

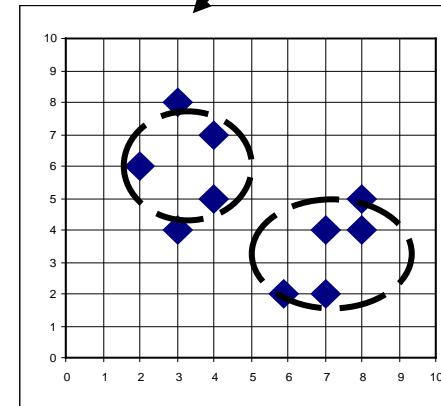
Clustering Feature Vector in BIRCH

❑ Clustering Feature (CF): $CF = (N, LS, SS)$

❑ N : Number of data points

❑ LS : linear sum of N points: $\sum_{i=1}^N X_i$

❑ SS : square sum of N points: $\sum_{i=1}^N X_i^2$



$CF = (5, (16,30),(54,190))$

(3,4)

(2,6)

(4,5)

(4,7)

(3,8)

❑ Clustering feature:

❑ Summary of the statistics for a given sub-cluster: the 0-th, 1st, and 2nd moments of the sub-cluster from the statistical point of view

❑ Registers crucial measurements for computing cluster and utilizes storage efficiently