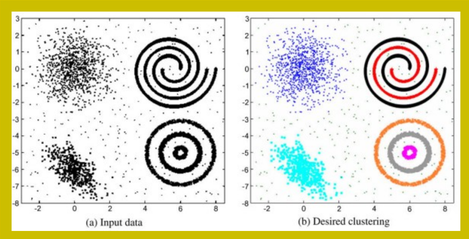
What's After K-Means?

TEXAS STATE UNIVERSITY
The rising STAR *of Texas*

50 years beyond K-means

- Authored By: Anil K. Jain
- Published in Pattern Recognition Letters 31, 2009
- Emphasizes the importance of grouping data.
- Two types of algorithms:
  - Partitional: builds clusters simultaneously
  - Hierarchical: joins or divides clusters.



Global Kernel K-Means

- Presented by: Tzortzis and Likas
- IEEE Joint Conference on Neural Networks in 2008
- Kernel K-Means finds non linear separable clusters.
- Global K-Means minimizes sensitivity to initialization.
- Comes in "fast" and "near optimal" variety.

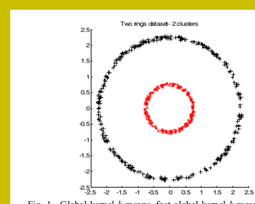


Fig. 1. Global kernel k-means, fast global kernel k-means and kernel k-means (run with minimum clustering error) on the two rings dataset.

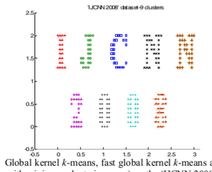


Fig. 2. Global kernel k-means, fast global kernel k-means and kernel k-means (run with minimum clustering error) on the 'IJCNN 2008' logo.

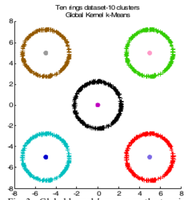


Fig. 3. Global kernel k-means on the ten rings dataset.

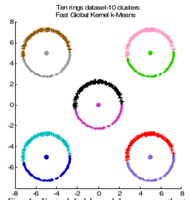


Fig. 4. Fast global kernel k-means on the ten rings dataset.

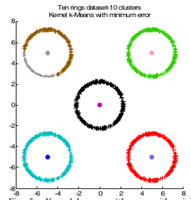


Fig. 5. Kernel k-means (the run with minimum clustering error) on the ten rings dataset.

BIRCH

- Presented by: Zhang, Ramakrishnan, and Livny
- SIGMOD 1996
- Aimed to handle big data within limited memory.
- Compares data points locally rather than globally.
- Can produce good results on a single scan of data.
- Captures hierarchy of the data as a CF Tree.

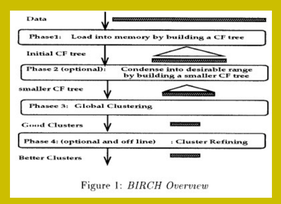


Figure 1: *BIRCH Overview*

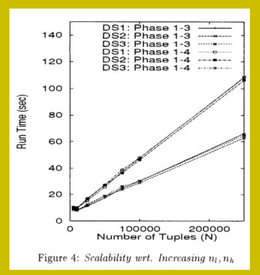


Figure 4: *Scalability wrt. Increasing* n_i, n_k

A Density Based Algorithm

- Presented by: Ester, Kriegel, Sander and Xu.
- Association for the Advancement of AI in 1996
- Only takes one input parameter.
- Clusters by density of points rather than centroids.
- Can define clusters of arbitrary shapes.
- This is a partitioning algorithm.

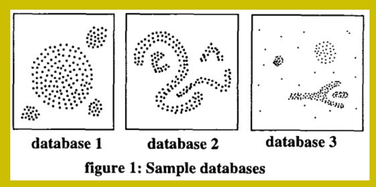


figure 1: Sample databases

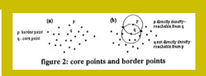


figure 2: core points and border points

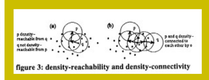
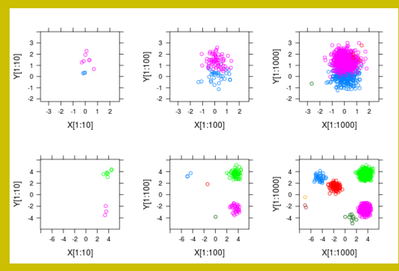


figure 3: density-reachability and density-connectivity

Dynamic Clustering

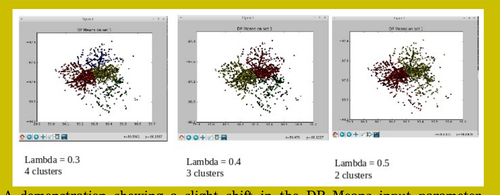
- Presented by: Campbell, Liu, Kulis, How, and Carin.
- ANIPS, 2013
- Based on Dirichlet Processes
- Non-parametric: model grows with data.
- Clusters can be created, eliminated, and altered.
- Promises high speed for time-sensitive applications.



https://en.wikipedia.org/wiki/Dirichlet_process#/media/File:DP_clustering_simulation.png

So What Comes Next?

- Focus on Billion-Scale data.
- Solve the "Curse of Dimensionality" rather than avoiding.
- Distance measures break down at high dimension
- Centroids make assumptions about structure of data.
- Parallelism is a must.
- Fuzzy clustering should be a possibility.
- **Good algorithms do not make assumptions about data.**



Lambda = 0.3
4 clusters

Lambda = 0.4
3 clusters

Lambda = 0.5
2 clusters

A demonstration showing a slight shift in the DP Means input parameter finding 2, 3, or 4 clusters on Austin traffic accidents.

Presenter: Gentry Atkinson

Faculty Mentors: Dr. Tesic and Dr. Tamir