to be eliminated, i.e., low or high SSE for all clusters, since they are not useful for clustering. Attributes with high SSE for all clusters
are particularly troublesome if they have a relatively high SSE with respect to other attributes (perhaps because of their scale) since they introduce a lot of noise into the computation of the overall SSE.