CS685: Data Mining
Clustering

Arnab Bhattacharya
arnabb@cse.iitk.ac.in

Computer Science and Engineering,
Indian Institute of Technology, Kanpur
http://web.cse.iitk.ac.in/~cs685/

1st semester, 2020-21
Mon 1030-1200 (online)

Clustering

- A dataset of n objects $O_i, i = 1, \ldots, n$
- Partitioning of the dataset into k clusters or groups
- Can be
  - Crisp: Each object belongs to one and only one cluster
  - Fuzzy: An object belongs to a cluster with a probability; such probabilities add up to 1
- Unsupervised learning
- Sometimes very useful to learn structures in the data

Clustering

- A dataset of n objects $O_i, i = 1, \dots, n$
- Partitioning of the dataset into k clusters or groups
- Can be
  - Crisp: Each object belongs to one and only one cluster
  - Fuzzy: An object belongs to a cluster with a probability; such probabilities add up to 1
- Unsupervised learning
- Sometimes very useful to learn structures in the data
- Five main types
  - Partitioning-based
  - Hierarchical
    - *Agglomerative* or bottom-up
    - *Divisive* or top-down
  - Density-based
  - Grid-based
  - Model-based

Clustering Issues

- Scalability
  - Dealing with large datasets

Clustering Issues

- Scalability
  - Dealing with large datasets
- Types of objects
  - Vectors
  - Images
  - Documents
  - Sets

Clustering Issues

- Scalability
  - Dealing with large datasets
- Types of objects
  - Vectors
  - Images
  - Documents
  - Sets
- Measure of similarity
  - Euclidean distance
  - Similarity (or equivalently, distance) matrix

Clustering Issues

- Scalability
  - Dealing with large datasets
- Types of objects
  - Vectors
  - Images
  - Documents
  - Sets
- Measure of similarity
  - Euclidean distance
  - Similarity (or equivalently, distance) matrix
- Shape of cluster
  - Convex versus arbitrary

Clustering Issues

- Scalability
  - Dealing with large datasets
- Types of objects
  - Vectors
  - Images
  - Documents
  - Sets
- Measure of similarity
  - Euclidean distance
  - Similarity (or equivalently, distance) matrix
- Shape of cluster
  - Convex versus arbitrary
- Incremental
  - Ability to handle new data objects

Clustering Issues

- Scalability
  - Dealing with large datasets
- Types of objects
  - Vectors
  - Images
  - Documents
  - Sets
- Measure of similarity
  - Euclidean distance
  - Similarity (or equivalently, distance) matrix
- Shape of cluster
  - Convex versus arbitrary
- Incremental
  - Ability to handle new data objects
- Data input order
  - Sensitivity to order of input

Clustering Issues

- Scalability
  - Dealing with large datasets
- Types of objects
  - Vectors
  - Images
  - Documents
  - Sets
- Measure of similarity
  - Euclidean distance
  - Similarity (or equivalently, distance) matrix
- Shape of cluster
  - Convex versus arbitrary
- Incremental
  - Ability to handle new data objects
- Data input order
  - Sensitivity to order of input
- Noise
  - Detection of outliers
  - Noise objects as separate cluster

Cluster Evaluation Methods

- When *ground truth* is known, extrinsic methods or *supervised* methods are used

Cluster Evaluation Methods

- When *ground truth* is known, extrinsic methods or *supervised* methods are used
- Extrinsic methods evaluate based on four criteria

Cluster Evaluation Methods

- When *ground truth* is known, extrinsic methods or *supervised* methods are used
- Extrinsic methods evaluate based on four criteria
  - *Cluster homogeneity*: A purer cluster is better

Cluster Evaluation Methods

- When *ground truth* is known, extrinsic methods or *supervised* methods are used
- Extrinsic methods evaluate based on four criteria
  - *Cluster homogeneity*: A purer cluster is better
  - *Cluster completeness*: A more complete cluster, i.e., one that contains more points from the same category, is better

Cluster Evaluation Methods

- When *ground truth* is known, extrinsic methods or *supervised* methods are used
- Extrinsic methods evaluate based on four criteria
  - *Cluster homogeneity*: A purer cluster is better
  - *Cluster completeness*: A more complete cluster, i.e., one that contains more points from the same category, is better
  - *Rag bag*: It is better to cluster "heterogeneous" points in a separate "others" cluster

Cluster Evaluation Methods

- When *ground truth* is known, extrinsic methods or *supervised* methods are used
- Extrinsic methods evaluate based on four criteria
  - *Cluster homogeneity*: A purer cluster is better
  - *Cluster completeness*: A more complete cluster, i.e., one that contains more points from the same category, is better
  - *Rag bag*: It is better to cluster "heterogeneous" points in a separate "others" cluster
  - *Small cluster preservation*: Smaller clusters should be preserved more as otherwise they break into noise pieces

Cluster Evaluation Methods

- When *ground truth* is known, extrinsic methods or *supervised* methods are used
- Extrinsic methods evaluate based on four criteria
  - *Cluster homogeneity*: A purer cluster is better
  - *Cluster completeness*: A more complete cluster, i.e., one that contains more points from the same category, is better
  - *Rag bag*: It is better to cluster "heterogeneous" points in a separate "others" cluster
  - *Small cluster preservation*: Smaller clusters should be preserved more as otherwise they break into noise pieces
- When no "ground" truth is available, i.e, the actual clusters are not known,

Cluster Evaluation Methods

- When *ground truth* is known, extrinsic methods or *supervised* methods are used
- Extrinsic methods evaluate based on four criteria
  - *Cluster homogeneity*: A purer cluster is better
  - *Cluster completeness*: A more complete cluster, i.e., one that contains more points from the same category, is better
  - *Rag bag*: It is better to cluster "heterogeneous" points in a separate "others" cluster
  - *Small cluster preservation*: Smaller clusters should be preserved more as otherwise they break into noise pieces
- When no "ground" truth is available, i.e, the actual clusters are not known, use silhouette coefficient
- These are called intrinsic methods or *unsupervised* methods

Neighboring Cluster of an Object

- Suppose object O_i is in cluster A, i.e., $O_i \in A$
- Define a_i as the *average* distance of O_i to A

$$a_i = \frac{\sum_{p \in A} d(o_i, p)}{|A|}$$

- Similarly, define $d_i(C)$ to be the *average* distance of O_i to any other cluster C

$$d_i(C) = \frac{\sum_{q \in C} d(o_i, q)}{|C|}$$

Neighboring Cluster of an Object

- Suppose object O_i is in cluster A, i.e., $O_i \in A$
- Define a_i as the *average* distance of O_i to A

$$a_i = \frac{\sum_{p \in A} d(o_i, p)}{|A|}$$

- Similarly, define $d_i(C)$ to be the *average* distance of O_i to any other cluster C

$$d_i(C) = \frac{\sum_{q \in C} d(o_i, q)}{|C|}$$

- Suppose B is the cluster that minimizes this distance and b_i be the corresponding distance, i.e.,

$$b_i = \min d_i(C)$$
$$B = \arg\min d_i(C)$$

Neighboring Cluster of an Object

- Suppose object O_i is in cluster A, i.e., $O_i \in A$
- Define a_i as the *average* distance of O_i to A

$$a_i = \frac{\sum_{p \in A} d(o_i, p)}{|A|}$$

- Similarly, define $d_i(C)$ to be the *average* distance of O_i to any other cluster C

$$d_i(C) = \frac{\sum_{q \in C} d(o_i, q)}{|C|}$$

- Suppose B is the cluster that minimizes this distance and b_i be the corresponding distance, i.e.,

$$b_i = \min d_i(C)$$
$$B = \arg\min d_i(C)$$

- In a sense, cluster B is the "neighbor" of O_i
- O_i could have been in cluster B instead of A

Silhouette Index

- Silhouette index or silhouette coefficient of object O_i captures the difference of these two distances

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}$$

Silhouette Index

- Silhouette index or silhouette coefficient of object O_i captures the difference of these two distances

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}$$

- $-1 \leq s_i \leq +1$
    - If $s_i \rightarrow +1$, $b_i \gg a_i$ and O_i is in a good cluster
    - If $s_i \approx 0$, $b_i \approx a_i$ and O_i could have been in B as well
    - If $s_i < 0$, $b_i < a_i$ and O_i is better in B than current cluster

Silhouette Index

- Silhouette index or silhouette coefficient of object O_i captures the difference of these two distances

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}$$

- $-1 \le s_i \le +1$
  - If $s_i \to +1$, $b_i \gg a_i$ and O_i is in a good cluster
  - If $s_i \approx 0$, $b_i \approx a_i$ and O_i could have been in B as well
  - If $s_i < 0$, $b_i < a_i$ and O_i is better in B than current cluster
- Using these, average silhouette width of a cluster and of the entire dataset can be defined

Silhouette Index

- Silhouette index or silhouette coefficient of object O_i captures the difference of these two distances

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}$$

- $-1 \leq s_i \leq +1$
  - If $s_i \to +1$, $b_i \gg a_i$ and O_i is in a good cluster
  - If $s_i \approx 0$, $b_i \approx a_i$ and O_i could have been in B as well
  - If $s_i < 0$, $b_i < a_i$ and O_i is better in B than current cluster
- Using these, average silhouette width of a cluster and of the entire dataset can be defined
- Choose k that *maximizes* average silhouette width of the dataset, $\bar{s}_k$

Silhouette Index

- Silhouette index or silhouette coefficient of object O_i captures the difference of these two distances

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}$$

- $-1 \leq s_i \leq +1$
  - If $s_i \to +1$, $b_i \gg a_i$ and O_i is in a good cluster
  - If $s_i \approx 0$, $b_i \approx a_i$ and O_i could have been in B as well
  - If $s_i < 0$, $b_i < a_i$ and O_i is better in B than current cluster
- Using these, average silhouette width of a cluster and of the entire dataset can be defined
- Choose k that *maximizes* average silhouette width of the dataset, $\bar{s}_k$
- Different ranges of silhouette index
  - > 0.75: strong clustering
  - $0.5 - 0.75$: reasonable clustering
  - $0.25 - 0.5$: weak clustering
  - < 0.25: no structure

BCubed Measures

- C is a clustering on D
- Labels $l(O_i)$ are given as ideal (ground truth) for each $O_i \in D$
- For points O_i and O_j, correctness is agreement in ground truth

$$correctness(O_i, O_j) = \begin{cases} 1 & \text{if } l(O_i) = l(O_j) \Leftrightarrow C(O_i) = C(O_j) \\ 0 & \text{otherwise} \end{cases}$$

BCubed Measures

- C is a clustering on D
- Labels $I(O_i)$ are given as ideal (ground truth) for each $O_i \in D$
- For points O_i and O_j, correctness is agreement in ground truth

$$correctness(O_i, O_j) = \begin{cases} 1 & \text{if } I(O_i) = I(O_j) \Leftrightarrow C(O_i) = C(O_j) \\ 0 & \text{otherwise} \end{cases}$$

- BCubed precision measures fraction of same-cluster points that agree in ground truth

$$bcprecision = \frac{1}{n} \sum_{i=1}^{n} \frac{\sum_{O_j, i \neq j, C(O_i) = C(O_j)} correctness(O_i, O_j)}{|\{O_j, i \neq j, C(O_i) = C(O_j)\}|}$$

BCubed Measures

- C is a clustering on D
- Labels $l(O_i)$ are given as ideal (ground truth) for each $O_i \in D$
- For points O_i and O_j, correctness is agreement in ground truth

$$correctness(O_i, O_j) = \begin{cases} 1 & \text{if } l(O_i) = l(O_j) \Leftrightarrow C(O_i) = C(O_j) \\ 0 & \text{otherwise} \end{cases}$$

- BCubed precision measures fraction of same-cluster points that agree in ground truth

$$bcprecision = \frac{1}{n} \sum_{i=1}^{n} \frac{\sum_{O_j, i \neq j, C(O_i) = C(O_j)} correctness(O_i, O_j)}{|\{O_j, i \neq j, C(O_i) = C(O_j)\}|}$$

- BCubed recall measures fraction of same-ground truth points that agree in clustering

$$bcrecall = \frac{1}{n} \sum_{i=1}^{n} \frac{\sum_{O_j, i \neq j, l(O_i) = l(O_j)} correctness(O_i, O_j)}{|\{O_j, i \neq j, l(O_i) = l(O_j)\}|}$$

- Can define BCubed F-measure using these

RAND Index

- RAND Index or RAND Measure measures how similar two cluster outputs are
- If *ideal* clusters are known, this measures *quality* of a cluster output

RAND Index

- RAND Index or RAND Measure measures how similar two cluster outputs are
- If *ideal* clusters are known, this measures *quality* of a cluster output
- Suppose (ideal) clustering is $I = I_1, \dots, I_m$ where I_i are partitions
- Cluster to be measured is $C = C_1, \dots, C_k$
- Consider pairs of objects
  - a: Number of object pairs that are in the same cluster in both I and C
  - b: Number of object pairs that are in different clusters in both I and C
  - c: Number of object pairs that are in the same cluster in I but not in C
  - d: Number of object pairs that are in the same cluster in C but not in I

RAND Index

- RAND Index or RAND Measure measures how similar two cluster outputs are
- If *ideal* clusters are known, this measures *quality* of a cluster output
- Suppose (ideal) clustering is $I = I_1, \ldots, I_m$ where I_i are partitions
- Cluster to be measured is $C = C_1, \ldots, C_k$
- Consider pairs of objects
  - a: Number of object pairs that are in the same cluster in both I and C
  - b: Number of object pairs that are in different clusters in both I and C
  - c: Number of object pairs that are in the same cluster in I but not in C
  - d: Number of object pairs that are in the same cluster in C but not in I
- RAND Index is

$$R = \frac{a+b}{a+b+c+d} = \frac{a+b}{\binom{n}{2}} = \frac{TP+TN}{D}$$

Adjusted RAND Index

- Adjusted RAND Index (ARI) guards against random matches

Adjusted RAND Index

- Adjusted RAND Index (ARI) guards against random matches

$$ARI = \frac{\text{RAND Index} - \text{Expected RAND Index}}{\text{Maximum RAND Index} - \text{Expected RAND Index}}$$

Adjusted RAND Index

- Adjusted RAND Index (ARI) guards against random matches

$$ARI = \frac{\text{RAND Index} - \text{Expected RAND Index}}{\text{Maximum RAND Index} - \text{Expected RAND Index}}$$

- Contingency table of common objects

Clusters	C_1	$\cdots$	C_k	Total
I_1	n_{11}	$\cdots$	n_{1k}	i_1
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
I_m	n_{m1}	$\cdots$	n_{mk}	i_m
Total	c_1	$\cdots$	c_k	n

Adjusted RAND Index

- Adjusted RAND Index (ARI) guards against random matches

$$ARI = \frac{\text{RAND Index} - \text{Expected RAND Index}}{\text{Maximum RAND Index} - \text{Expected RAND Index}}$$

- Contingency table of common objects

Clusters	C_1	$\cdots$	C_k	Total
I_1	n_{11}	$\cdots$	n_{1k}	i_1
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
I_m	n_{m1}	$\cdots$	n_{mk}	i_m
Total	c_1	$\cdots$	c_k	n

- Expected number of pair matches assuming the same total distribution is

$$E\left[\sum_{i,j} \binom{n_{ij}}{2}\right] = \left[\sum_i \binom{n_{i\cdot}}{2} \cdot \sum_j \binom{n_{\cdot j}}{2}\right] \Big/ \binom{n}{2}$$

Adjusted RAND Index

- ARI can be written as

$$ARI = \frac{\sum_{i,j}\binom{n_{ij}}{2} - \left[\sum_i \binom{n_{i\cdot}}{2} \cdot \sum_j \binom{n_{\cdot j}}{2}\right] / \binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{n_{i\cdot}}{2} + \sum_j \binom{n_{\cdot j}}{2}\right] - \left[\sum_i \binom{n_{i\cdot}}{2} \cdot \sum_j \binom{n_{\cdot j}}{2}\right] / \binom{n}{2}}$$

- RAND Index is always between 0 and 1
- ARI is between -1 and $+1$
- ARI is negative when clustering is worse than random

Example

Clusters	C_1	C_2	C_3	Total
l_1	1	1	0	2
l_2	1	2	1	4
l_3	0	0	4	4
Total	2	3	5	10

- Number of pairs agreeing, i.e., $a = \binom{1}{2} + \cdots + \binom{4}{2} = 7$

Example

Clusters	C_1	C_2	C_3	Total
I_1	1	1	0	2
I_2	1	2	1	4
I_3	0	0	4	4
Total	2	3	5	10

- Number of pairs agreeing, i.e., $a = \binom{1}{2} + \cdots + \binom{4}{2} = 7$
- Number of pairs agreeing in I but not in C, i.e., $c = \binom{2}{2} + \binom{4}{2} + \binom{4}{2} - 7 = 13 - 7 = 6$
- Number of pairs agreeing in C but not in I, i.e., $d = \binom{2}{2} + \binom{3}{2} + \binom{5}{2} - 7 = 14 - 7 = 7$

Example

Clusters	C_1	C_2	C_3	Total
I_1	1	1	0	2
I_2	1	2	1	4
I_3	0	0	4	4
Total	2	3	5	10

- Number of pairs agreeing, i.e., $a = \binom{1}{2} + \cdots + \binom{4}{2} = 7$
- Number of pairs agreeing in I but not in C, i.e., $c = \binom{2}{2} + \binom{4}{2} + \binom{4}{2} - 7 = 13 - 7 = 6$
- Number of pairs agreeing in C but not in I, i.e., $d = \binom{2}{2} + \binom{3}{2} + \binom{5}{2} - 7 = 14 - 7 = 7$
- Therefore, number of pairs not agreeing, i.e., $b = \binom{10}{2} - 7 - 6 - 7 = 45 - 20 = 25$

Example

Clusters	C_1	C_2	C_3	Total
I_1	1	1	0	2
I_2	1	2	1	4
I_3	0	0	4	4
Total	2	3	5	10

- Number of pairs agreeing, i.e., $a = \binom{1}{2} + \cdots + \binom{4}{2} = 7$
- Number of pairs agreeing in I but not in C, i.e., $c = \binom{2}{2} + \binom{4}{2} + \binom{4}{2} - 7 = 13 - 7 = 6$
- Number of pairs agreeing in C but not in I, i.e., $d = \binom{2}{2} + \binom{3}{2} + \binom{5}{2} - 7 = 14 - 7 = 7$
- Therefore, number of pairs not agreeing, i.e., $b = \binom{10}{2} - 7 - 6 - 7 = 45 - 20 = 25$
- RAND Index is $\frac{7+25}{45} = 0.71$

Example

Clusters	C_1	C_2	C_3	Total
I_1	1	1	0	2
I_2	1	2	1	4
I_3	0	0	4	4
Total	2	3	5	10

- Number of pairs agreeing, i.e., $a = \binom{1}{2} + \cdots + \binom{4}{2} = 7$
- Number of pairs agreeing in I but not in C, i.e., $c = \binom{2}{2} + \binom{4}{2} + \binom{4}{2} - 7 = 13 - 7 = 6$
- Number of pairs agreeing in C but not in I, i.e., $d = \binom{2}{2} + \binom{3}{2} + \binom{5}{2} - 7 = 14 - 7 = 7$
- Therefore, number of pairs not agreeing, i.e., $b = \binom{10}{2} - 7 - 6 - 7 = 45 - 20 = 25$
- RAND Index is $\frac{7+25}{45} = 0.71$
- ARI is $\frac{7-(13\times14)/45}{(13+14)/2-(13\times14)/45} = 0.31$

Clustering Tendency

- Uniform data is poorly clustered
- Clustering tendency measures how likely that clusters exist

Clustering Tendency

- Uniform data is poorly clustered
- Clustering tendency measures how likely that clusters exist
- Measures how close it is to the *uniform distribution*

Clustering Tendency

- Uniform data is poorly clustered
- Clustering tendency measures how likely that clusters exist
- Measures how close it is to the *uniform distribution*
- Hopkin's statistic
- Generate p uniformly random points from the data space
- For each point, measure its distance u_i to nearest neighbor

Clustering Tendency

- Uniform data is poorly clustered
- Clustering tendency measures how likely that clusters exist
- Measures how close it is to the *uniform distribution*
- Hopkin's statistic
- Generate p uniformly random points from the data space
- For each point, measure its distance u_i to nearest neighbor
- Sample p points from the dataset
- For each point, measure its distance w_i to nearest neighbor

Clustering Tendency

- Uniform data is poorly clustered
- Clustering tendency measures how likely that clusters exist
- Measures how close it is to the *uniform distribution*
- Hopkin's statistic
- Generate p uniformly random points from the data space
- For each point, measure its distance u_i to nearest neighbor
- Sample p points from the dataset
- For each point, measure its distance w_i to nearest neighbor
- Hopkin's statistic is

$$H = \frac{\sum_{i=1}^{p} w_i}{\sum_{i=1}^{p} u_i + \sum_{i=1}^{p} w_i}$$

Clustering Tendency

- Uniform data is poorly clustered
- Clustering tendency measures how likely that clusters exist
- Measures how close it is to the *uniform distribution*
- Hopkin's statistic
- Generate p uniformly random points from the data space
- For each point, measure its distance u_i to nearest neighbor
- Sample p points from the dataset
- For each point, measure its distance w_i to nearest neighbor
- Hopkin's statistic is

$$H = \frac{\sum_{i=1}^{p} w_i}{\sum_{i=1}^{p} u_i + \sum_{i=1}^{p} w_i}$$

- If $H \approx 0.5$,

Clustering Tendency

- Uniform data is poorly clustered
- Clustering tendency measures how likely that clusters exist
- Measures how close it is to the *uniform distribution*
- Hopkin's statistic
- Generate p uniformly random points from the data space
- For each point, measure its distance u_i to nearest neighbor
- Sample p points from the dataset
- For each point, measure its distance w_i to nearest neighbor
- Hopkin's statistic is

$$H = \frac{\sum_{i=1}^{p} w_i}{\sum_{i=1}^{p} u_i + \sum_{i=1}^{p} w_i}$$

- If $H \approx 0.5$, then data is mostly uniform
- If $H \to 0$,

Clustering Tendency

- Uniform data is poorly clustered
- Clustering tendency measures how likely that clusters exist
- Measures how close it is to the *uniform distribution*
- Hopkin's statistic
- Generate p uniformly random points from the data space
- For each point, measure its distance u_i to nearest neighbor
- Sample p points from the dataset
- For each point, measure its distance w_i to nearest neighbor
- Hopkin's statistic is

$$H = \frac{\sum_{i=1}^{p} w_i}{\sum_{i=1}^{p} u_i + \sum_{i=1}^{p} w_i}$$

- If $H \approx 0.5$, then data is mostly uniform
- If $H \to 0$, then data is clustered
- If $H \to 1$,

Clustering Tendency

- Uniform data is poorly clustered
- Clustering tendency measures how likely that clusters exist
- Measures how close it is to the *uniform distribution*
- Hopkin's statistic
- Generate p uniformly random points from the data space
- For each point, measure its distance u_i to nearest neighbor
- Sample p points from the dataset
- For each point, measure its distance w_i to nearest neighbor
- Hopkin's statistic is

$$H = \frac{\sum_{i=1}^{p} w_i}{\sum_{i=1}^{p} u_i + \sum_{i=1}^{p} w_i}$$

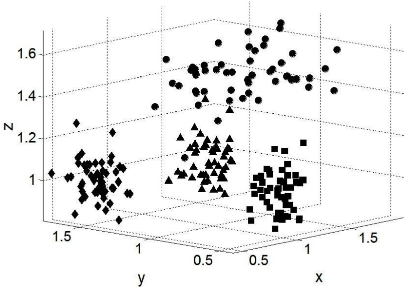
- If $H \approx 0.5$, then data is mostly uniform
- If $H \to 0$, then data is clustered
- If $H \to 1$, then data is regular or sparse

Subspace Clustering

- Clustering is typically applied on all the dimensions
- May *miss* out structures present in lower dimensional subspaces
- Clusters in lower dimensional spaces are *not* identified in higher dimensions due to randomness in other values

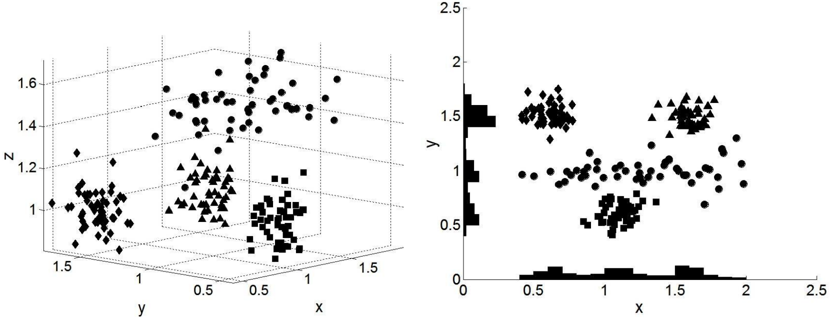
Subspace Clustering

- Clustering is typically applied on all the dimensions
- May *miss* out structures present in lower dimensional subspaces
- Clusters in lower dimensional spaces are *not* identified in higher dimensions due to randomness in other values
- $\diamond$, $\square$ and $\triangle$ points cluster
- $\bigcirc$ points weakly cluster



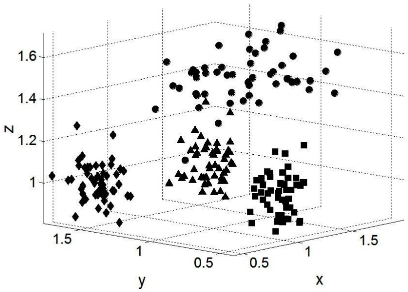
Clusters in xy Subspace

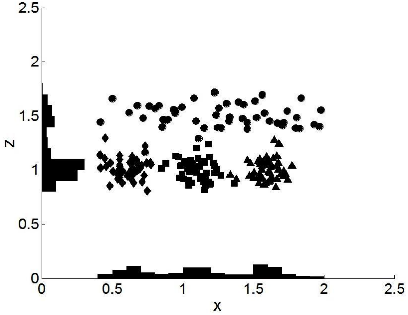
- In x space: $\diamond$, $\square$ and $\triangle$ points cluster but $\bigcirc$ acts as noise
- In y space: $\square$ separate, $\bigcirc$ separate, $\diamond$ and $\triangle$ together
- In xy space: $\diamond$, $\square$ and $\triangle$ points cluster and $\bigcirc$ points weakly cluster



Clusters in xz Subspace

- In x space: $\diamond$, $\square$ and $\triangle$ points cluster but $\bigcirc$ acts as noise
- In z space: $\bigcirc$ separate, $\square$, $\diamond$ and $\triangle$ together
- In xz space: $\diamond$, $\square$ and $\triangle$ points cluster and $\bigcirc$ points weakly cluster





Clusters in yz Subspace

- In y space: $\bigcirc$ separate, $\square$ separate, $\diamond$ and $\triangle$ together
- In z space: $\bigcirc$ separate, $\square$, $\diamond$ and $\triangle$ together
- In yz space: $\bigcirc$ separate, $\square$ separate, $\diamond$ and $\triangle$ together