More on Algorithm Design for Divisive Clustering

❑ Choosing which cluster to split

   ❑ Check the sums of squared errors of the clusters and choose the one with the largest value

❑ Splitting criterion: Determining how to split

   ❑ One may use Ward's criterion to chase for greater reduction in the difference in the SSE criterion as a result of a split

   ❑ For categorical data, Gini-index can be used

❑ Handling the noise

   ❑ Use a threshold to determine the termination criterion (do not generate clusters that are too small because they contain mainly noises)