APPENDIX A

Algorithm 1 is shown below:

Algorithm 1: The GB generation algorithm

Input: Dataset $X = \{x_1, x_2, \ldots, x_n\}$, bandwidth h, merging threshold ε
Output: GB Set $GBs = \{gb_1, gb_2, \ldots, gb_m\}$

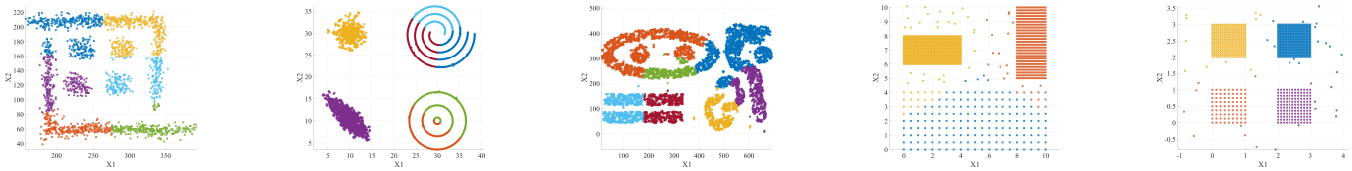
Initialize: $n = |X|$, $gb = X$, $GBs = \emptyset$, $M = \emptyset$, $P = \emptyset$;
Create an empty queue Q, add gb to Q;
while Q is not empty **do**
  Take the first element gb from Q and remove it from Q;
  if $|gb| > \sqrt{n}$ **then**
    for each $o_j \in gb$ **do**
      Calculate: kernel density estimate $\hat{f}(x_j)$ using Eq. (6);
      Calculate: covariance using Eq. (16);
      Calculate: effective dimension D_{eff} using Eq. (15);
      Calculate: adaptive bandwidth h_i using Eq. (17);
      Calculate: mean shift vector $m_{h_i}(o_j)$ using Eq. (7);
      Update: sample position $o_j^{(t+1)} = o_j^{(t)} + m_{h_i}(o_j^{(t)})$ using Eq. (12);
      if $\left\| o_j^{(t+1)} - o_j^{(t)} \right\| < 10^{-3} \cdot h_i$ **then**
        Add $o_j^{(t+1)}$ to set M;
      end
    end
    for each $o_j^* \in M$ **do**
      for each $o_s^* \in M, s \neq j$ **do**
        if $\left\| o_j^* - o_s^* \right\| < \varepsilon$ **then**
          Calculate $p_j = \frac{o_j^* + o_s^*}{2}$ and add p_j to set P;
        end
      end
    end
    $P1 = unique(P)$
    $k = |P1|$;
    if $k > 0$ **then**
      Use elements in $P1$ as initial cluster centers for k-means;
      Divide gb into k sub-GBs $\{sgb_1, sgb_2, \ldots, sgb_k\}$;
      Add $\{sgb_1, sgb_2, \ldots, sgb_k\}$ to the end of Q;
    else
      Add gb to the end of GBs directly;
    end
  else
    Calculate: the center c and radius r of GB using Eq. (2) and Eq. (10)
    $GBs = GBs \cup gb$;
  end
end
Return: The GB set GBs.

The complete clustering procedure is formalized in Algorithm 2.

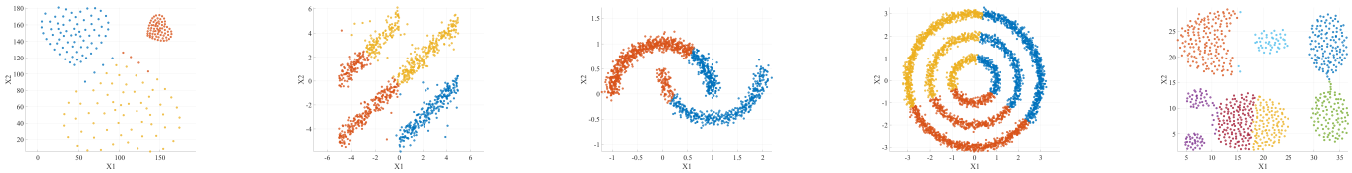
Algorithm 2: GB clustering algorithm

Input: GB set $GBs = \{gb_1, gb_2, \ldots, gb_m\}$, radii r, centers c, initial local density peak set $Modes$
Output: Cluster labels CL

Initialize: $CL = \emptyset$, $GB_{nom} = \emptyset$, $GB_{out} = \emptyset$, $comp = \emptyset$, $cluster = \emptyset$
for $gb_i \in GBs$ **do**
  if $|gb_i| > 1$ **then**
    $GB_{nom} = GB_{nom} \cup \{gb_i\}$
  else
    $GB_{out} = GB_{out} \cup \{gb_i\}$
  end
end
for $gb_i \in GB_{nom}$ **do**
  for $gb_j \in GB_{nom}$ **do**
    Calculate distance $d(gb_i, gb_j)$ by Eq. (21)
    Calculate normalized center distance s_{ij} using Eq. (22)
    Calculate similarity cd_{ij} by Eq. (23)
  end
end
for $gb_i \in GB_{nom}$ **do**
  for $mode \in Modes$ **do**
    Determine if gb_i is a core GB using Eq. (25)
  end
end
Let GBs_{core} be the set of core GBs
$comp = \text{BFS}(cd_{ij}, GBs_{core})$
$cluster = \text{assignClusters}(comp)$
for $gb_i \in GB_{out}$ **do**
  $cluster = cluster \cup \{gb_i\}$
end
for $i = 1$ **to** $|cluster|$ **do**
  if $|cluster(i)| = 1$ **then**
    Let gb_s be the only GB in $cluster(i)$
    Calculate distances from gb_s to every other GB in GBs using Eq. (21)
    Find the nearest GB $gb_{nearest}$
    Assign the cluster label of $gb_{nearest}$ to gb_s
  end
end
for each data point x **do**
  if x is covered by some GB gb_k **then**
    Assign label of gb_k to x
  else
    Calculate distances from x to all GBs by Eq. (26)
    Assign label of the nearest GB to x
  end
end
Return: Cluster labels CL for all data points.

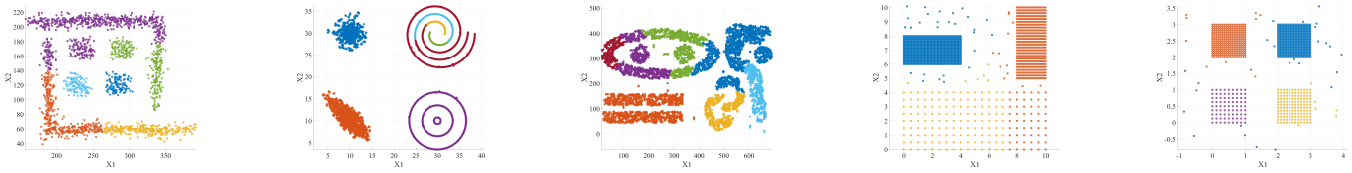


(a) (b) (c) (d) (e)

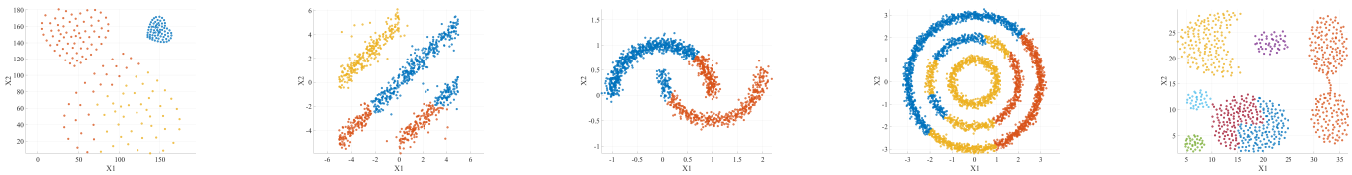


(f) (g) (h) (i) (j)

Fig. A.1: Clustering visualization of k-means on synthetic datasets

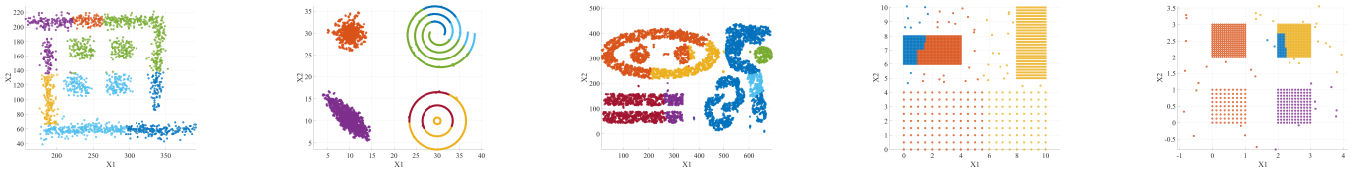


(a) (b) (c) (d) (e)

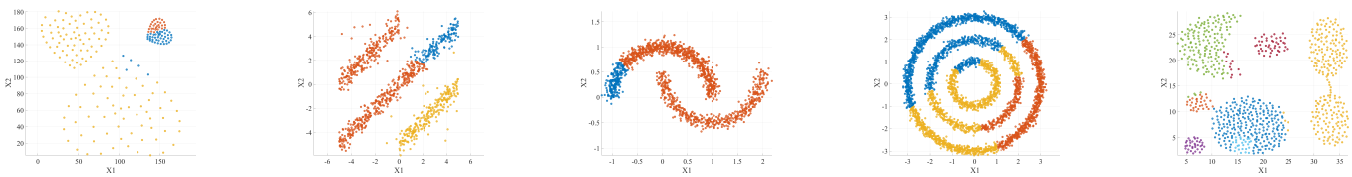


(f) (g) (h) (i) (j)

Fig. A.2: Clustering visualization of DP on synthetic datasets

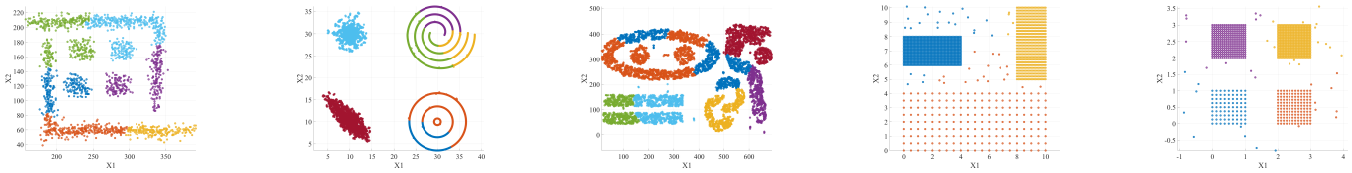


(a) (b) (c) (d) (e)

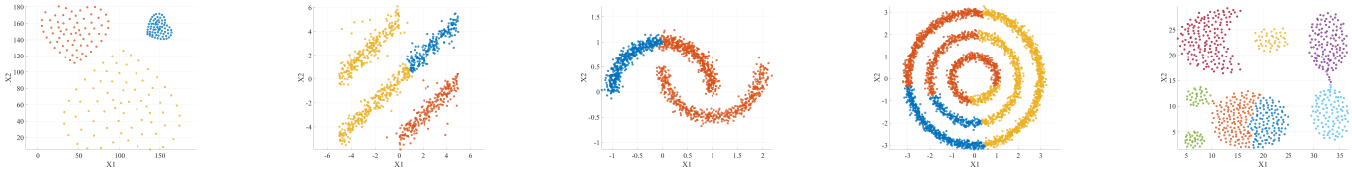


(f) (g) (h) (i) (j)

Fig. A.3: Clustering visualization of GB-DP on synthetic datasets

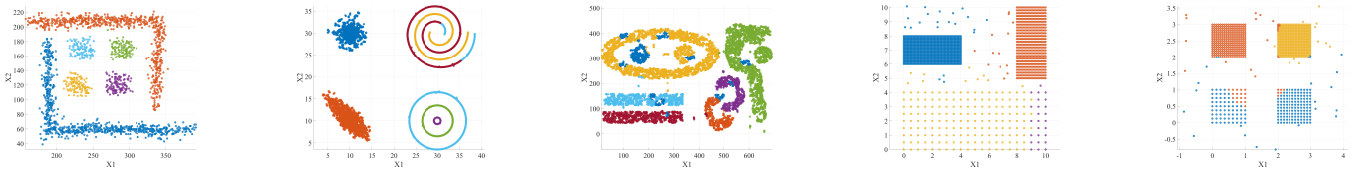


(a) (b) (c) (d) (e)

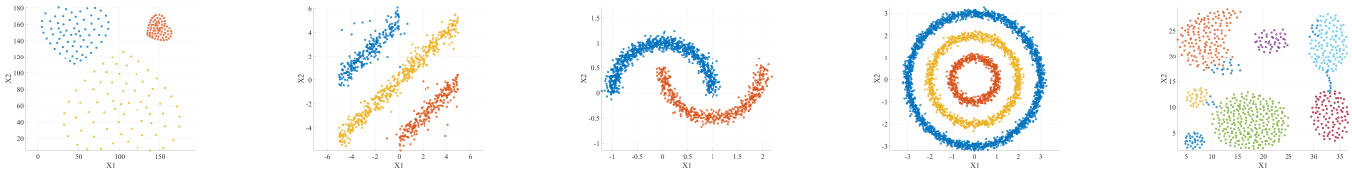


(f) (g) (h) (i) (j)

Fig. A.4: Clustering visualization of HC on synthetic datasets

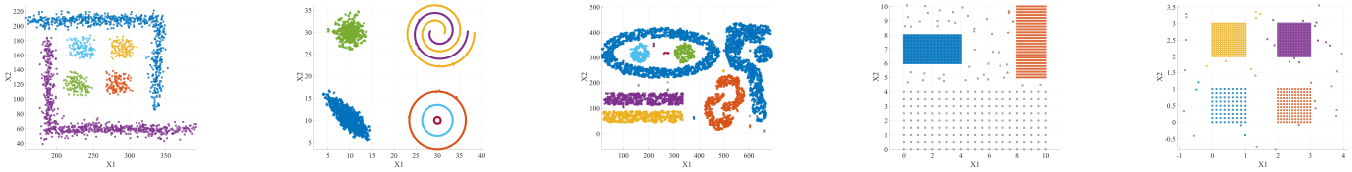


(a) (b) (c) (d) (e)

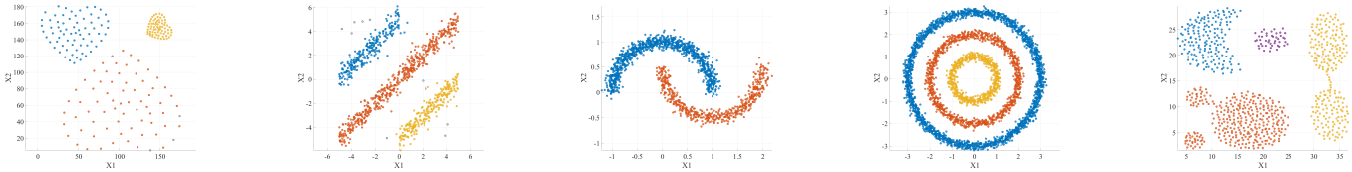


(f) (g) (h) (i) (j)

Fig. A.5: Clustering visualization of NaGB-DBSCAN on synthetic datasets

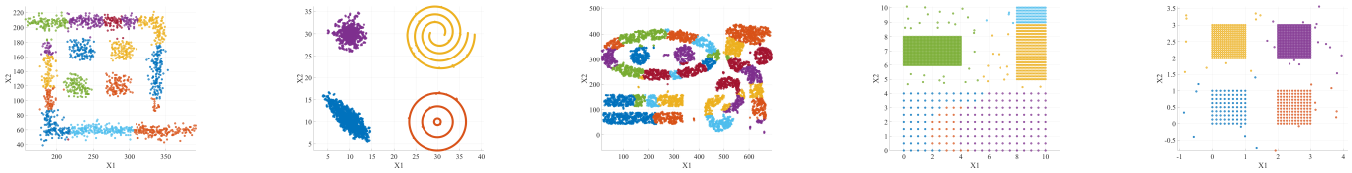


(a) (b) (c) (d) (e)

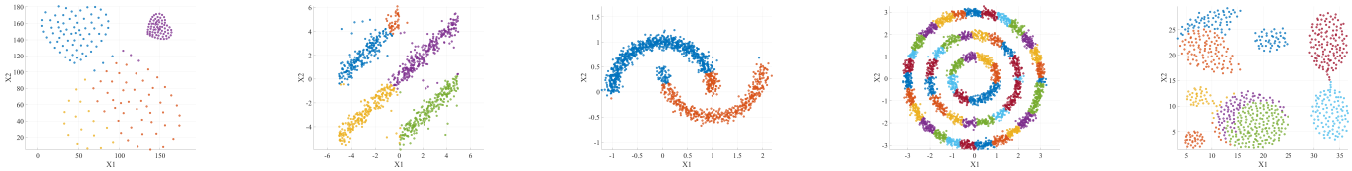


(f) (g) (h) (i) (j)

Fig. A.6: Clustering visualization of DBSCAN on synthetic datasets

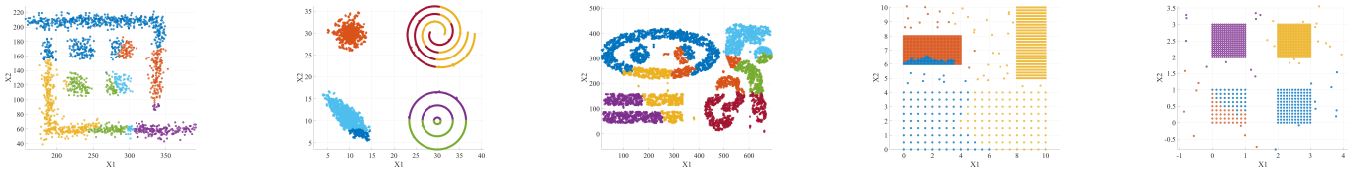


(a) (b) (c) (d) (e)

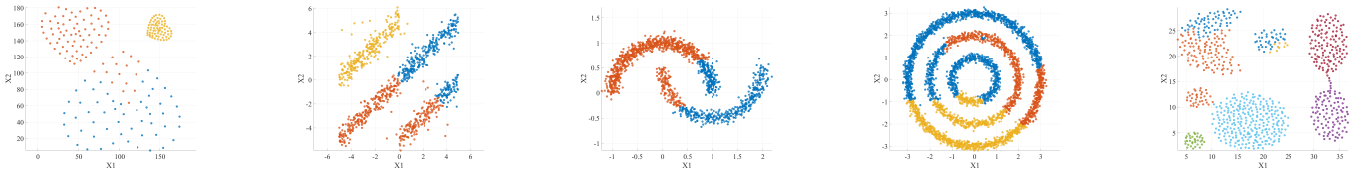


(f) (g) (h) (i) (j)

Fig. A.7: Clustering visualization of mean shift on synthetic datasets

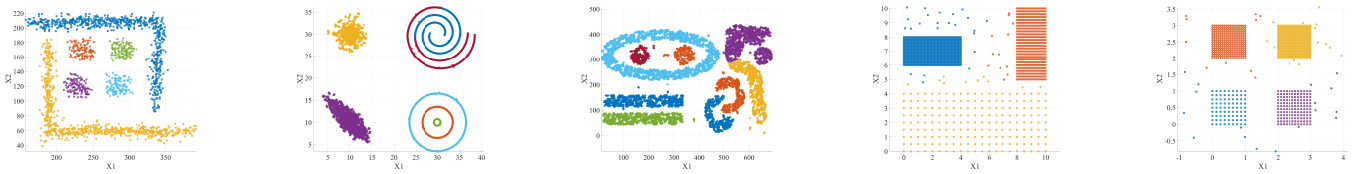


(a) (b) (c) (d) (e)

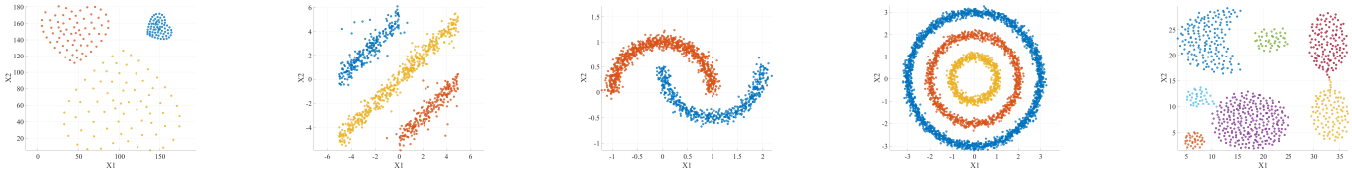


(f) (g) (h) (i) (j)

Fig. A.8: Clustering visualization of GBK-DPC on synthetic datasets

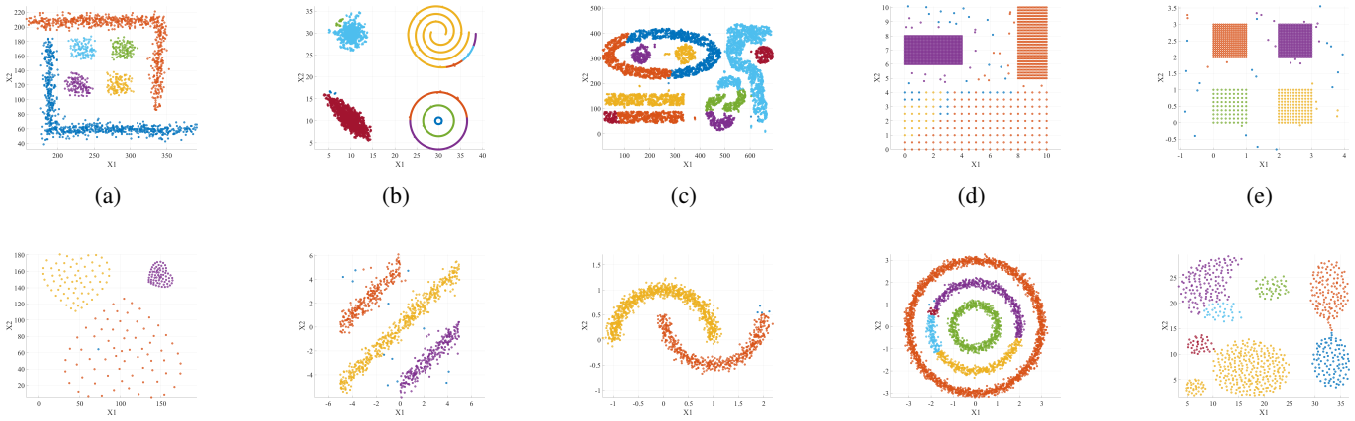


(a) (b) (c) (d) (e)



(f) (g) (h) (i) (j)

Fig. A.9: Clustering visualization of GBSC on synthetic datasets



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Fig. A.10: Clustering visualization of GBCT on synthetic datasets