

Handling Outliers: From *K-Means* to *K-Medoids*

- ❑ The *K-Means* algorithm is sensitive to outliers!—since an object with an extremely large value may substantially distort the distribution of the data
- ❑ *K-Medoids*: Instead of taking the **mean** value of the object in a cluster as a reference point, **medoids** can be used, which is the **most centrally located** object in a cluster
- ❑ The *K-Medoids* clustering algorithm:
 - ❑ Select K points as the initial representative objects (i.e., as initial K medoids)
 - ❑ **Repeat**
 - ❑ Assigning each point to the cluster with the closest medoid
 - ❑ Randomly select a non-representative object o_i
 - ❑ Compute the total cost S of swapping the medoid m with o_i
 - ❑ If $S < 0$, then swap m with o_i to form the new set of medoids
 - ❑ **Until** convergence criterion is satisfied