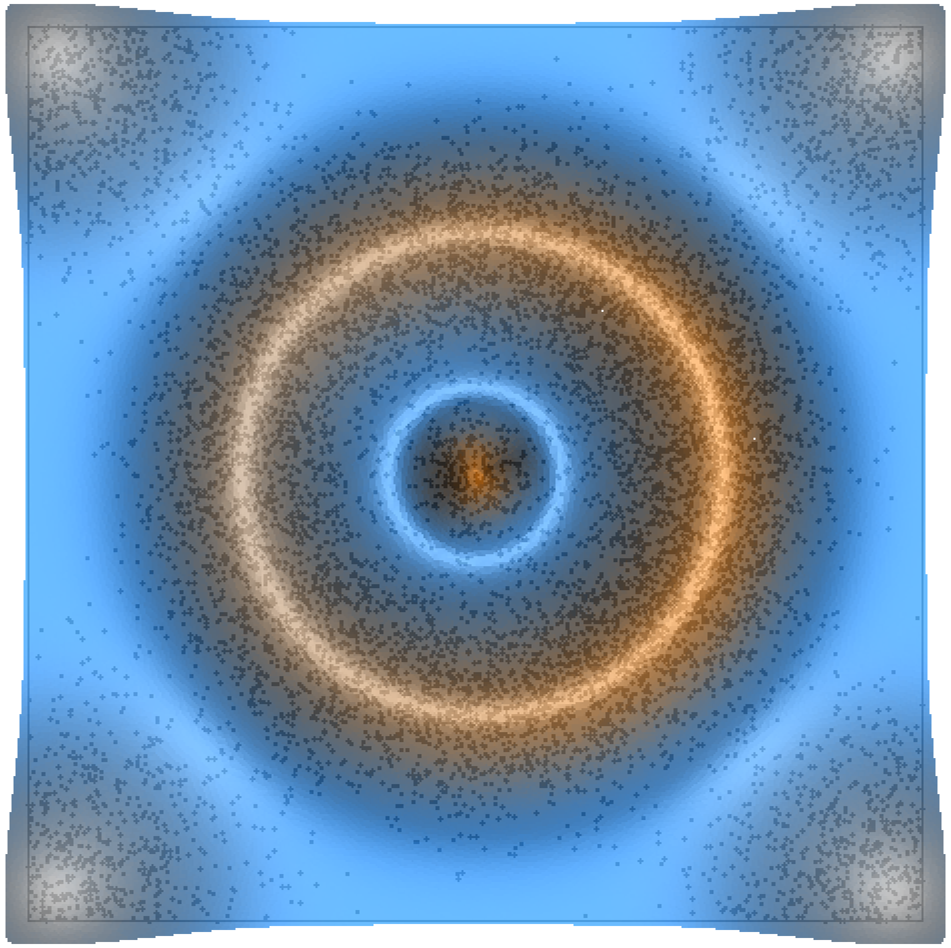
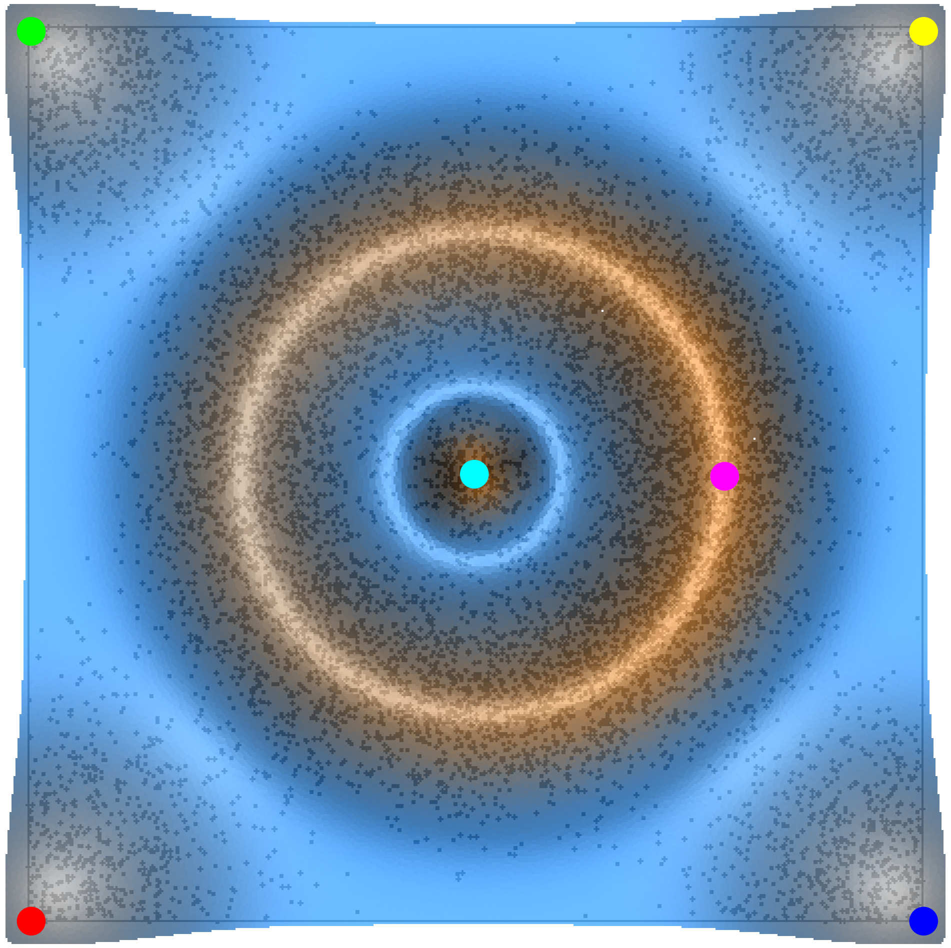
Mode-Seeking Paradigm

- Assume the data points are sampled from some unknown probability distribution
- Partition the data according to the basins of attraction of the peaks of the density



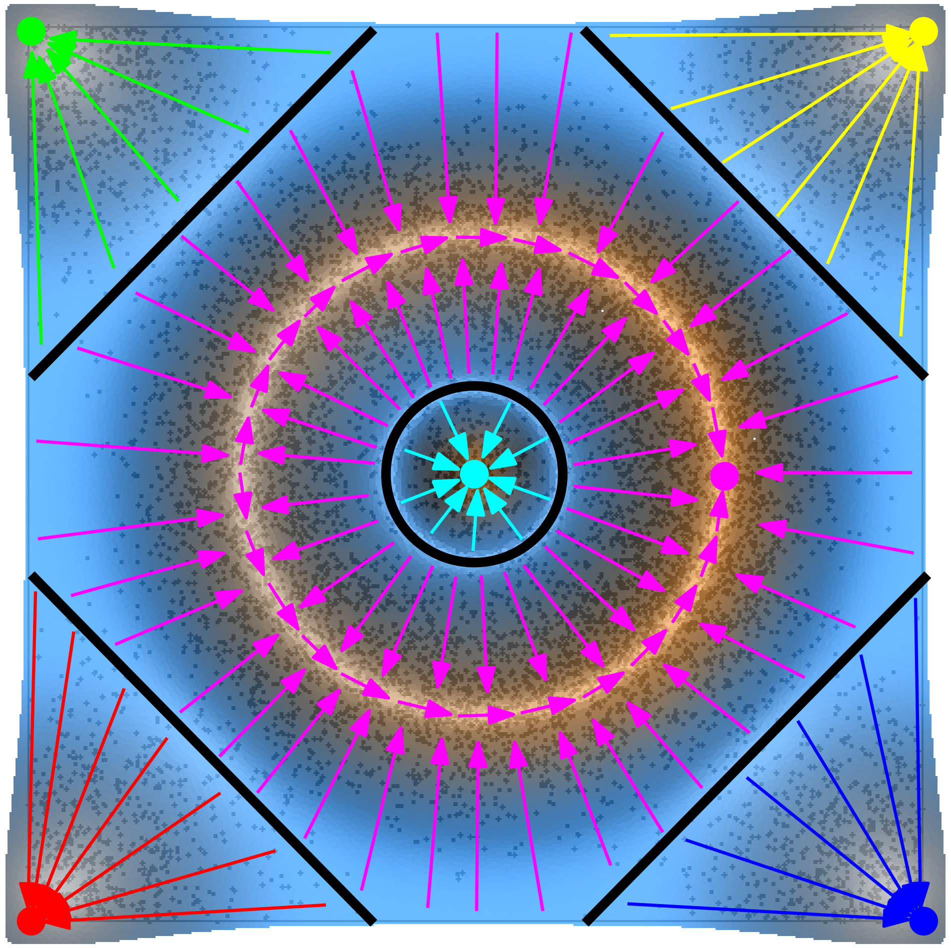
Mode-Seeking Paradigm

- Assume the data points are sampled from some unknown probability distribution
- Partition the data according to the basins of attraction of the peaks of the density



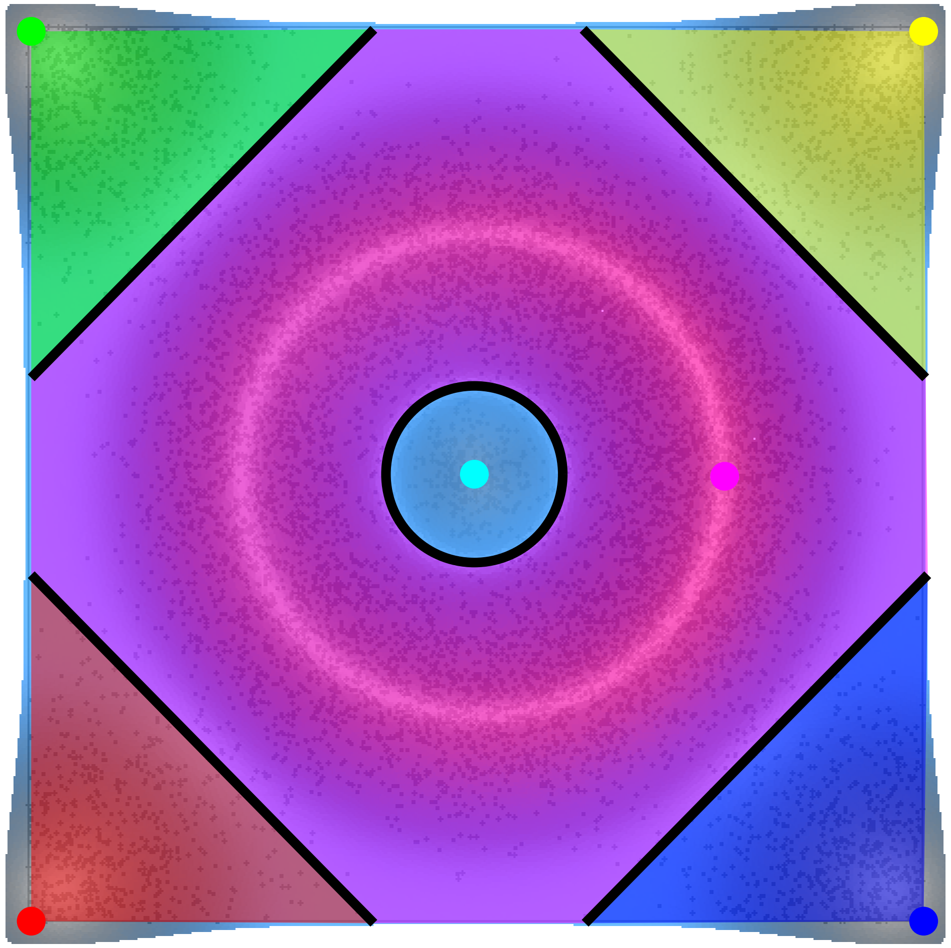
Mode-Seeking Paradigm

- Assume the data points are sampled from some unknown probability distribution
- Partition the data according to the basins of attraction of the peaks of the density



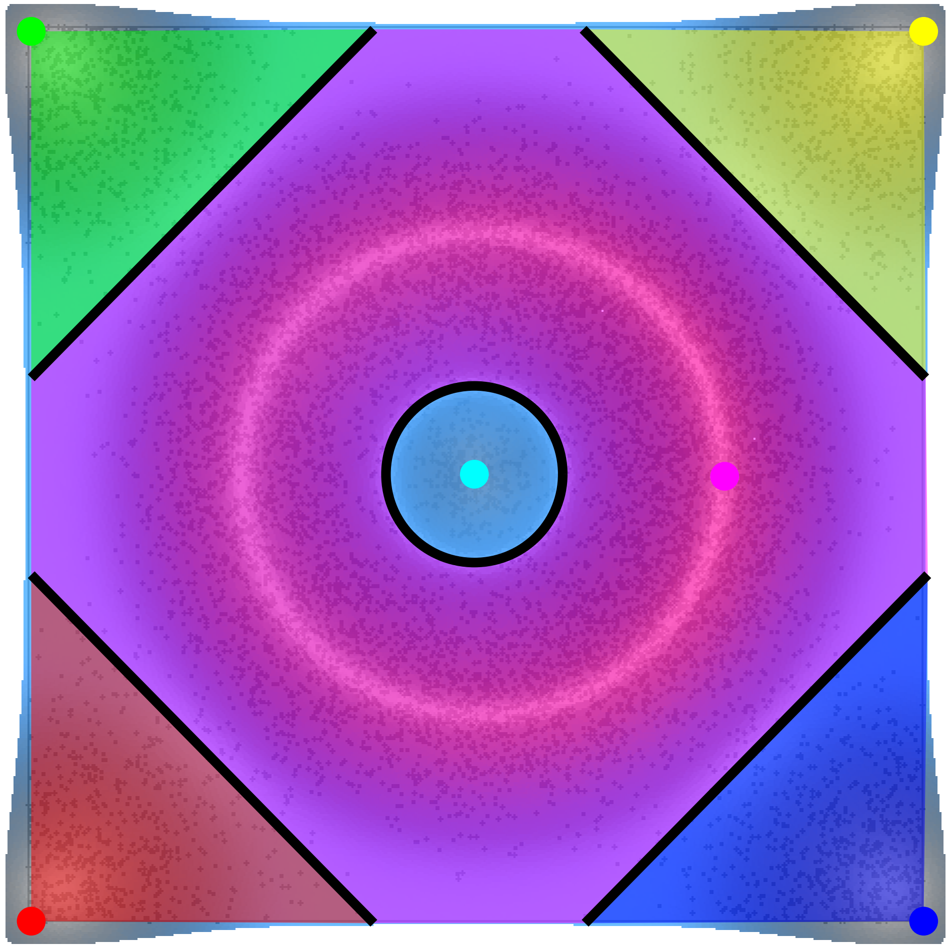
Mode-Seeking Paradigm

- Assume the data points are sampled from some unknown probability distribution
- Partition the data according to the basins of attraction of the peaks of the density



Mode-Seeking Paradigm

- Assume the data points are sampled from some $\boxed{\text{unknown}}$ probability distribution
- Partition the data according to the basins of attraction of the peaks of the density

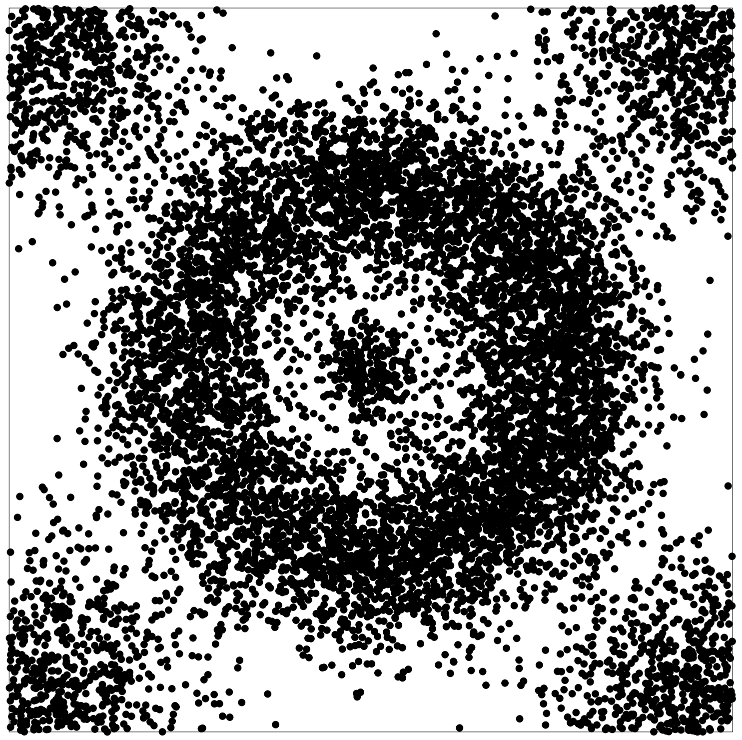


Hill-Climbing Schemes

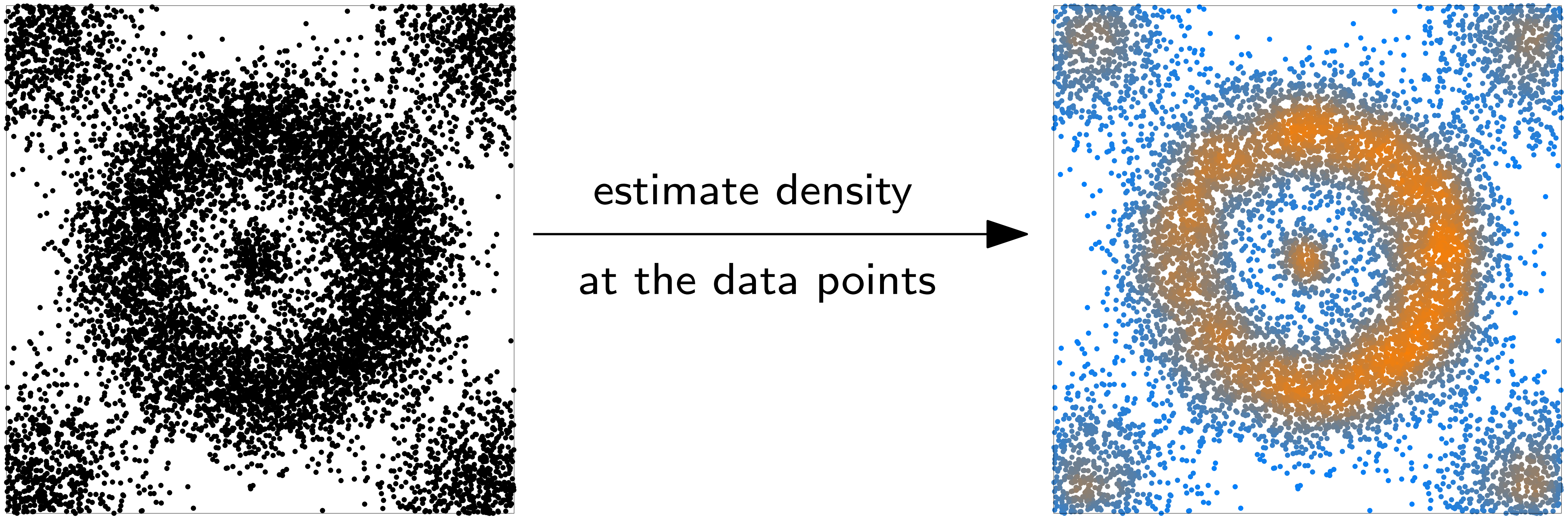
- **Iterative**, e.g. D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(5):603619, May 2002.

- **Non-iterative**, e.g. W. L. Koontz, P. M. Narendra, and K. Fukunaga. A graph-theoretic approach to nonparametric cluster analysis. *IEEE Trans. on Computers*, 24:936944, September 1976.

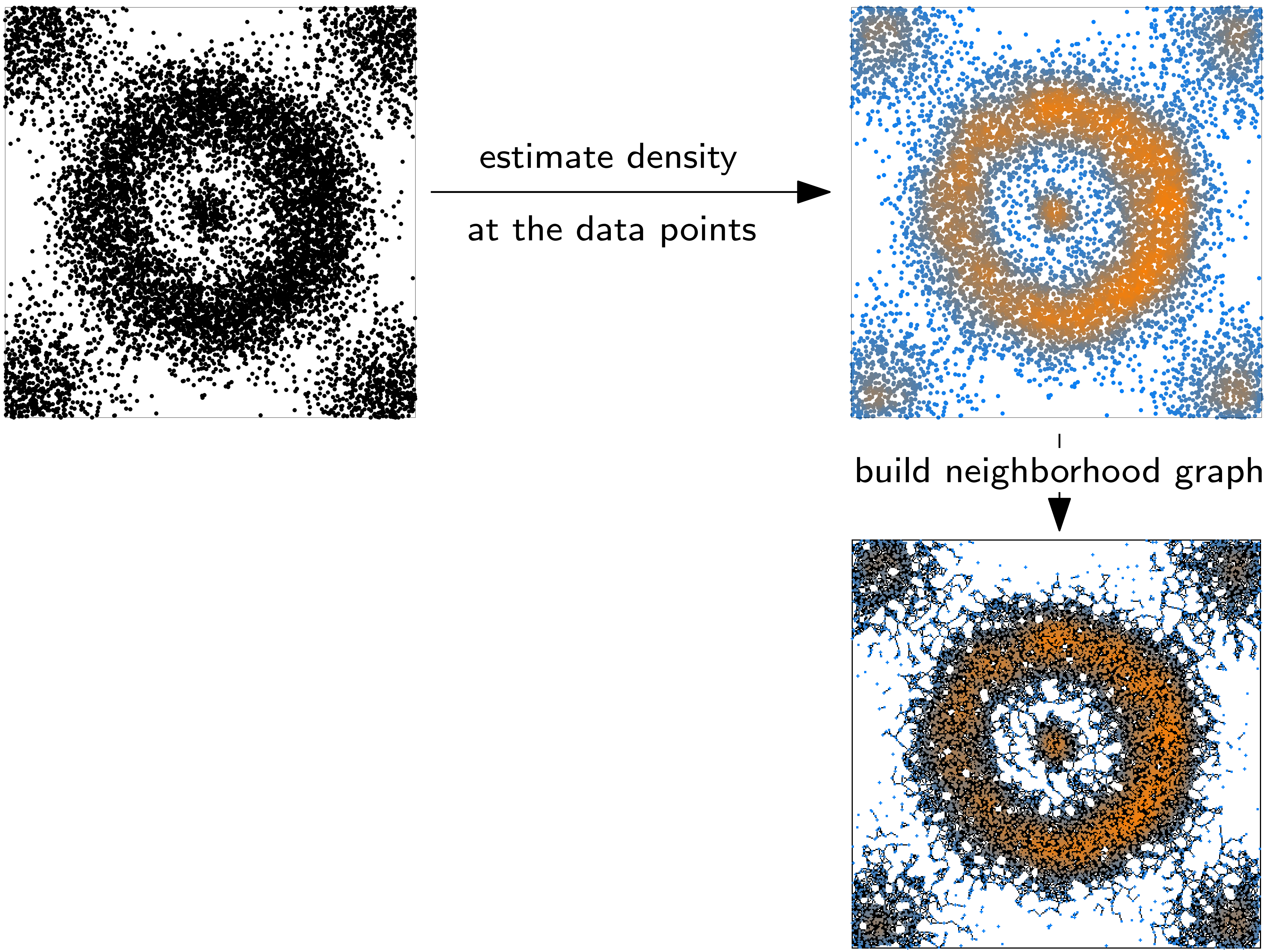
[Koontz, Narendra, Fukunaga'76] in a Nutshell



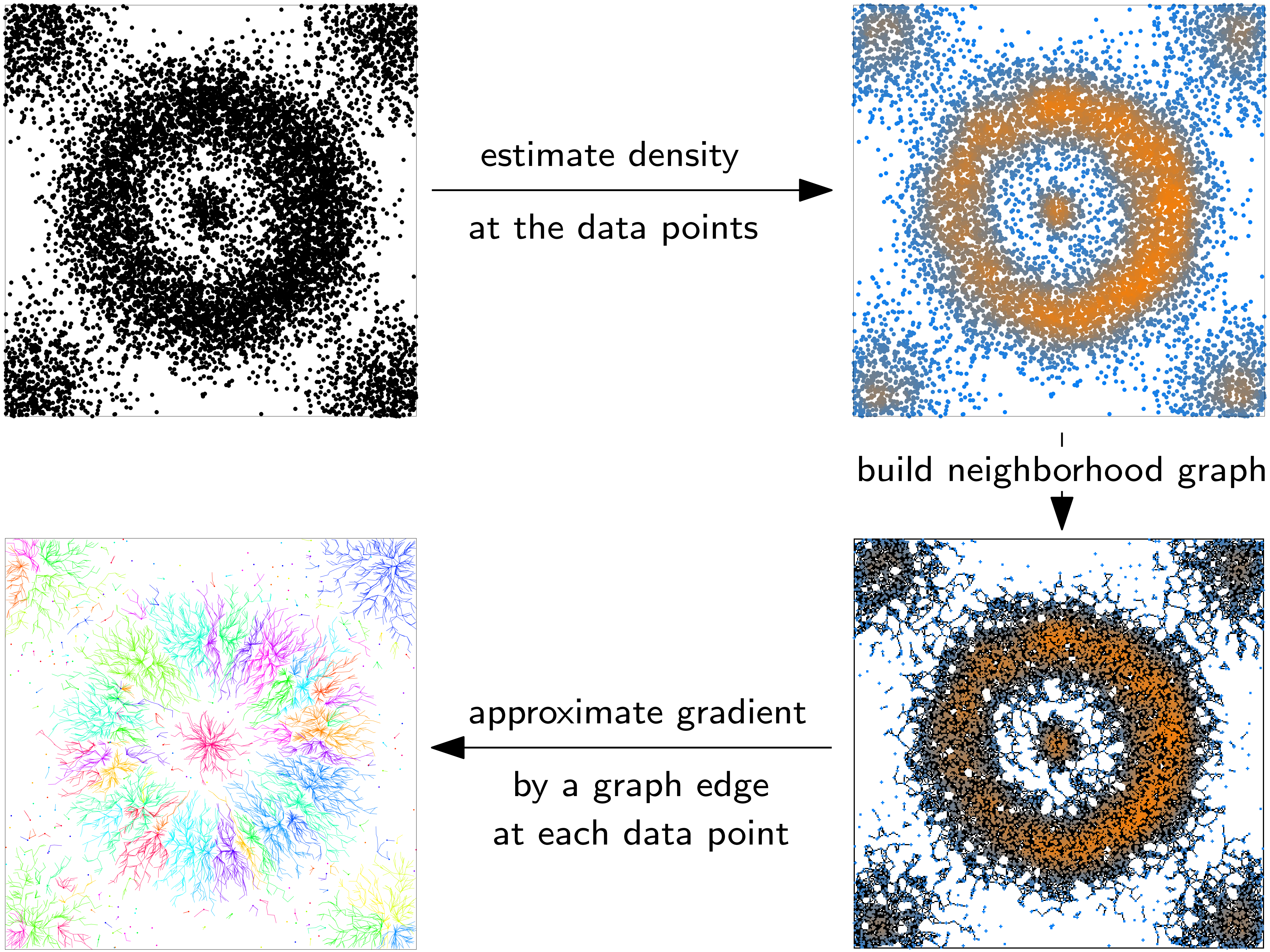
[Koontz, Narendra, Fukunaga'76] in a Nutshell



[Koontz, Narendra, Fukunaga'76] in a Nutshell

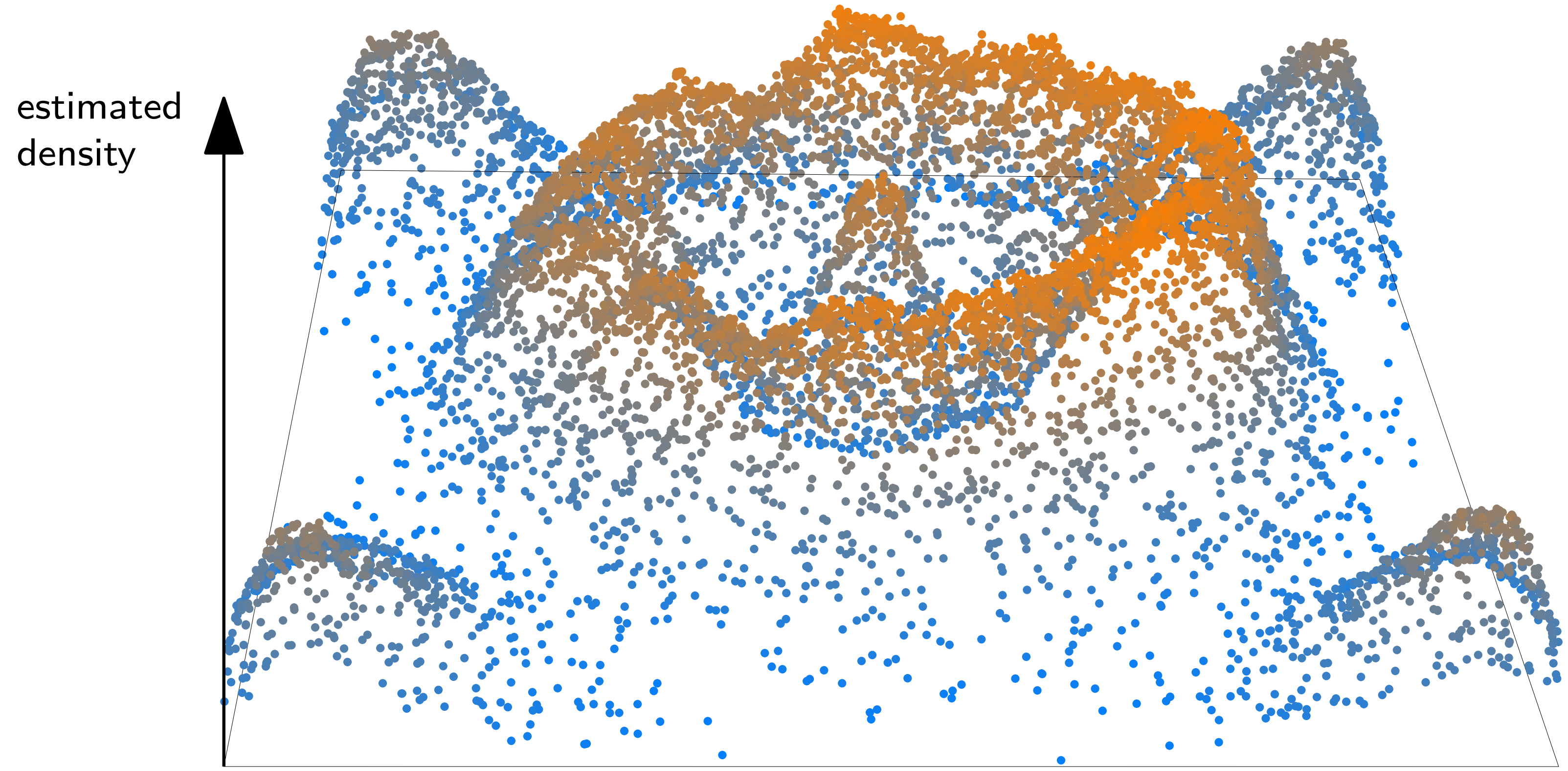


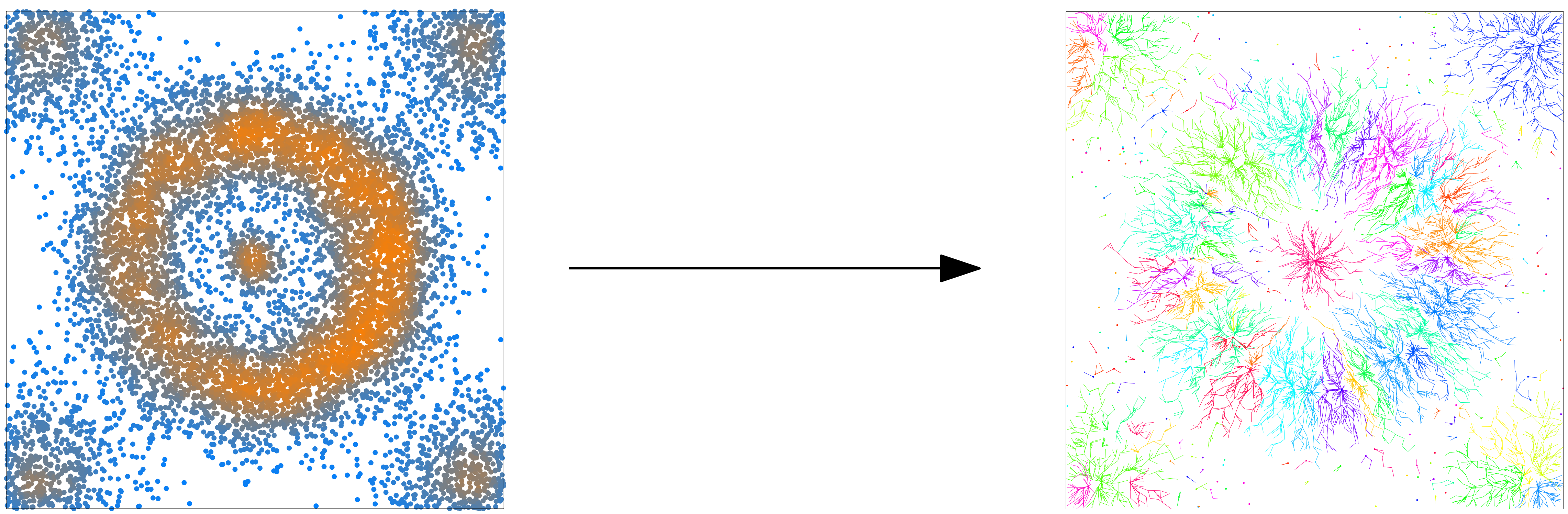
[Koontz, Narendra, Fukunaga'76] in a Nutshell



Why things are likely to go ill

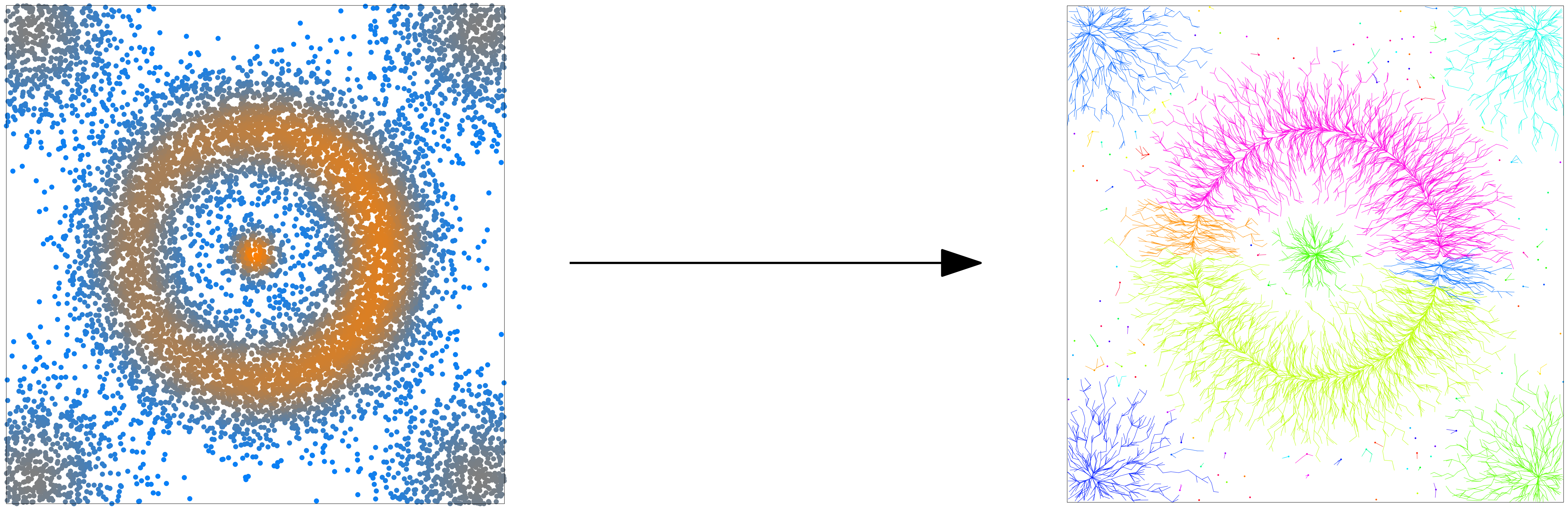
- Noisy estimator





Why things are likely to go ill

- Noisy estimator
- Neighborhood graph



Why things are likely to go ill

- Noisy estimator
- Neighborhood graph

Solutions:

1. **Be proactive**: act on approximate gradient flow (Mean-Shift [CM'02])

    $\rightarrow$ use kernel density estimator, with smoothing window parameter

    $\rightarrow$ work in ambient space to circumvent neighborhood graph issue

Why things are likely to go ill

- Noisy estimator
- Neighborhood graph

Solutions:

1. **Be proactive**: act on approximate gradient flow (Mean-Shift [CM'02])

   $\rightarrow$ use kernel density estimator, with smoothing window parameter

   $\rightarrow$ work in ambient space to circumvent neighborhood graph issue

2. **Be reactive**: merge clusters after clustering, to regain some stability

   $\rightarrow$ repeat mode-seeking until convergence (Medoid-Shift [SKK'07])

   $\rightarrow$ use topological persistence to guide a single-pass merging step

Why things are likely to go ill

- Noisy estimator
- Neighborhood graph

Solutions:

1. **Be proactive**: act on approximate gradient flow (Mean-Shift [CM'02])

    $\rightarrow$ use kernel density estimator, with smoothing window parameter

    $\rightarrow$ work in ambient space to circumvent neighborhood graph issue

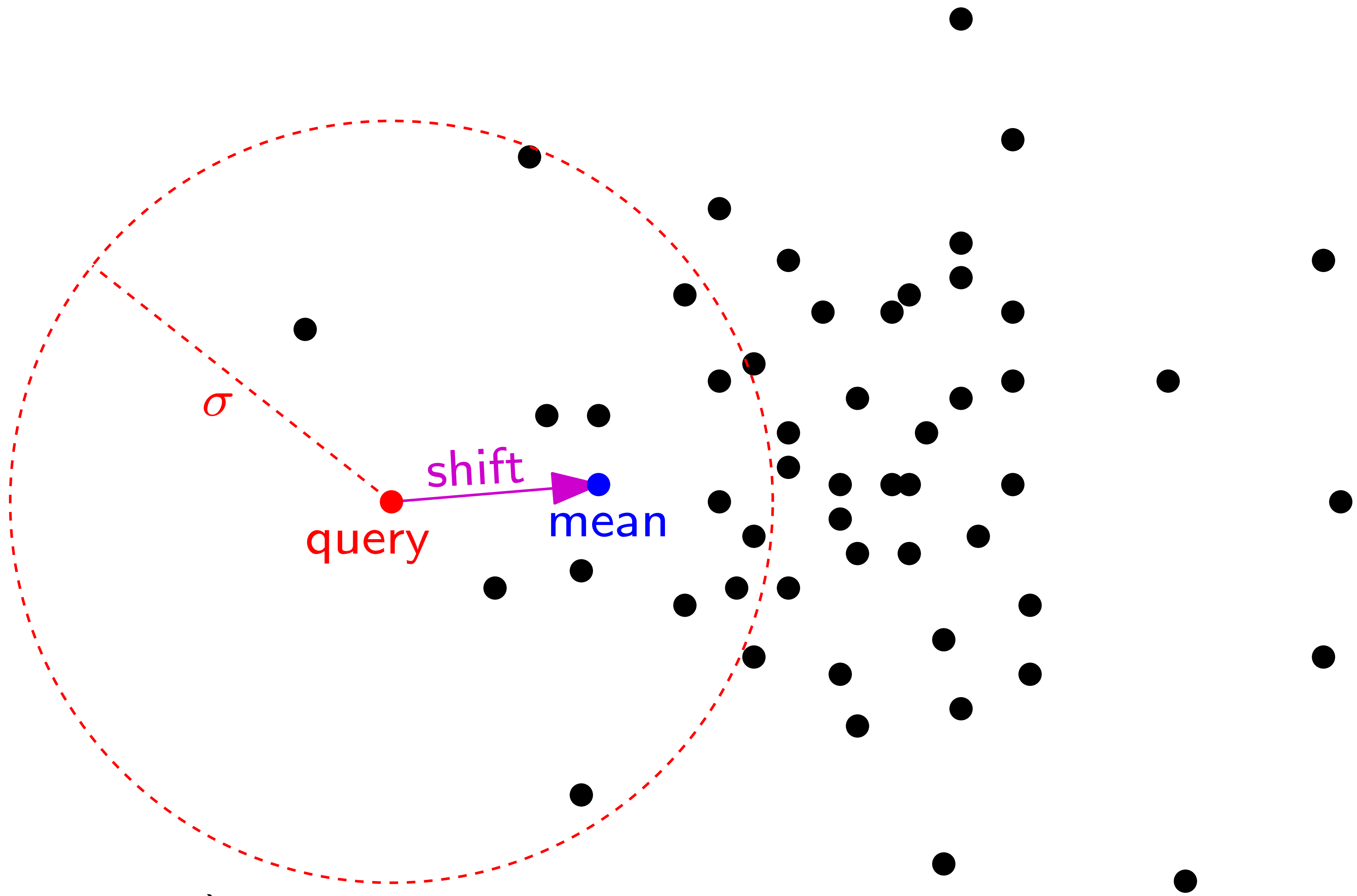
2. **Be reactive**: merge clusters after clustering, to regain some stability

    $\rightarrow$ repeat mode-seeking until convergence (Medoid-Shift [SKK'07])

    $\rightarrow$ use topological persistence to guide a single-pass merging step

Mean-Shift in practice

- Apply Mean-Shift hill-climbing to each input point $p_i \in P$



(Epanechnikov kernel)

Mean-Shift in practice

- Apply Mean-Shift hill-climbing to each input point $p_i \in P$
- Epanechnikov kernel $\Rightarrow$ convergence in finite time

  $\rightarrow$ may converge outside the set of critical points of the estimator

  $\rightarrow$ use variant to guarantee cvgence to maximum [Huang et al. 2017]

Mean-Shift in practice

- Apply Mean-Shift hill-climbing to each input point $p_i \in P$
- Epanechnikov kernel $\Rightarrow$ convergence in finite time

  $\rightarrow$ may converge outside the set of critical points of the estimator

  $\rightarrow$ use variant to guarantee cvgence to maximum [Huang et al. 2017]

- Gaussian kernel $\Rightarrow$ convergence at the limit (infinite time)

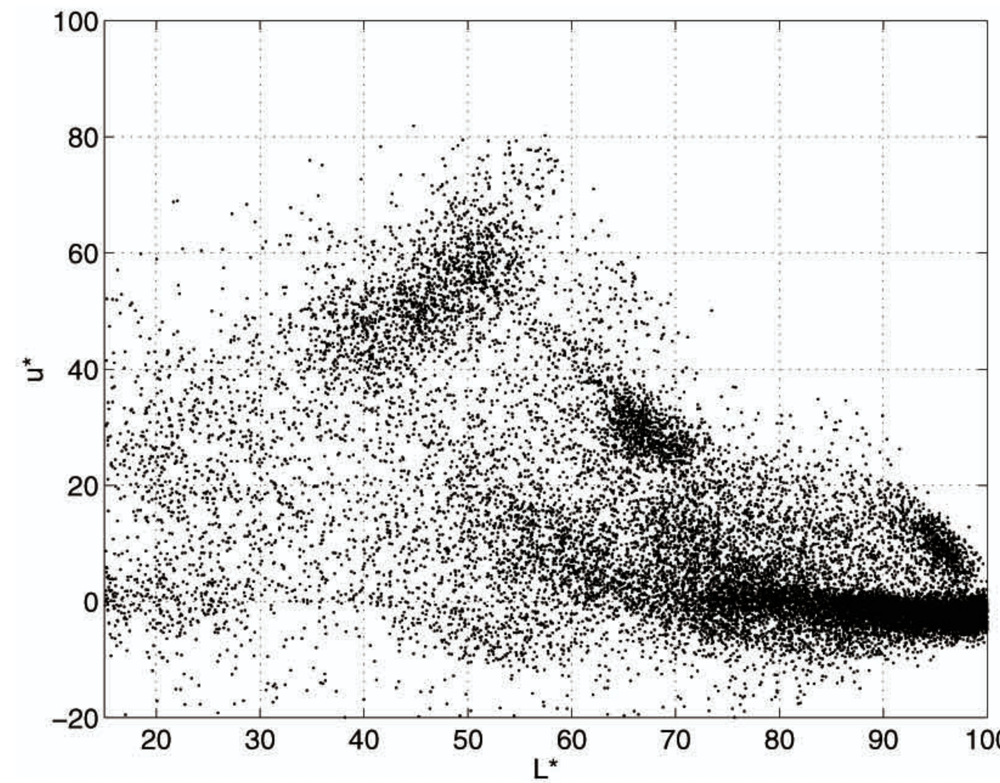
  $\rightarrow$ stopping criterion (convergence radius)

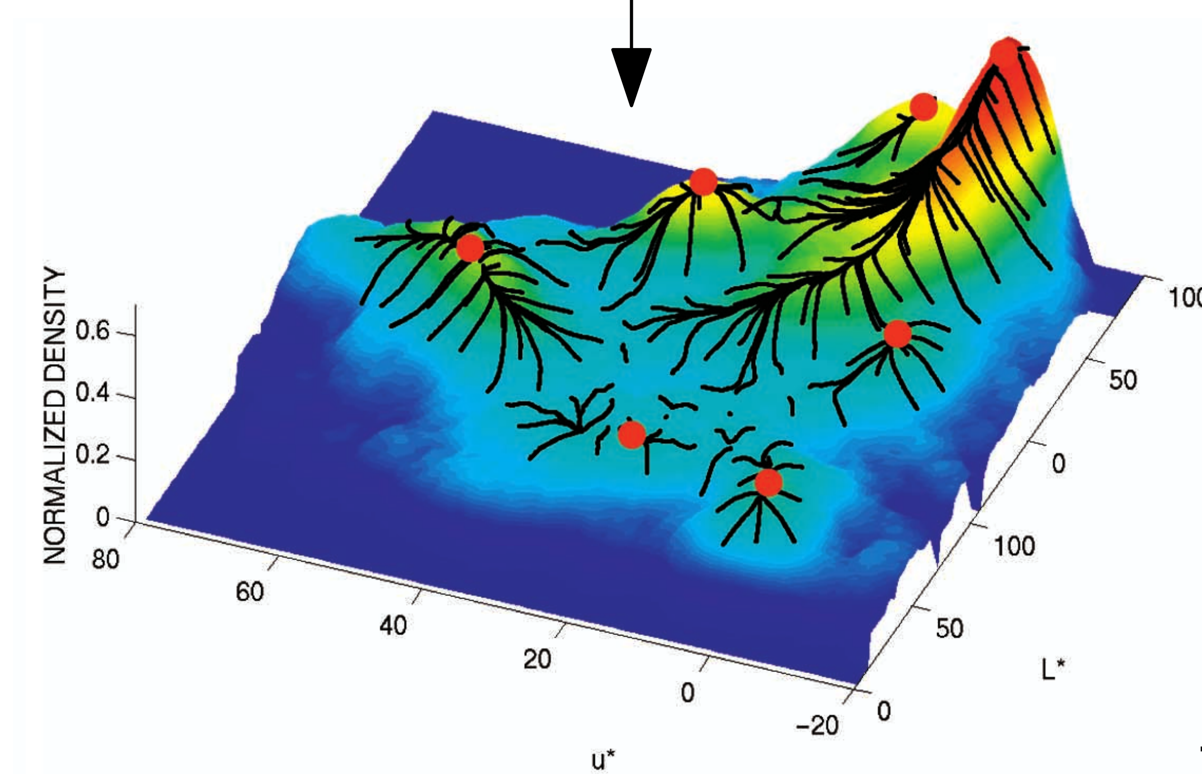
  $\rightarrow$ identification of modes (mode radius)

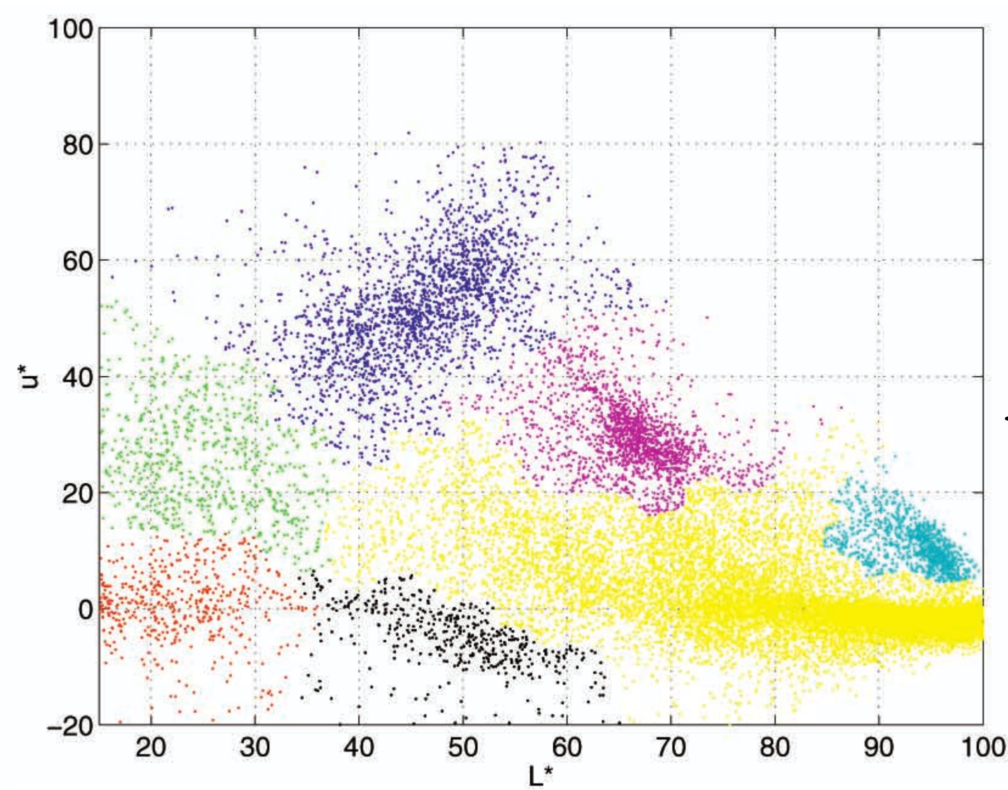
  $\rightarrow$ speed-up: hill-climbing gathers neighboring points (gathering radius)

$\rightsquigarrow$ heuristic: make these radii proportional to the estimator's bandwidth σ

Examples [Comaniciu, Meer 2002]







Examples [Comaniciu, Meer 2002]

Original

$(h_s, h_r) = (8, 8)$

$(h_s, h_r) = (8, 16)$

$(h_s, h_r) = (16, 4)$

$(h_s, h_r) = (16, 8)$

$(h_s, h_r) = (16, 16)$

$(h_s, h_r) = (32, 4)$

$(h_s, h_r) = (32, 8)$

$(h_s, h_r) = (32, 16)$