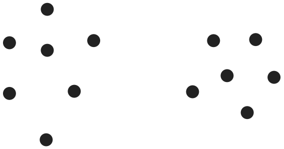
What do we seek in a hierarchical clustering?

Justin Eldridge, Mikhail Belkin, Yusu Wang
The Ohio State University

May 13, 2015

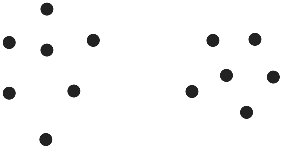
The goal of clustering:

Identify structure in data by grouping it into *clusters*

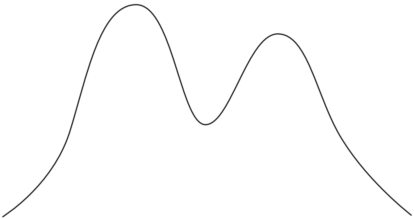


The goal of clustering:

Identify structure in data by grouping it into *clusters*



Assumption: data is generated by some source with structure. This structure is what we *actually* want to recover.



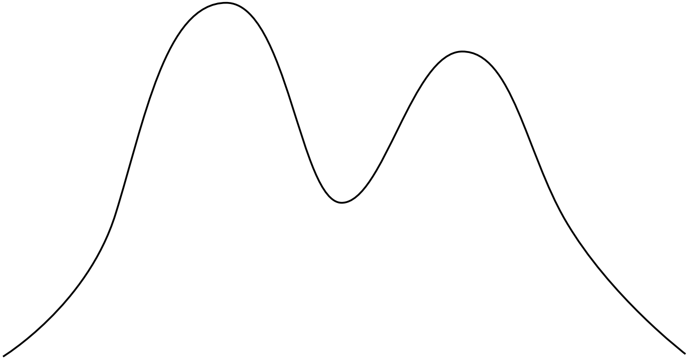
Theory of clustering

Given a data source (i.e., a density):

- How is a cluster defined?
- What cluster structure do we wish to recover?

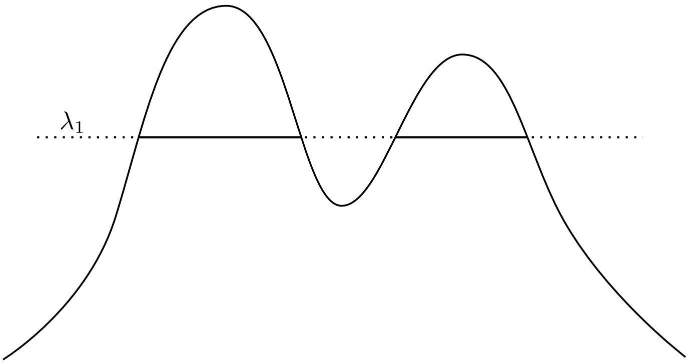
How do we define a cluster of a density f?

A region of high density: Hartigan (1981), Wishart (1969)...



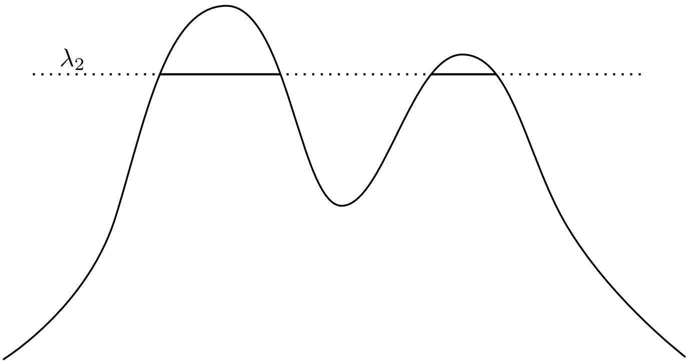
High-density clusters

Connected components of $\{f \geq \lambda_1\}$?



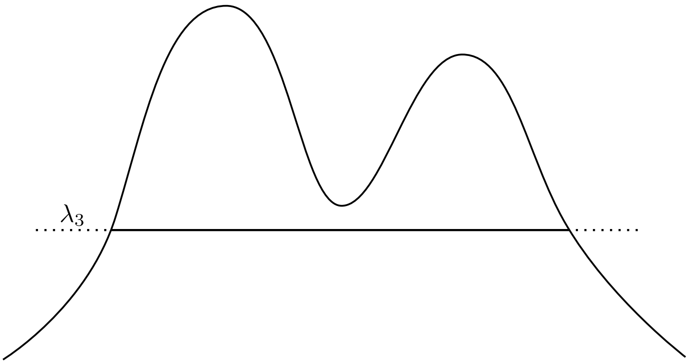
High-density clusters

Connected components of $\{f \geq \lambda_2\}$?



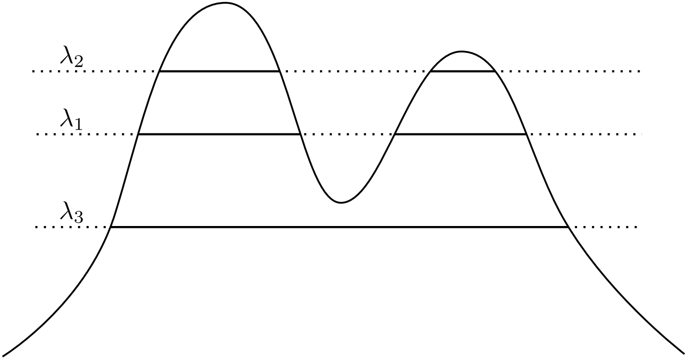
High-density clusters

Connected components of $\{f \geq \lambda_3\}$?



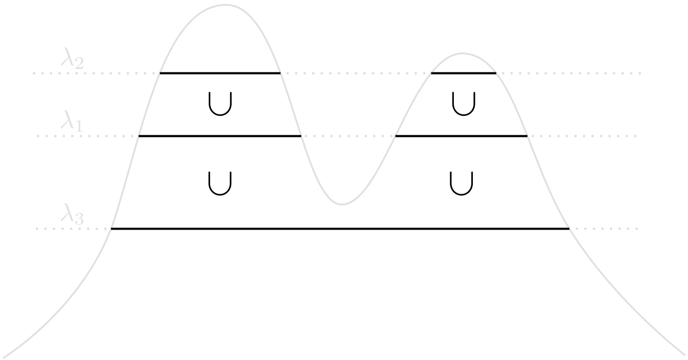
High-density clusters

A cluster is a connected component of $\{f \geq \lambda\}$ for any $\lambda > 0$.



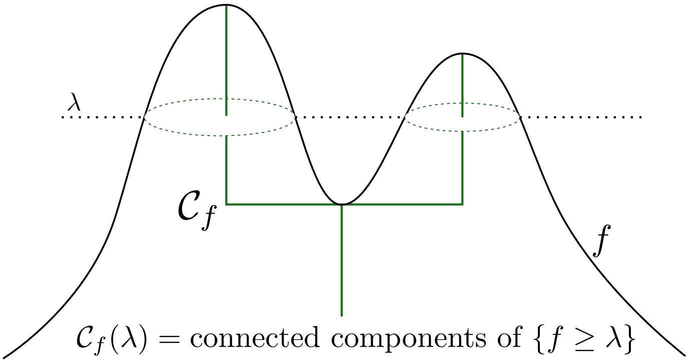
A hierarchy of clusters

Clusters from higher levels nest within clusters from lower levels.



The density cluster tree

This gives rise to a tree structure called the *density cluster tree*.



Theory of clustering a density

Given a density:

- How is a cluster defined?

Theory of clustering a density

Given a density:

- How is a cluster defined?
    - A region of high density.

Theory of clustering a density

Given a density:

- How is a cluster defined?
  - A region of high density.
  - Formally: A connected component of a superlevel set.

Theory of clustering a density

Given a density:

- How is a cluster defined?
  - A region of high density.
  - Formally: A connected component of a superlevel set.
- What cluster structure do we wish to recover?

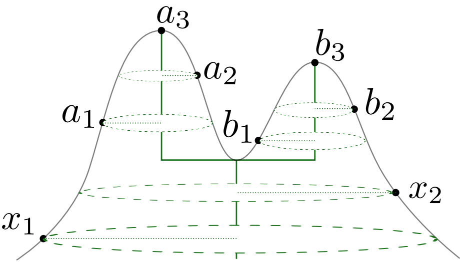
Theory of clustering a density

Given a density:

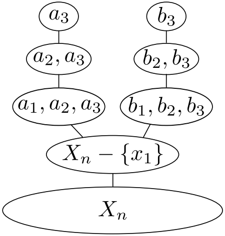
- How is a cluster defined?
  - A region of high density.
  - Formally: A connected component of a superlevel set.
- What cluster structure do we wish to recover?
  - The density cluster tree.

Theory of clustering a density

- But... We typically do not have access to the density.
- Recover density cluster tree $\mathcal{C}_f$ by clustering finite data.
- Algorithm outputs finite cluster tree $\hat{\mathcal{C}}_{f,n}$ whose nodes are *empirical clusters*.



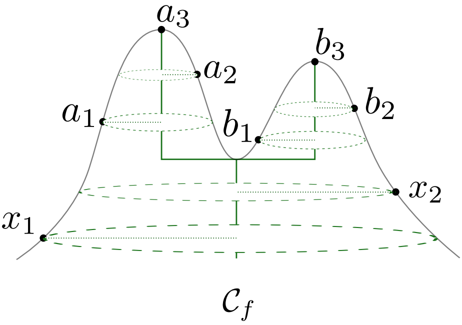
$\mathcal{C}_f$

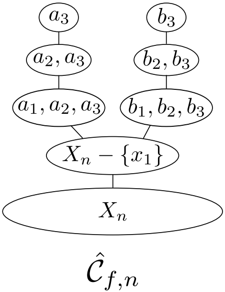


$\hat{\mathcal{C}}_{f,n}$

Theory of clustering a density

- But... We typically do not have access to the density.
- Recover density cluster tree $\mathcal{C}_f$ by clustering finite data.
- Algorithm outputs finite cluster tree $\hat{\mathcal{C}}_{f,n}$ whose nodes are *empirical clusters*.





$\hat{\mathcal{C}}_{f,n}$ is a collection of clusters with hierarchical structure

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?

2. How "close" is a clustering to the ideal density cluster tree?

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
2. How "close" is a clustering to the ideal density cluster tree?
3. Do algorithms exist which have these properties/converge to the true density cluster tree?

Theory of clustering a density: our contributions

We answer these questions by:

Theory of clustering a density: our contributions

We answer these questions by:

1. identifying the properties desirable in a clustering,

Theory of clustering a density: our contributions

We answer these questions by:

1. identifying the properties desirable in a clustering,
2. introducing a metric on cluster trees,

Theory of clustering a density: our contributions

We answer these questions by:

1. identifying the properties desirable in a clustering,
2. introducing a metric on cluster trees,
3. showing that convergence implies our properties,

Theory of clustering a density: our contributions

We answer these questions by:

1. identifying the properties desirable in a clustering,
2. introducing a metric on cluster trees,
3. showing that convergence implies our properties,
4. proving convergence for two algorithms.

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?

2. How "close" is a clustering to the ideal density cluster tree?

3. Do algorithms exist which have these properties/converge to the true density cluster tree?

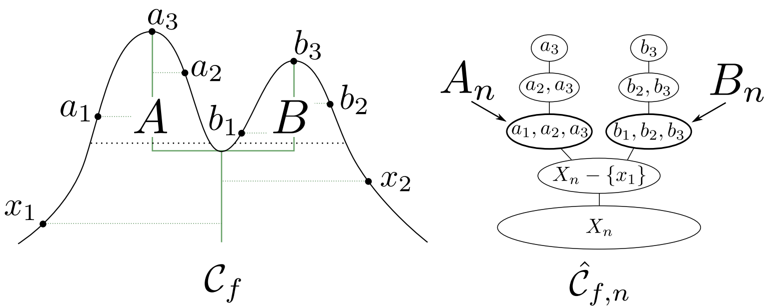
Hartigan consistency

To answer this, Hartigan (1981) defined notion of *consistency*:

- A density f supported on $\mathcal{X}$
- A sample $X_n \sim f$
- A method producing an estimate $\hat{\mathcal{C}}_{f,n}$ of the density cluster tree

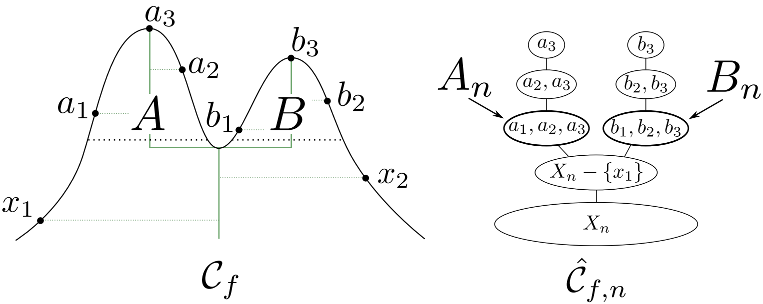
A method is *Hartigan consistent* if as $n \to \infty$, any two disjoint clusters in the density cluster tree of f are kept separate by $\hat{\mathcal{C}}_{f,n}$

Hartigan consistency



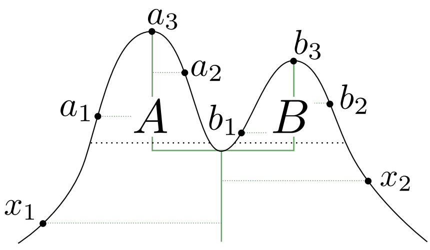
- Notation: For any set $A \subset \mathcal{X}$, let A_n denote the smallest cluster of $\hat{\mathcal{C}}_{f,n}$ containing $A \cap X_n$

Hartigan consistency

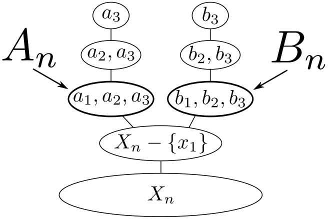


- Notation: For any set $A \subset \mathcal{X}$, let A_n denote the smallest cluster of $\hat{\mathcal{C}}_{f,n}$ containing $A \cap X_n$
- A_n is the "tightest" empirical cluster recovering A

Hartigan consistency



$\mathcal{C}_f$



$\hat{\mathcal{C}}_{f,n}$

- Notation: For any set $A \subset \mathcal{X}$, let A_n denote the smallest cluster of $\hat{\mathcal{C}}_{f,n}$ containing $A \cap X_n$
- A_n is the "tightest" empirical cluster recovering A
- Consistency: whenever A and B are different connected components of $\{f \geq \lambda\}$, $\Pr(A_n \text{ is disjoint from } B_n) \to 1$ as $n \to \infty$.

Are there algorithms which are Hartigan consistent?

- Hartigan analyzed single linkage, showed that it is not consistent in $d > 1$

Are there algorithms which are Hartigan consistent?

- Hartigan analyzed single linkage, showed that it is not consistent in $d > 1$
- 30 years pass...

Are there algorithms which are Hartigan consistent?

- Hartigan analyzed single linkage, showed that it is not consistent in $d > 1$
- 30 years pass...
- Several algorithms have been shown to be consistent:
  - Robust single linkage (Chaudhuri and Dasgupta, 2010)
  - Tree pruning (Kpotufe and von Luxburg, 2011)

Are there algorithms which are Hartigan consistent?

- Hartigan analyzed single linkage, showed that it is not consistent in $d > 1$
- 30 years pass...
- Several algorithms have been shown to be consistent:
  - Robust single linkage (Chaudhuri and Dasgupta, 2010)
  - Tree pruning (Kpotufe and von Luxburg, 2011)
- And so these capture the structure of the density cluster tree

Are there algorithms which are Hartigan consistent?

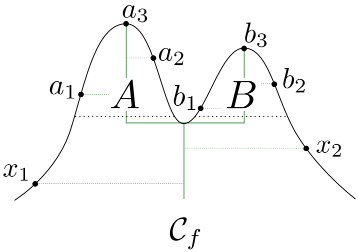
- Hartigan analyzed single linkage, showed that it is not consistent in $d > 1$
- 30 years pass...
- Several algorithms have been shown to be consistent:
  - Robust single linkage (Chaudhuri and Dasgupta, 2010)
  - Tree pruning (Kpotufe and von Luxburg, 2011)
- And so these capture the structure of the density cluster tree
- *Or maybe not...*

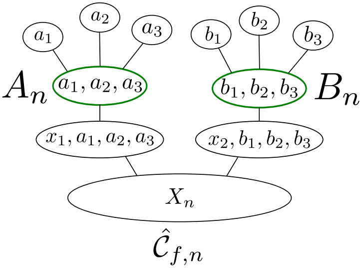
- Hartigan consistency is clearly *desirable*

- Hartigan consistency is clearly *desirable*
- *But is it sufficient?*

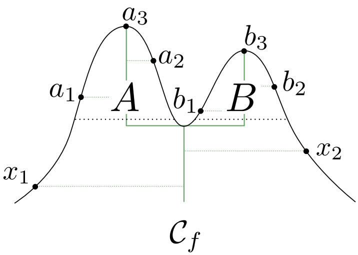
- Hartigan consistency is clearly *desirable*
- *But is it sufficient?*
- Three ways to be consistent, yet very different than true tree:
  1. Over-segmentation
  2. Improper nesting
  3. Laziness

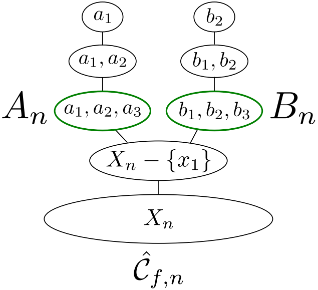
Issue #1: Over-segmentation



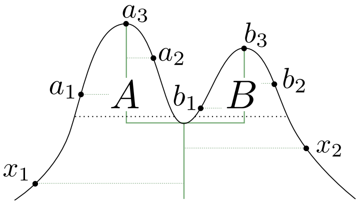


Issue #2: Improper nesting

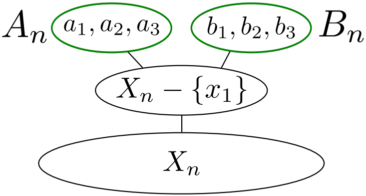




Issue #3: Laziness



$\mathcal{C}_f$



$\hat{\mathcal{C}}_{f,n}$

The root cause: non-minimality

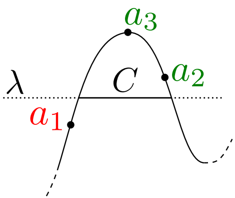
We identify over-segmentation, improper nesting, and laziness as manifestations of one issue: *non-minimality*

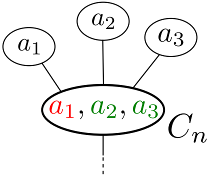
The root cause: non-minimality

We identify over-segmentation, improper nesting, and laziness as manifestations of one issue: *non-minimality*

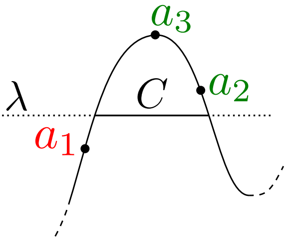
Definition (Non-minimality)

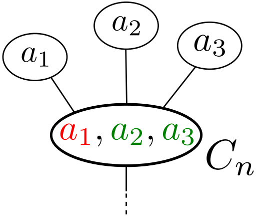
- Let C be a connected component of $\{f \geq \lambda\}$.
- Let C_n be the smallest empirical cluster containing all of $C \cap X_n$
- C_n is *non-minimal* if it contains extra points that aren't in C



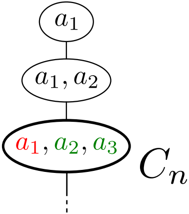


Non-minimality

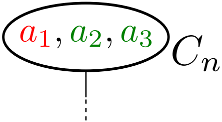




over-segmentation

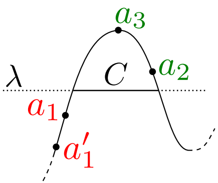


improper nesting



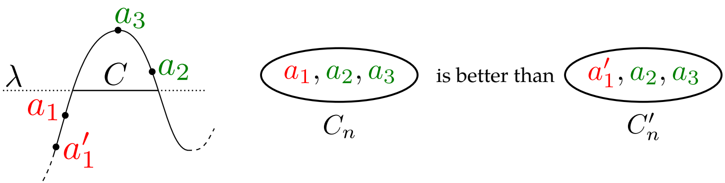
laziness

δ-non-minimality



a_1, a_2, a_3 (C_n) is better than a'_1, a_2, a_3 (C'_n)

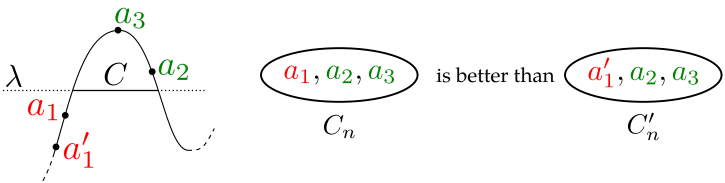
δ-non-minimality



Definition (δ-non-minimality)

- Let C be a connected component of $\{f \geq \lambda\}$.
- Let C_n be the smallest empirical cluster containing all of $C \cap X_n$
- We say that C_n is *δ-non-minimal* if $\min_{x \in C_n} f(x) < \lambda - \delta$.

δ-non-minimality



Definition (δ-non-minimality)

- Let C be a connected component of $\{f \geq \lambda\}$.
- Let C_n be the smallest empirical cluster containing all of $C \cap X_n$
- We say that C_n is *δ-non-minimal* if $\min_{x \in C_n} f(x) < \lambda - \delta$.

We want $\delta \to 0$ as $n \to \infty$!

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
   - Hartigan consistency is not sufficient

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
   - Hartigan consistency is not sufficient

- We introduce two *new* properties: *minimality* and *separation*

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
   - Hartigan consistency is not sufficient

- We introduce two *new* properties: *minimality* and *separation*
- We show that *minimality* + *separation* $\implies$ consistency

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
    - Hartigan consistency is not sufficient

- We introduce two *new* properties: *minimality* and *separation*
- We show that *minimality* + *separation* $\Longrightarrow$ consistency
- But first we need some definitions...

Cluster tree with height function

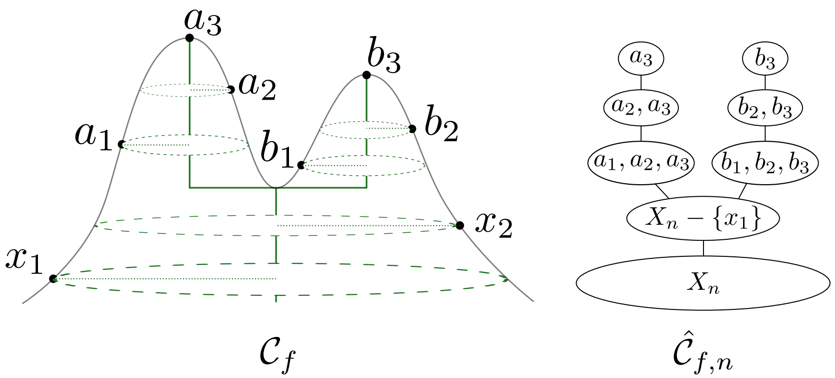
Definition

A cluster tree with a height function is a triple $\mathsf{C} = (X, \mathcal{C}, h)$, where:

- X is a set of objects,
- $\mathcal{C}$ is a cluster tree of X,
- and $h : X \to \mathbb{R}$ is a height function mapping each point in X to a "height".

We define the height of a cluster $C \in \mathcal{C}$ to be the infimal height of any point in the cluster. That is, $h(C) = \inf_{x \in C} h(x)$.

Cluster tree with height function



$h(\{b_3\}) = f(b_3)$ $\quad$ $h(\{b_2, b_3\}) = f(b_2)$ $\quad$ $h(\{b_1, b_2, b_3\}) = f(b_1)$

Connectedness and separation

Definition

- Points x and x' are *connected at level λ* if there is a cluster C containing both, with $h(C) \geq \lambda$
- Otherwise they are *separated at level λ*

Connectedness and separation

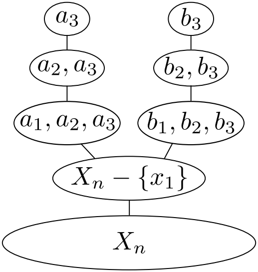
Definition

- Points x and x' are *connected at level λ* if there is a cluster C containing both, with $h(C) \geq \lambda$
- Otherwise they are *separated at level λ*
- A set S is *connected at level λ* if any two points $s, s' \in S$ are connected at level λ
- Sets S and S' are *separated at level λ* if for any $s \in S$, $s' \in S'$, s and s' are separated at level λ

Merge height

Definition

The *merge height* of two points x and x', written $m_C(x, x')$, is the height of the smallest cluster containing both.



$\hat{\mathcal{C}}_{f,n}$

$$m_{\hat{\mathcal{C}}_{f,n}}(a_3, b_3) = h(X_n - \{x_1\})$$
$$m_{\hat{\mathcal{C}}_{f,n}}(a_2, a_3) = h(\{a_2, a_3\})$$

Non-minimality revisited

Recall:

Definition (δ-non-minimality)

- Let A be a connected component of $\{f \geq \lambda\}$.
- Let A_n be the smallest cluster of $\hat{\mathcal{C}}_{f,n}$ containing all of $A \cap X_n$
- We say that A_n is *δ-non-minimal* if $\min_{x \in A_n} f(x) < \lambda - \delta$.

In other words:

- We say that A_n is *δ-non-minimal* if $h(A_n) < \lambda - \delta$
- That is, $A \cap X_n$ is not connected at level $\lambda - \delta$

Non-minimality revisited

Recall:

Definition (δ-non-minimality)

- Let A be a connected component of $\{f \geq \lambda\}$.
- Let A_n be the smallest cluster of $\hat{\mathcal{C}}_{f,n}$ containing all of $A \cap X_n$
- We say that A_n is *δ-non-minimal* if $\min_{x \in A_n} f(x) < \lambda - \delta$.

In other words:

- We say that A_n is *δ-non-minimal* if $h(A_n) < \lambda - \delta$
- That is, $A \cap X_n$ is not connected at level $\lambda - \delta$

We want $A \cap X_n$ to be connected at level $\lambda - \delta$, with δ small.

Minimality

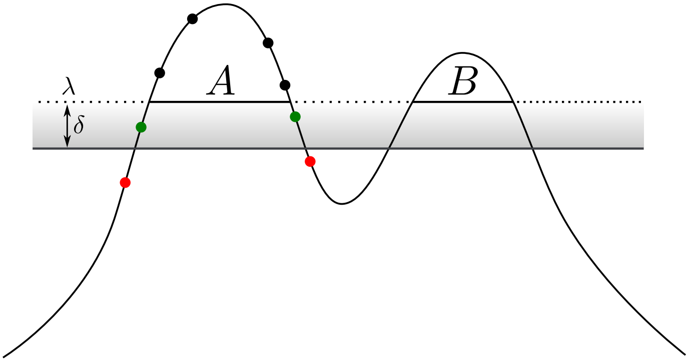
Definition (Minimality)

A method ensures *minimality* if given any cluster A of $\{f \geq \lambda\}$, $A \cap X_n$ is connected at level $\lambda - \delta$ in $\hat{C}_{f,n}$ for any $\delta > 0$ as $n \to \infty$

Where:

- f is a density supported on $\mathcal{X}$
- $X_n \sim f$
- $\hat{C}_{f,n}$ is an estimate of the true cluster tree, equipped with f as a height function

Minimality



Separation

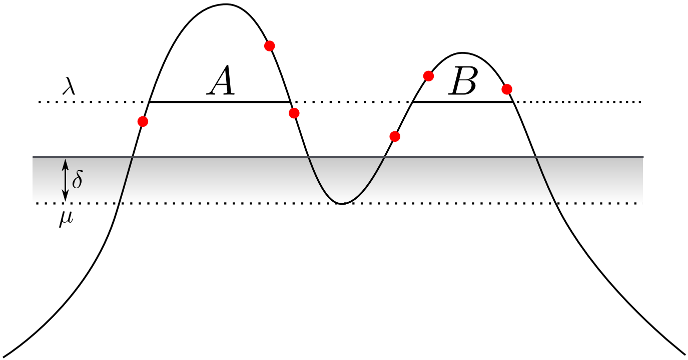
Definition (Separation)

A method ensures *separation* if when A and B are two disjoint connected components of $\{f \geq \lambda\}$ merging at $\mu = m_{C_f}(A \cup B)$, $A \cap X_n$ and $B \cap X_n$ are separated at level $\mu + \delta$ in $\hat{C}_{f,n}$ for any $\delta > 0$ as $n \to \infty$.

Where:

- f is a density supported on $\mathcal{X}$
- $X_n \sim f$
- $\hat{C}_{f,n}$ is an estimate of the true cluster tree, equipped with f as a height function

Separation



Minimality + separation $\implies$ Hartigan consistency

Theorem

If a hierarchical clustering method ensures both separation and minimality, then it is Hartigan consistent.

Uniform minimality and separation

Definition

We say that $\hat{C}_{f,n}$ ensures *uniform minimality* if given any $\delta > 0$ there exists an N depending only on δ such that for all $n \geq N$ and all λ, any cluster $A \in \{x \in \mathcal{X} : f(x) \geq \lambda\}$ is connected at level $\lambda - \delta$.

Definition

$\hat{C}_{f,n}$ is said to ensure *uniform separation* if given any $\delta > 0$ there exists an N depending only on δ such that for all $n \geq N$ and all μ, any two disjoint connected components merging in $\{x \in \mathcal{X} : f(x) \geq \mu\}$ are separated at level $\mu + \delta$.

Uniform minimality and separation

Theorem

If the density f is:

- *bounded from above,*
- *such that $\{f \geq \lambda\}$ contains finitely many connected components for any λ,*

then

- *minimality $\Longrightarrow$ uniform minimality*
- *separation $\Longrightarrow$ uniform separation*

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
   - *Minimality* and *separation*
2. How "close" is a clustering to the ideal density cluster tree?
3. Do algorithms exist which have these properties/converge to the true density cluster tree?

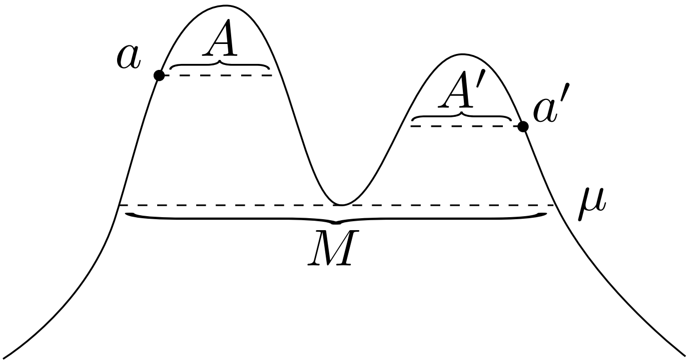
Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
   - Minimality and separation
2. How "close" is a clustering to the ideal density cluster tree?
3. Do algorithms exist which have these properties/converge to the true density cluster tree?

Merge distortion metric

- We have introduced two desirable limit properties: *minimality* and *separation*
- But we may want a *quantitative* measure of convergence
- We introduce a distance between the true cluster tree and an estimate
- Convergence will imply *minimality* and *separation*

Merge distortion metric



At what height do a and a' merge in $\hat{C}_{f,n}$?

- *Minimality*: M connected at $\mu - \delta$, with $\delta \to 0$
- *Separation*: A and A' separated at $\mu + \delta$, with $\delta \to 0$
- Minimality + separation $\Longrightarrow m_{\hat{C}_{f,n}}(a, a') = \mu \pm \delta$

Merge distortion metric

We define the *merge distortion metric* as:

$$d(\mathrm{C}_f, \hat{\mathrm{C}}_{f,n}) = \max_{x,x' \in X_n} |m_{\mathrm{C}_f}(x, x') - m_{\hat{\mathrm{C}}_{f,n}}(x, x')|.$$

where:

- $X_n \sim f$
- C_f is the true cluster tree of f
- $\hat{\mathrm{C}}_{f,n}$ is the estimated cluster tree
- Each tree is equipped with f as height function
- m_{C_f} and $m_{\hat{\mathrm{C}}_{f,n}}$ are the merge heights in C_f and $\hat{\mathrm{C}}_{f,n}$

Convergence to the density cluster tree

Definition

We say that a sequence of cluster trees $\{\hat{\mathrm{C}}_{f,n}\}$ converges to the high density cluster tree C_f of f, written $\hat{\mathrm{C}}_{f,n} \to \mathrm{C}_f$, if for any $\varepsilon > 0$ there exists an N such that for all $n \geq N$, $d(\hat{\mathrm{C}}_{f,n}, \mathrm{C}_f) < \varepsilon$.

Note

Convergence in the merge distortion metric implies that for any two points x and x', $|m_{\mathrm{C}_f}(x, x') - m_{\hat{\mathrm{C}}_{f,n}}(x, x')| \to 0$ as $n \to \infty$.

Properties of convergence

Theorem

$\hat{C}_{f,n} \to C_f$ *implies 1) uniform minimality and 2) uniform separation.*

Properties of convergence

Theorem
$\hat{C}_{f,n} \to C_f$ implies 1) uniform minimality and 2) uniform separation.

Theorem
If $\hat{C}_{f,n}$ ensures uniform separation and uniform minimality, then $\hat{C}_{f,n} \to C_f$.

Merge distortion metric (general cluster trees)

Definition

Let $\mathsf{C}_1 = (X_1, \mathcal{C}_1, h_1)$ and $\mathsf{C}_2 = (X_2, \mathcal{C}_2, h_2)$ be two hierarchical clusterings equipped with height functions. Let $S_1 \subset X_1$ and $S_2 \subset X_2$. Let $\gamma \subset S_1 \times S_2$ be a correspondence between S_1 and S_2. The distance between C_1 and C_2 with respect to γ is defined as

$$d_\gamma(\mathsf{C}_1, \mathsf{C}_2) = \max_{(x_1, x_2), (x_1', x_2') \in \gamma} |m_{\mathsf{C}_1}(x_1, x_1') - m_{\mathsf{C}_2}(x_2, x_2')|$$

L_∞-stability of true cluster tree

Theorem

- *Given a density* $f : \mathcal{X} \to \mathbb{R}$ *and a perturbed density* $\tilde{f} : \mathcal{X} \to \mathbb{R}$
- *Let* $\mathcal{C}_f$ *and* $\mathcal{C}_{\tilde{f}}$ *be density cluster trees*
- *Let* $\mathrm{C}_f := (\mathcal{X}, \mathcal{C}_f, f)$ *and* $\mathrm{C}_{\tilde{f}} := (\mathcal{X}, \mathcal{C}_{\tilde{f}}, \tilde{f})$ *be the cluster trees equipped with height functions*

Then we have $d(\mathrm{C}_f, \mathrm{C}_{\tilde{f}}) \leq \|f - \tilde{f}\|_\infty$

L_∞-stability w.r.t. f

Theorem

Let $\mathsf{C}_1 := (\mathcal{X}, \mathcal{C}, f_1)$ *and* $\mathsf{C}_2 := (\mathcal{X}, \mathcal{C}, f_2)$ *be two cluster trees with height functions. Then we have* $d(\mathsf{C}_1, \mathsf{C}_2) \leq 2\|f_1 - f_2\|_\infty$.

Some implications:

- If $\hat{f}$ is a consistent density estimate, then true cluster tree of $\hat{f}$ converges to true cluster tree of f
- Then estimated cluster tree of $\hat{f}$ converges to true cluster tree of f
- This justifies sampling points until distance between consecutive estimated cluster trees is small

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
    - Minimality and separation
2. How "close" is a clustering to the ideal density cluster tree?
    - *Merge distortion metric*
3. Do algorithms exist which have these properties/converge to the true density cluster tree?

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
   - Minimality and separation
2. How "close" is a clustering to the ideal density cluster tree?
   - Merge distortion metric
3. Do algorithms exist which have these properties/converge to the true density cluster tree?

Are there algorithms which converge in our metric?

- Convergence in the merge distortion metric implies several nice properties.
- Do algorithms exist which have these properties?

Are there algorithms which converge in our metric?

- Convergence in the merge distortion metric implies several nice properties.
- Do algorithms exist which have these properties?
- Kleinberg makes us nervous...

Are there algorithms which converge in our metric?

- Convergence in the merge distortion metric implies several nice properties.
- Do algorithms exist which have these properties?
- Kleinberg makes us nervous...
  - Gave three desirable properties
    1. Scale invariance
    2. Richness
    3. Consistency

Are there algorithms which converge in our metric?

- Convergence in the merge distortion metric implies several nice properties.
- Do algorithms exist which have these properties?
- Kleinberg makes us nervous...
  - Gave three desirable properties
    1. Scale invariance
    2. Richness
    3. Consistency
  - Showed that no algorithm exists

Are there algorithms which converge in our metric?

- Convergence in the merge distortion metric implies several nice properties.
- Do algorithms exist which have these properties?
- Kleinberg makes us nervous...
  - Gave three desirable properties
    1. Scale invariance
    2. Richness
    3. Consistency
  - Showed that no algorithm exists
- But indeed there are. We analyze two:
  - Split tree clustering
  - Robust single linkage (Chaudhuri and Dasgupta, 2010)

Robust single linkage (Chaudhuri and Dasgupta, 2010)

Given a sample X_n of n points, and parameters α and k:

1. For each $x_i \in X_n$, set $r_k(x_i)$ to be the distance from x_i to its kth neighbor.
2. As r grows from 0 to ∞:
   - Construct a graph G_r with nodes $\{x_i : r_k(x_i) \leq r\}$.
   - Include edge (x_i, x_j) if $\|x_i - x_j\| \leq \alpha r$.
   - The *clusters* at time r are the connected components of G_r.

Robust single linkage (Chaudhuri and Dasgupta, 2010)

Given a sample X_n of n points, and parameters α and k:

1. For each $x_i \in X_n$, set $r_k(x_i)$ to be the distance from x_i to its kth neighbor.
2. As r grows from 0 to ∞:
   - Construct a graph G_r with nodes $\{x_i : r_k(x_i) \leq r\}$.
   - Include edge (x_i, x_j) if $\|x_i - x_j\| \leq \alpha r$.
   - The *clusters* at time r are the connected components of G_r.

Output is a finite cluster tree.

Convergence of robust single linkage

- Robust single linkage is Hartigan consistent

Convergence of robust single linkage

- Robust single linkage is Hartigan consistent
- It also converges to the true tree in our metric

Convergence of robust single linkage

- Robust single linkage is Hartigan consistent
- It also converges to the true tree in our metric

Theorem

Suppose f is L-Lipschitz, compactly supported, and for any λ, $\{f \geq \lambda\}$ has finitely-many connected components
Then:

- *Robust single linkage ensures uniform minimality and uniform separation*
- *Therefore robust single linkage converges to the true cluster tree in the merge distortion metric*

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
   - ~~Hartigan consistency~~ (allows non-minimality)

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
   - ~~Hartigan consistency~~ (allows non-minimality)
   - Minimality and separation

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
   - ~~Hartigan consistency~~ (allows non-minimality)
   - Minimality and separation
   - Minimality + separation $\Longrightarrow$ Hartigan

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
    - ~~Hartigan consistency~~ (allows non-minimality)
    - Minimality and separation
    - Minimality + separation $\Longrightarrow$ Hartigan

2. How "close" is a clustering to the ideal density cluster tree?

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
   - ~~Hartigan consistency~~ (allows non-minimality)
   - Minimality and separation
   - Minimality + separation $\Longrightarrow$ Hartigan
2. How "close" is a clustering to the ideal density cluster tree?
   - Merge distortion metric

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
   - ~~Hartigan consistency~~ (allows non-minimality)
   - Minimality and separation
   - Minimality + separation $\implies$ Hartigan

2. How "close" is a clustering to the ideal density cluster tree?
   - Merge distortion metric
   - Convergence $\iff$ (uniform) minimality and separation

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
    - ~~Hartigan consistency~~ (allows non-minimality)
    - Minimality and separation
    - Minimality + separation $\implies$ Hartigan

2. How "close" is a clustering to the ideal density cluster tree?
    - Merge distortion metric
    - Convergence $\iff$ (uniform) minimality and separation

3. Do algorithms exist which have these properties/converge to the true density cluster tree?

Theory of clustering a density

1. What properties ensure that an algorithm captures the true density cluster tree?
   - ~~Hartigan consistency~~ (allows non-minimality)
   - Minimality and separation
   - Minimality + separation $\implies$ Hartigan

2. How "close" is a clustering to the ideal density cluster tree?
   - Merge distortion metric
   - Convergence $\iff$ (uniform) minimality and separation

3. Do algorithms exist which have these properties/converge to the true density cluster tree?
   - Robust single linkage