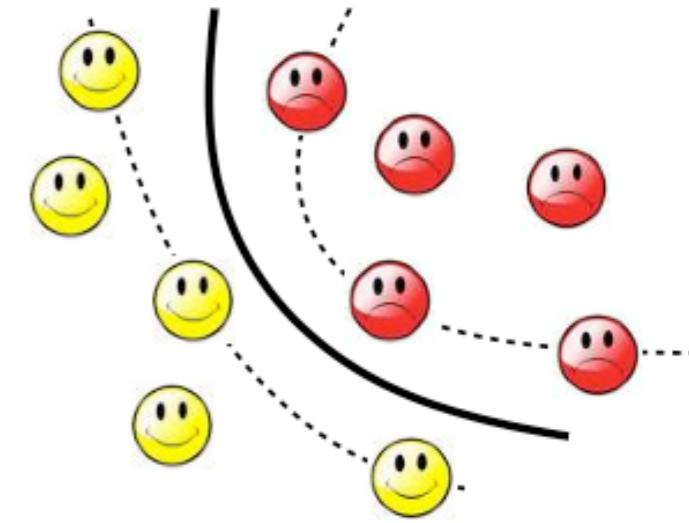




THE UNIVERSITY OF
SYDNEY



Machine Learning and Data Mining (COMP 5318)

Clustering and Expectation-Maximisation

Fabio Ramos
Roman Marchant



THE UNIVERSITY OF
SYDNEY

Clustering

C. Bishop, *Pattern Recognition and Machine Learning*,

Chapter 9: Mixture Models and EM

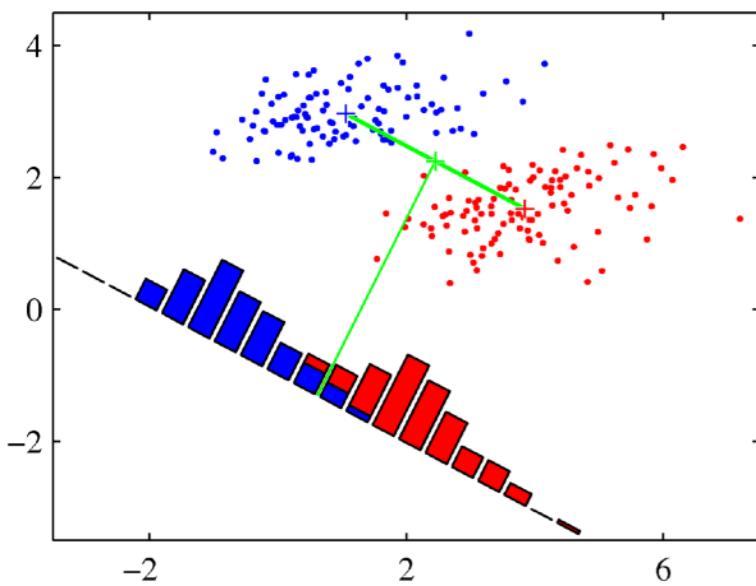
Springer New York, 2006

K.P. Murphy, *Machine Learning: a Probabilistic Perspective*,

Chapters 11 and 25, Massachusetts Institute of Technology, 2006

Types of Learning

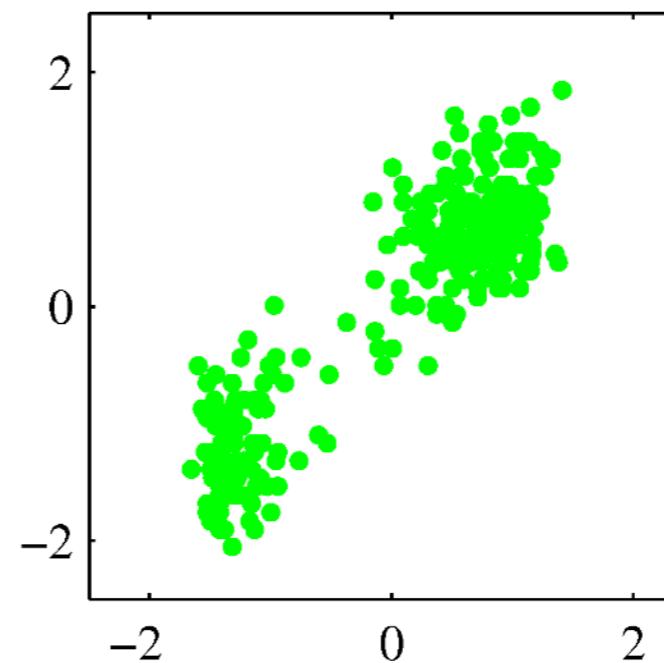
Supervised Learning



Learning input-output from examples.

Regression Classification

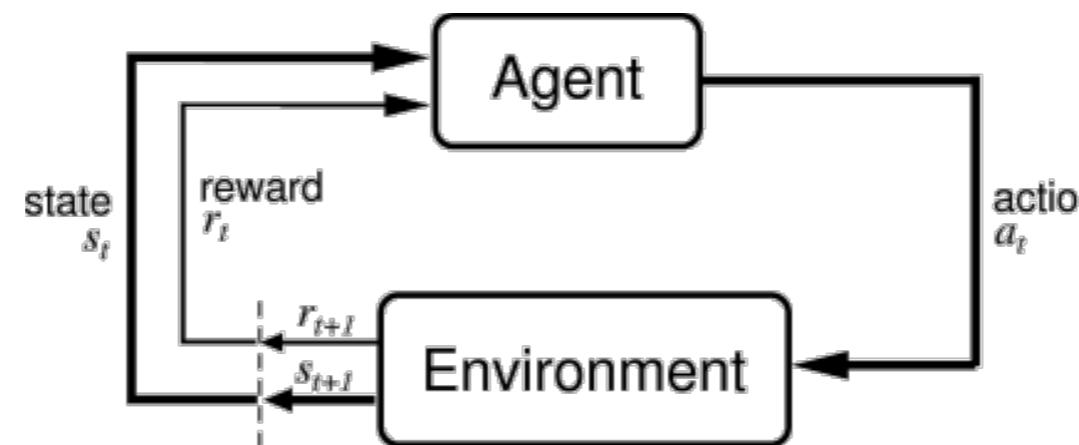
Unsupervised Learning



Learning underlying structure.

Clustering Density Estimation

Reinforcement Learning



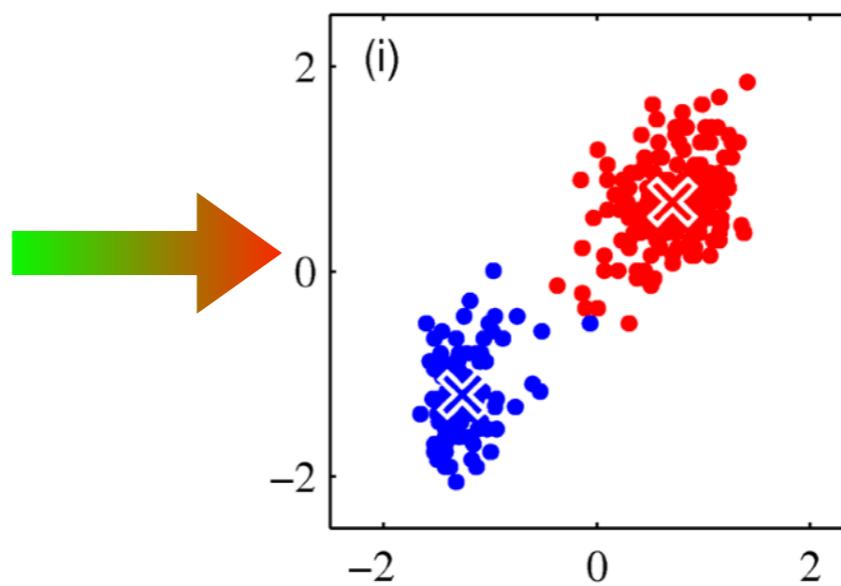
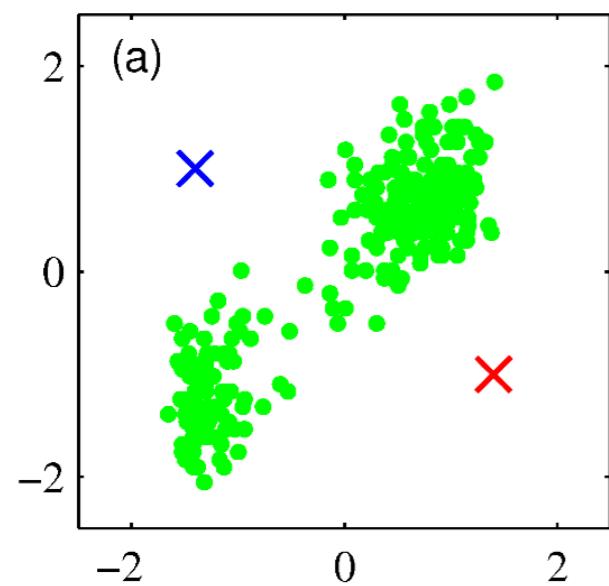
Learning policy from state-action-reward sequences.

Learning behaviour.



Clustering

Process of grouping similar objects together.



- ✖ Cluster 1 centre.
- ✖ Cluster 2 centre.
- Unlabelled data.
- Data belonging to Cluster 1.
- Data belonging to Cluster 2.

Learn a set of clusters and assign data to a specific cluster.

Deterministic: Hard assignment to each cluster (K-means).

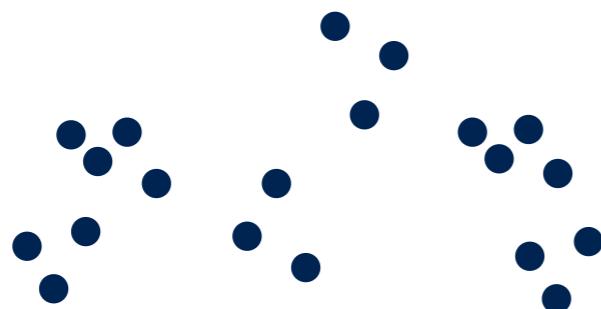
Probabilistic: Model assignment as a discrete latent variable.
(Mixtures of Gaussians, Dirichlet Process)



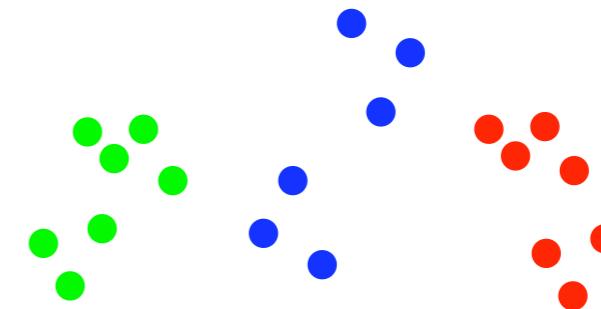
THE UNIVERSITY OF
SYDNEY

Clustering

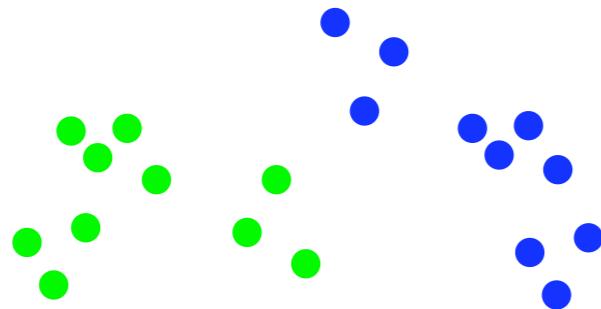
How many clusters?



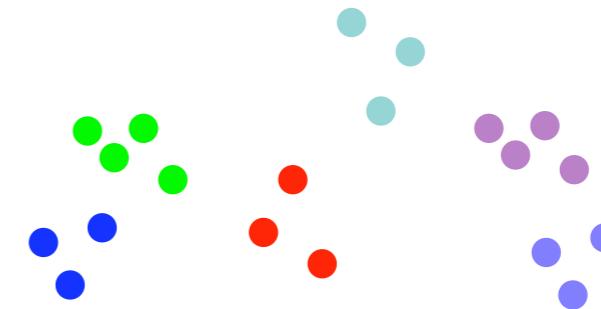
One



Three



Two



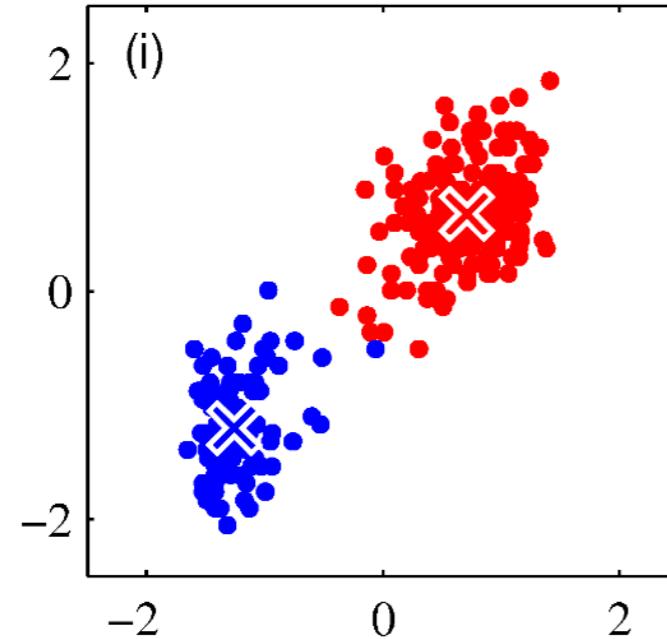
Six

Presence of ambiguous solutions.



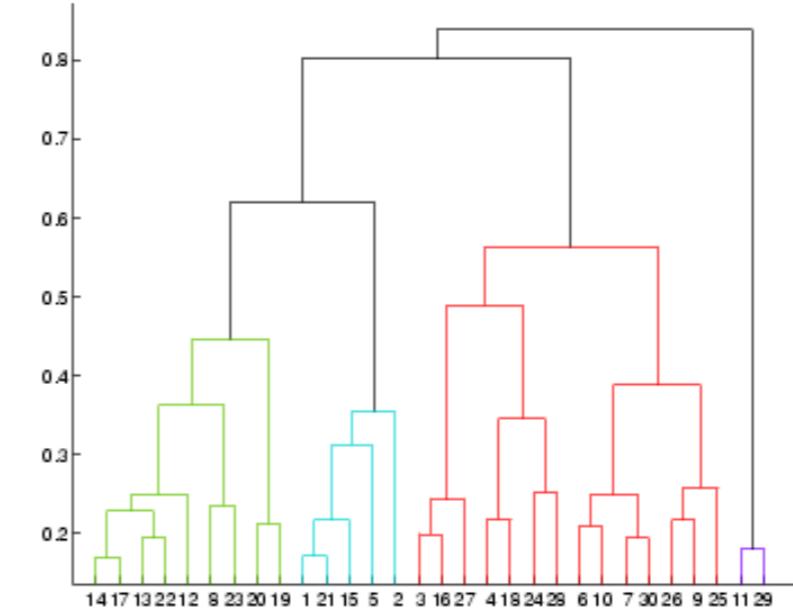
Types of Clustering

Partition Clustering



Partition de objects into disjoint sets.
Faster to create.
Sensible to initial conditions.
Model selection for K.

Hierarchical Clustering



Nested tree of partitions.
Slower to create.
Often more useful.
Do not require knowing the number of clusters.



Clustering

Dataset $\mathcal{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ with N Observations

Each datapoint is D dimensional: $\mathbf{x}_i = (x_1, \dots, x_D)$

Goal: Partition dataset into K clusters. (For now, assume K is given)

$\mu_k = (\mu_1, \dots, \mu_D)$ Prototype for each cluster. $k \in 1, \dots, K$

Binary indicator variables

$$r_{nk} = \begin{cases} 1, & \text{if datapoint } n \text{ belongs to cluster } k \\ 0, & \sim \end{cases}$$

If \mathbf{x}_n is assigned to cluster k, then $r_{nk} = 1 \wedge r_{nj} = 0 \forall j \neq k$



K-Means

Objective function:

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2$$

Represents the sum of the squares of the distances of each datapoint to its assigned prototype vector.

Goal: Find $\{r_{nk}\}$ and $\{\boldsymbol{\mu}_k\}$ that minimise J.

$$\{r_{nk}, \boldsymbol{\mu}_k\}^* = \operatorname{argmin}_{\{r_{nk}, \boldsymbol{\mu}_k\}} \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2$$



K-Means

Iterative solution to minimise J:

- 1 Data Preprocessing
- 2 Initialise $\{\mu_k\}$
- 3 Repeat until convergence or Max Iterations
 - 4 Minimise J w.r.t. $\{r_{nk}\}$ keeping $\{\mu_k\}$ fixed.
 - 5 Minimise J w.r.t. $\{\mu_k\}$ keeping $\{r_{nk}\}$ fixed.



THE UNIVERSITY OF
SYDNEY

K-Means

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2$$

Line 4 : Optimise w.r.t. r_{nk}

Each data point is independent, so we can optimise for each n separately:

$$r_{nk} = \begin{cases} 1 & \text{if } k = \operatorname{argmin}_j \|\mathbf{x}_n - \boldsymbol{\mu}_j\|^2 \\ 0 & \sim \end{cases}$$

Assign each data point to its closest centre.



K-Means

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2$$

Line 5 : Optimise w.r.t. $\boldsymbol{\mu}_k$

$$\frac{\partial J}{\partial \boldsymbol{\mu}_k} = 0$$

$$2 \sum_{n=1}^N r_{nk} (\mathbf{x}_n - \boldsymbol{\mu}_k) = 0$$

$$\sum_{n=1}^N r_{nk} \mathbf{x}_n = \sum_{n=1}^N r_{nk} \boldsymbol{\mu}_k$$

$$\frac{\sum_n r_{nk} \mathbf{x}_n}{\sum_n r_{nk}} = \boldsymbol{\mu}_k$$

Set $\boldsymbol{\mu}_k$ equal to the mean of all data points \mathbf{x}_n assigned to cluster k .

K-Means

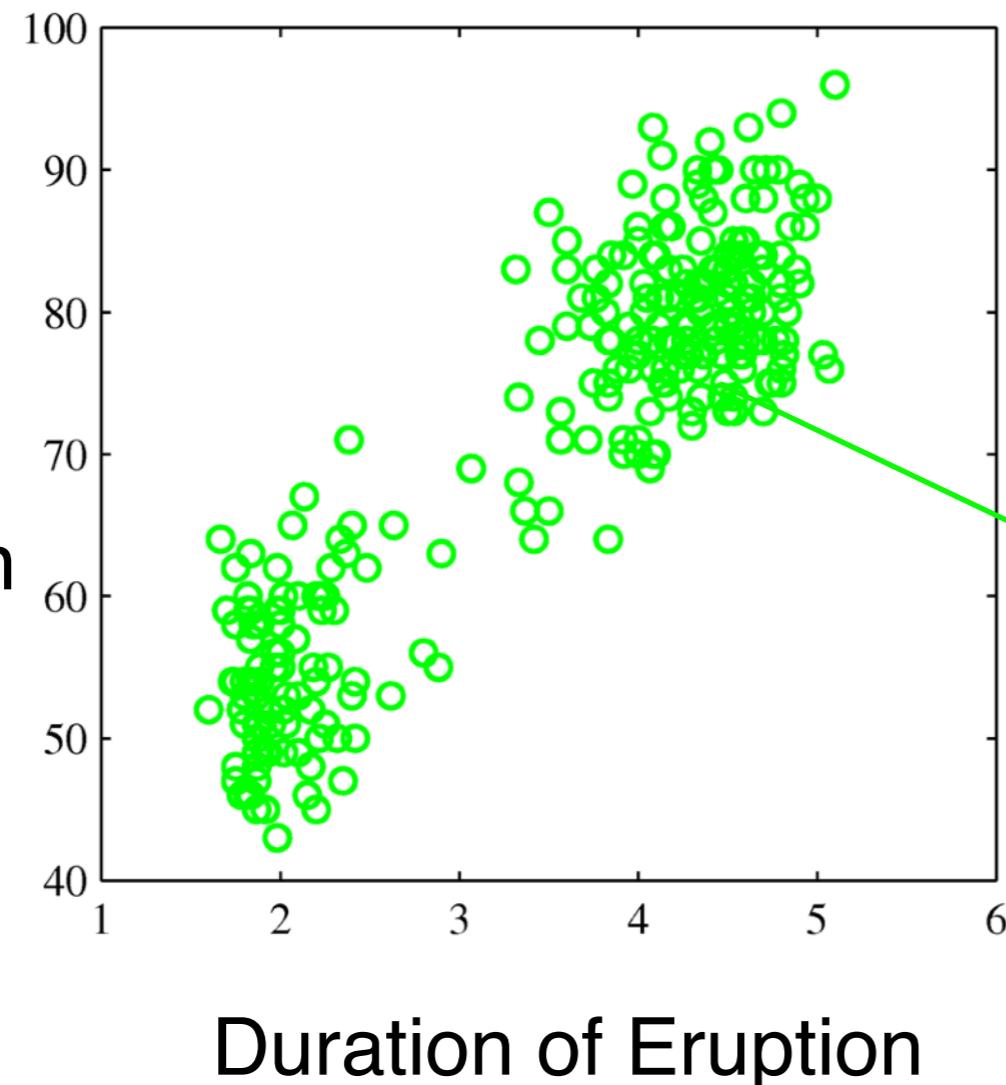


THE UNIVERSITY OF
SYDNEY

K-Means Example

Hydrothermal Geyser: Old Faithful

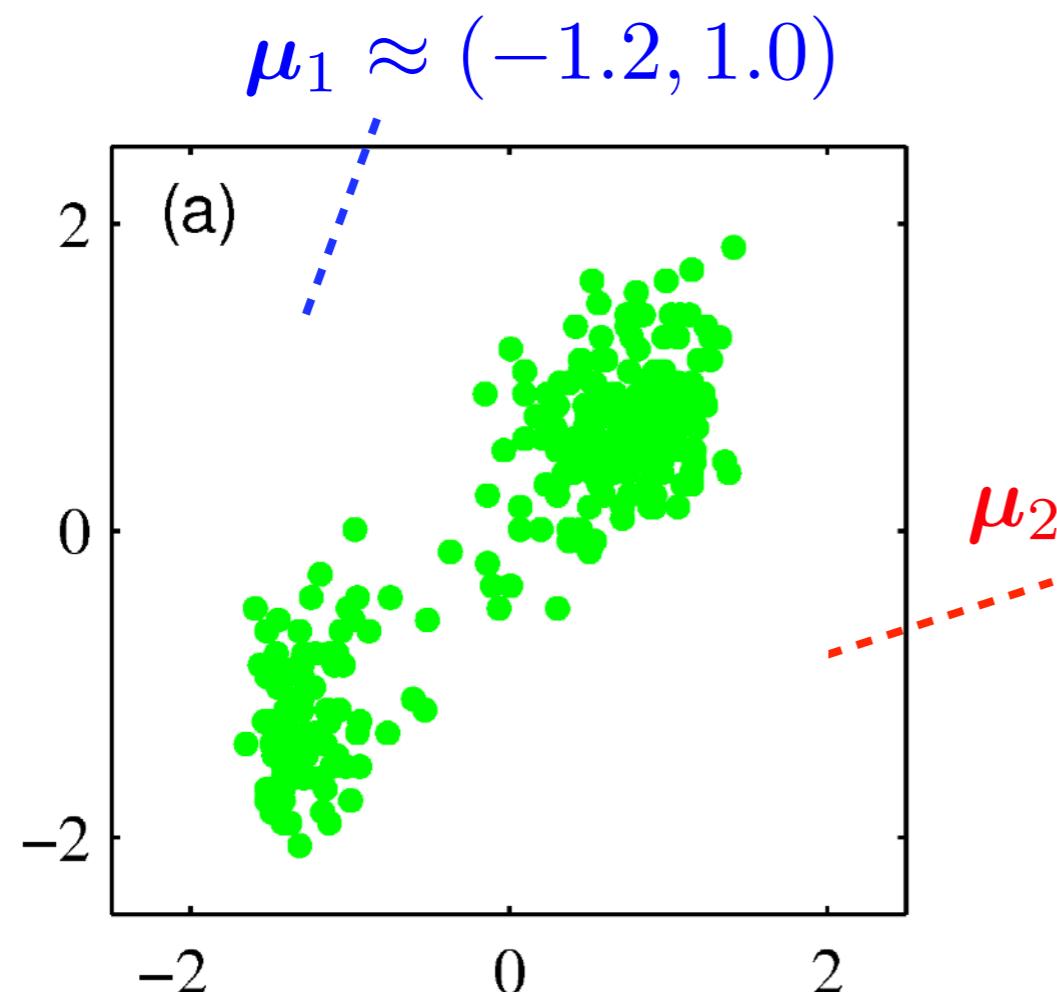
Time to
next
Eruption



Single Eruption
 $N = 272$ Observations

K-Means Example

Number of clusters: $K = 2$



- 1 Data Preprocessing
- 2 Initialise $\{\mu_k\}$
- 3 Repeat until convergence or Max Iterations
- 4 Minimise J w.r.t. $\{r_{nk}\}$ keeping $\{\mu_k\}$ fixed.
- 5 Minimise J w.r.t. $\{\mu_k\}$ keeping $\{r_{nk}\}$ fixed.

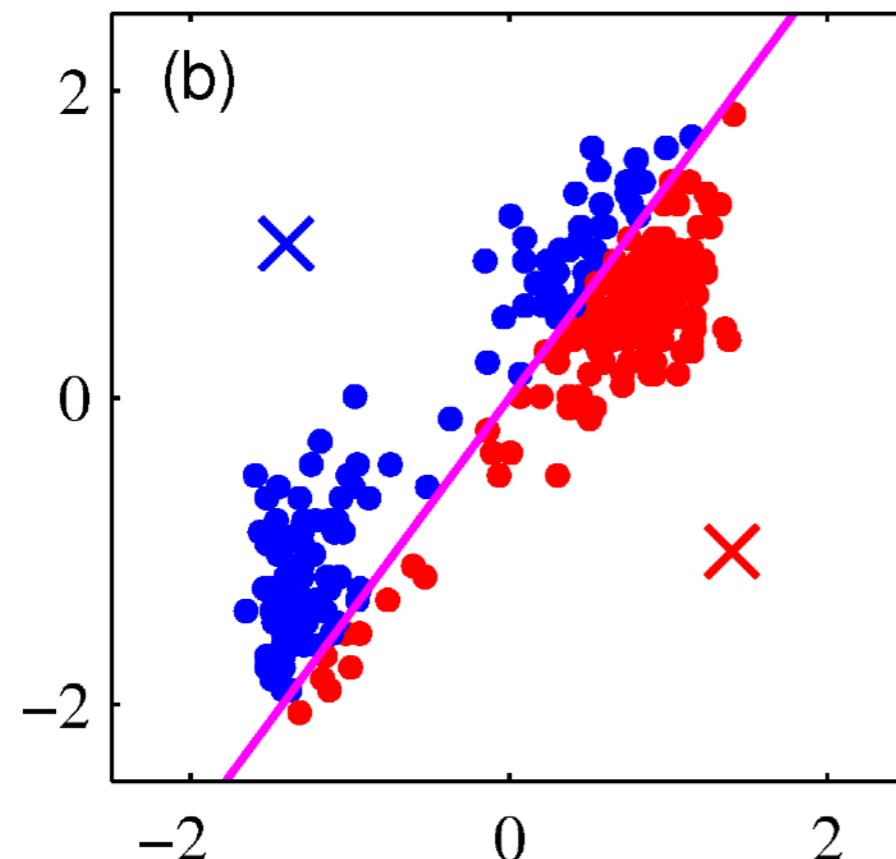
Standardise data:
Each dimension has zero mean and unit standard deviation.

Better initialisation:
Choose $\{\mu_k\}$ as average of a random subset.



K-Means Example

Number of clusters: $K = 2$



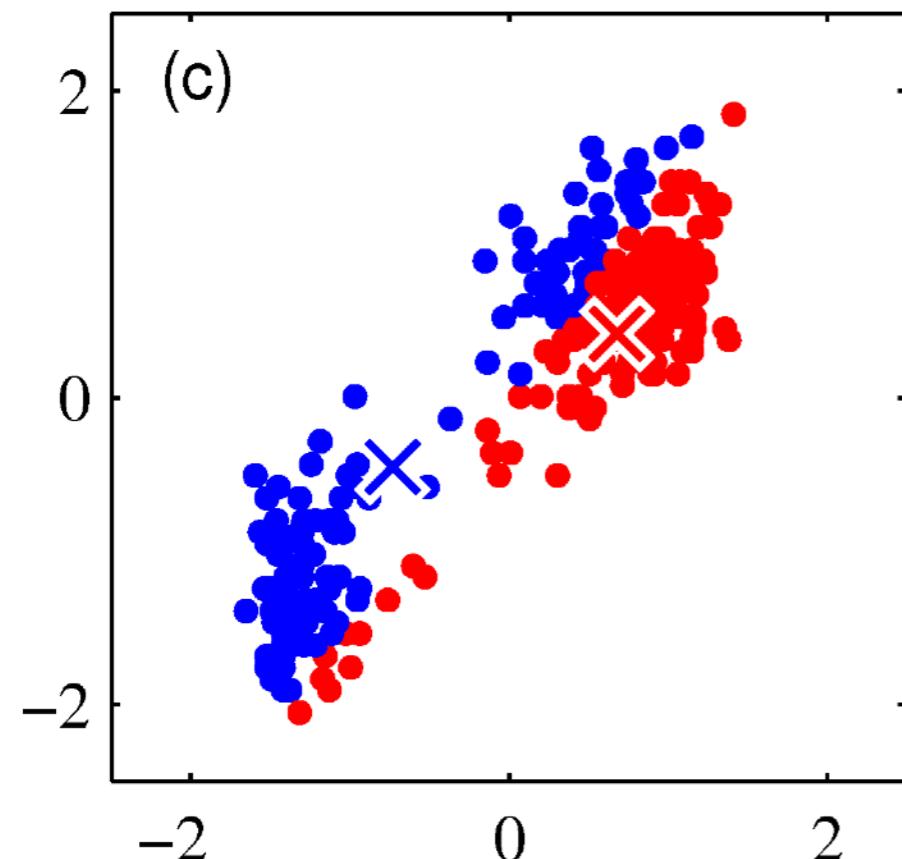
- 1 Data Preprocessing
- 2 Initialise $\{\mu_k\}$
- 3 Repeat until convergence or Max Iterations
- 4 Minimise J w.r.t. $\{r_{nk}\}$ keeping $\{\mu_k\}$ fixed.
- 5 Minimise J w.r.t. $\{\mu_k\}$ keeping $\{r_{nk}\}$ fixed.

Each data point is assigned to the closest cluster centre.



K-Means Example

Number of clusters: $K = 2$



- 1 Data Preprocessing
- 2 Initialise $\{\mu_k\}$
- 3 Repeat until convergence or Max Iterations
- 4 Minimise J w.r.t. $\{r_{nk}\}$ keeping $\{\mu_k\}$ fixed.
- 5 Minimise J w.r.t. $\{\mu_k\}$ keeping $\{r_{nk}\}$ fixed.

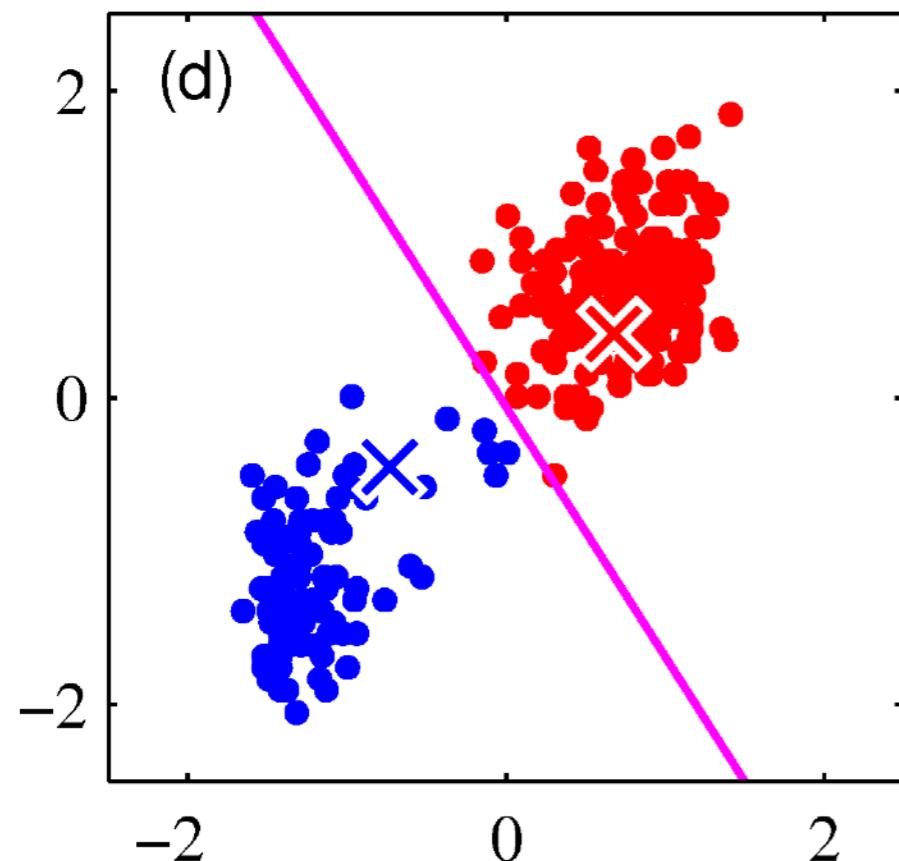
Re-compute each cluster centre to be the mean of the points previously assigned.



THE UNIVERSITY OF
SYDNEY

K-Means Example

Number of clusters: $K = 2$



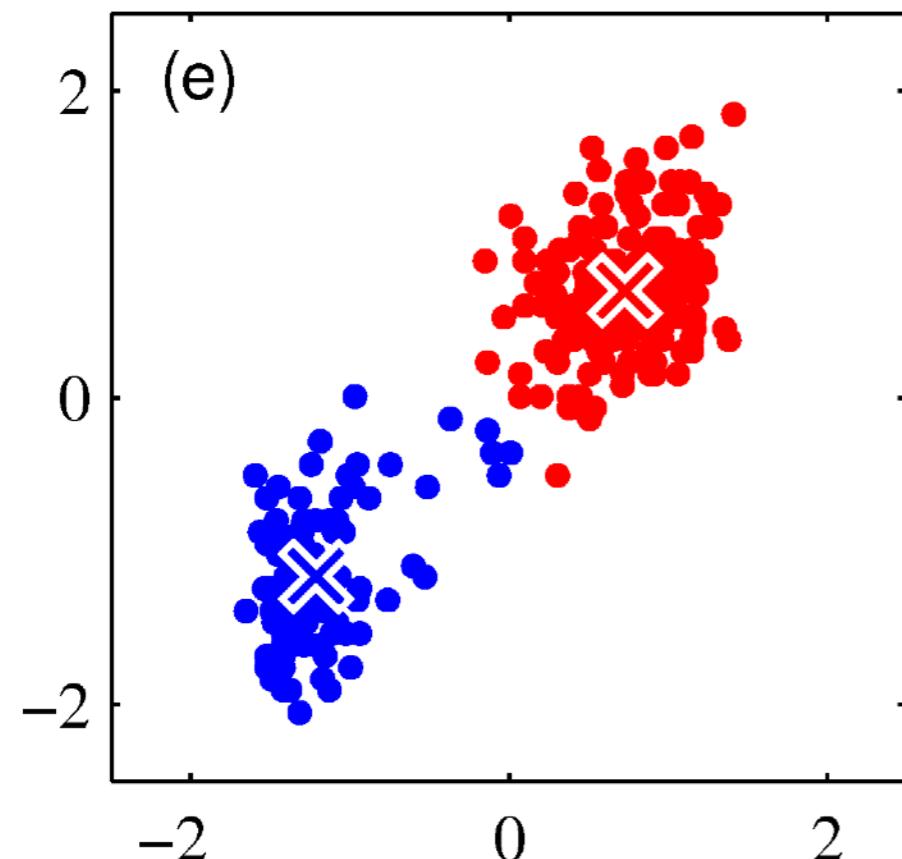
- 1 Data Preprocessing
- 2 Initialise $\{\mu_k\}$
- 3 Repeat until convergence or Max Iterations
- 4 Minimise J w.r.t. $\{r_{nk}\}$ keeping $\{\mu_k\}$ fixed.
- 5 Minimise J w.r.t. $\{\mu_k\}$ keeping $\{r_{nk}\}$ fixed.

Each data point is assigned to the closest cluster centre.



K-Means Example

Number of clusters: $K = 2$



- 1 Data Preprocessing
- 2 Initialise $\{\mu_k\}$
- 3 Repeat until convergence or Max Iterations
- 4 Minimise J w.r.t. $\{r_{nk}\}$ keeping $\{\mu_k\}$ fixed.
- 5 Minimise J w.r.t. $\{\mu_k\}$ keeping $\{r_{nk}\}$ fixed.

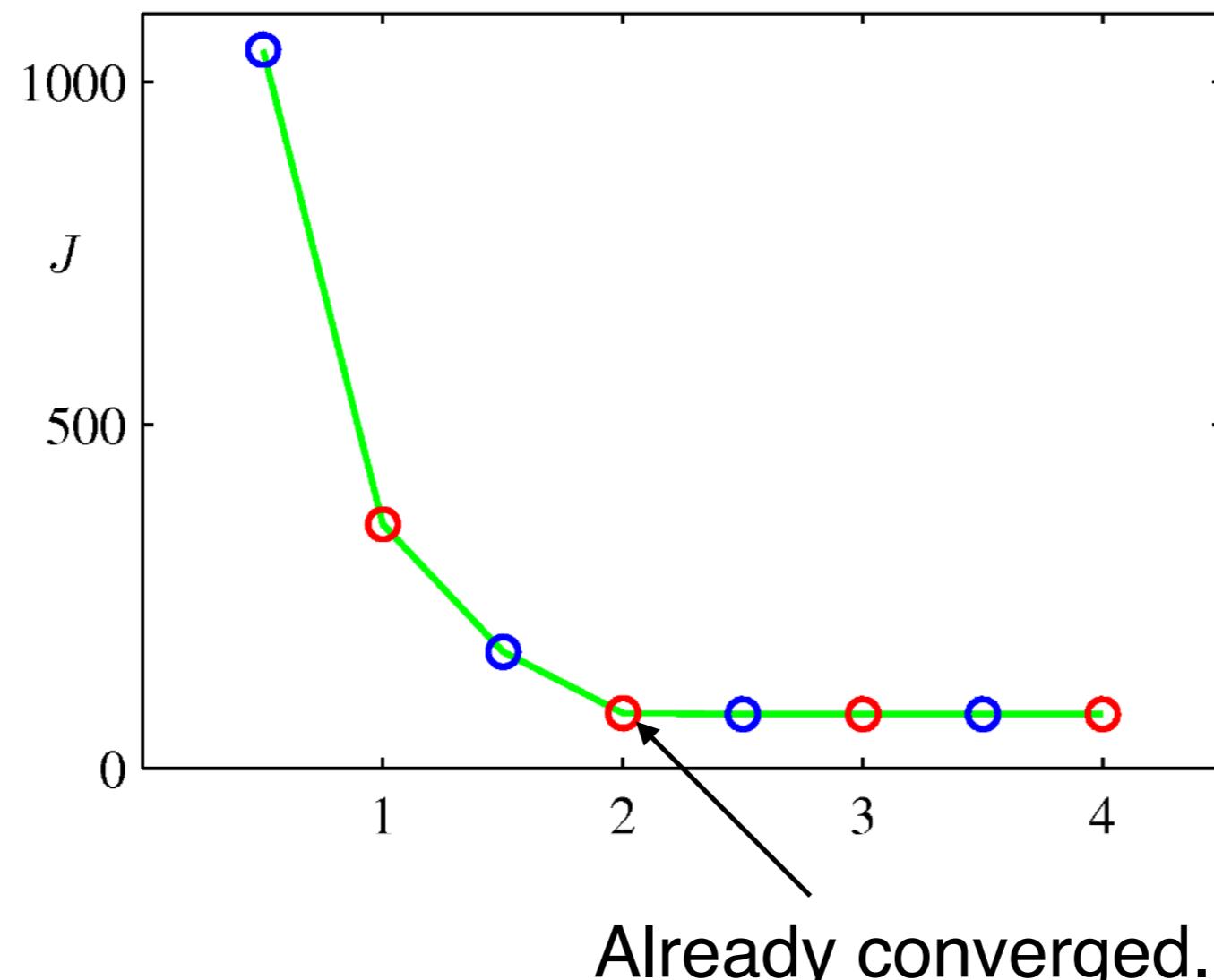
Re-compute each cluster centre to be the mean of the points previously assigned.



THE UNIVERSITY OF
SYDNEY

K-Means Example

Plot of the cost function for each iteration.





THE UNIVERSITY OF
SYDNEY

K-Means Example 2

Image segmentation and compression.

$K = 2$



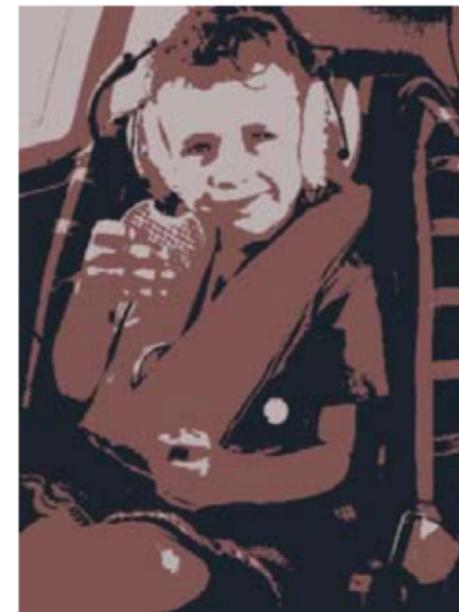
$K = 3$



$K = 10$



Original image

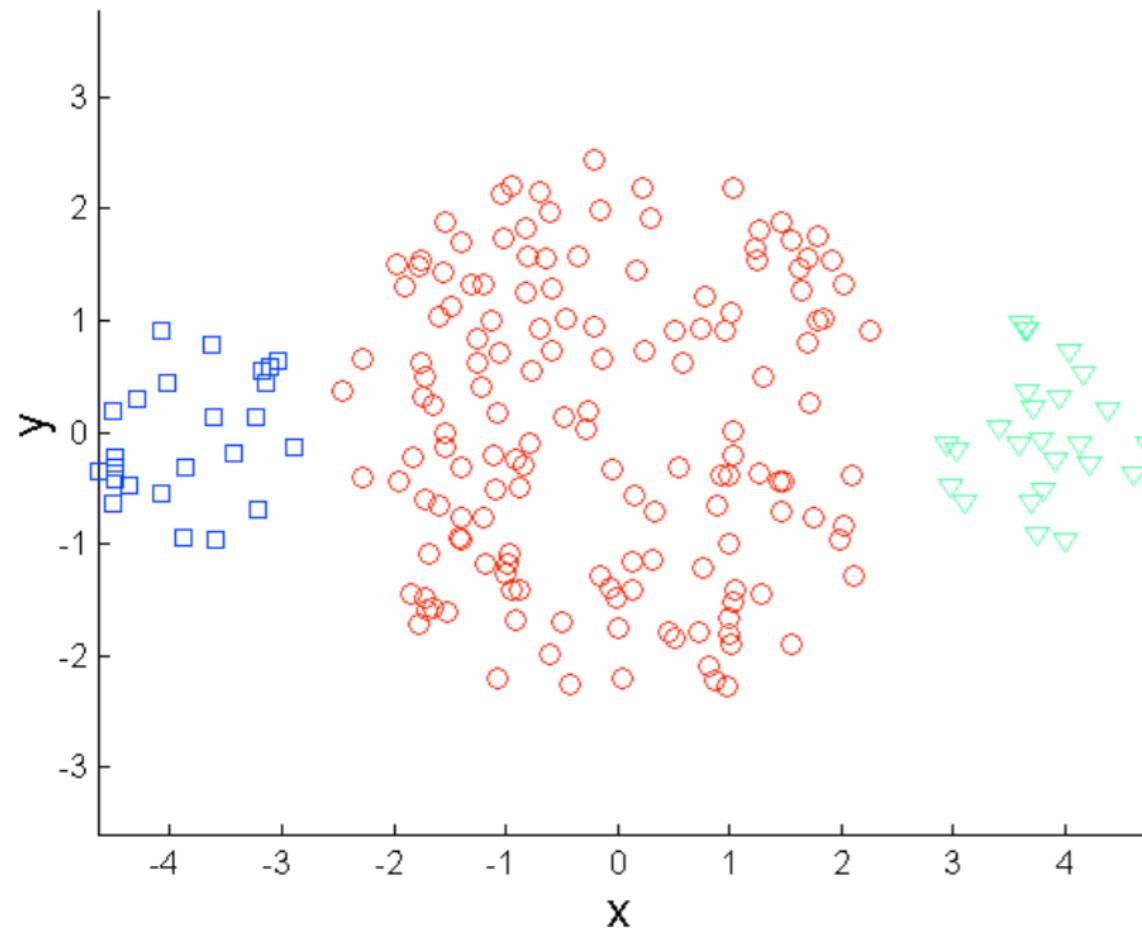




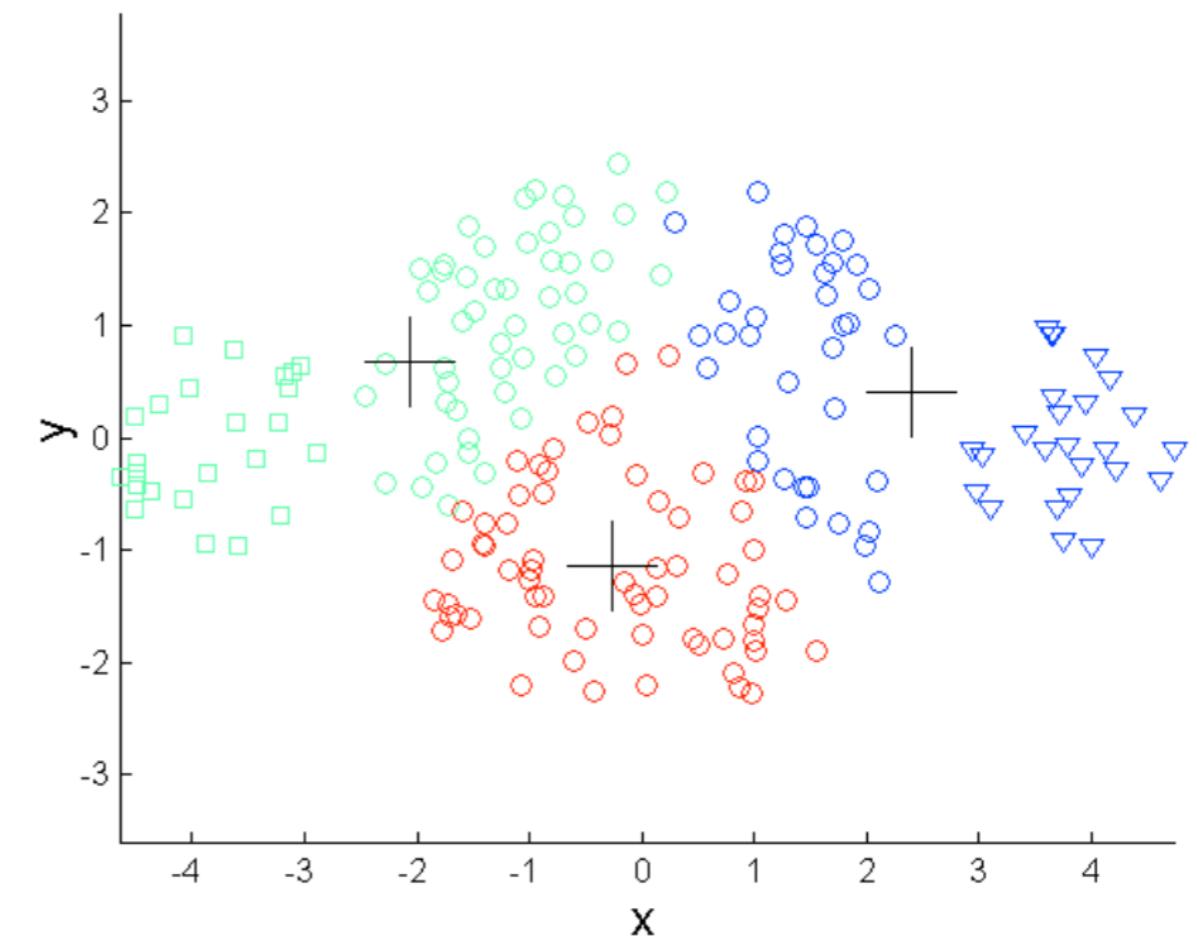
THE UNIVERSITY OF
SYDNEY

K-Means Limitations

Differing Sizes:



Original Points



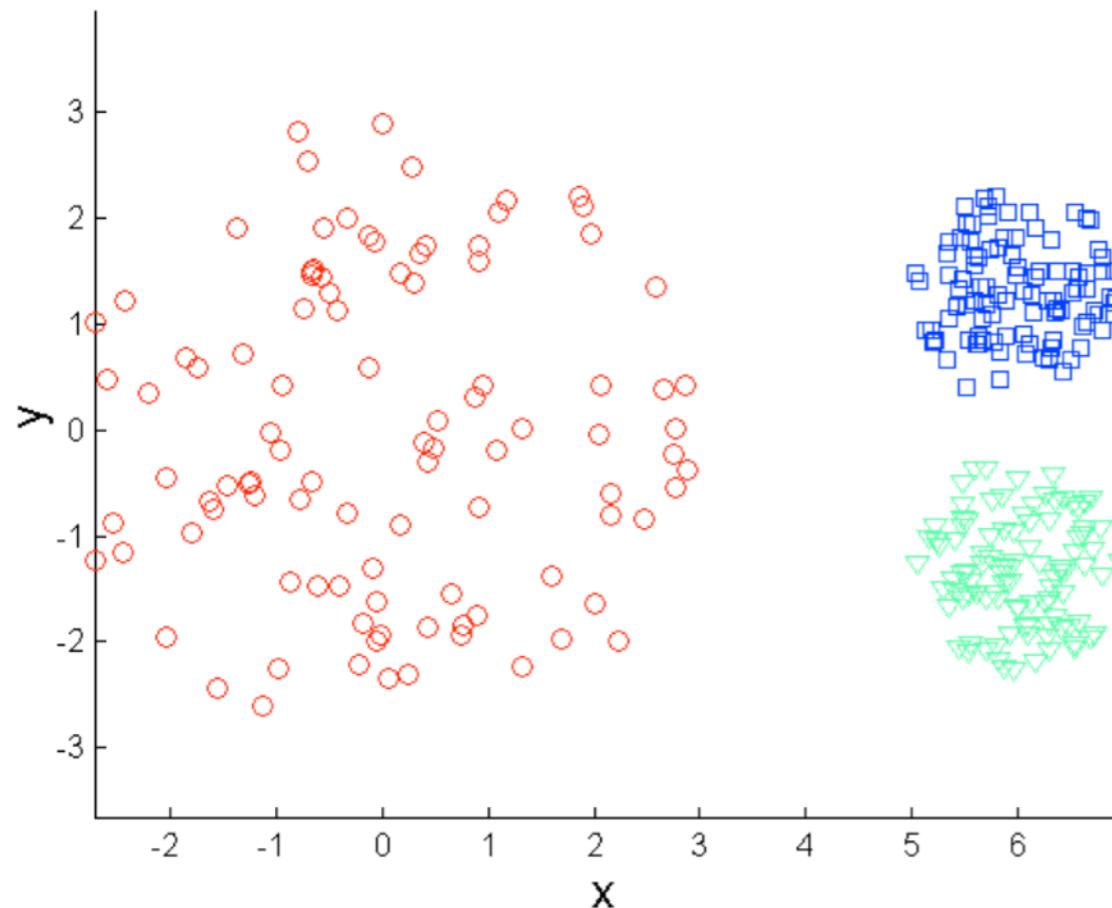
K-Means 3
Clusters



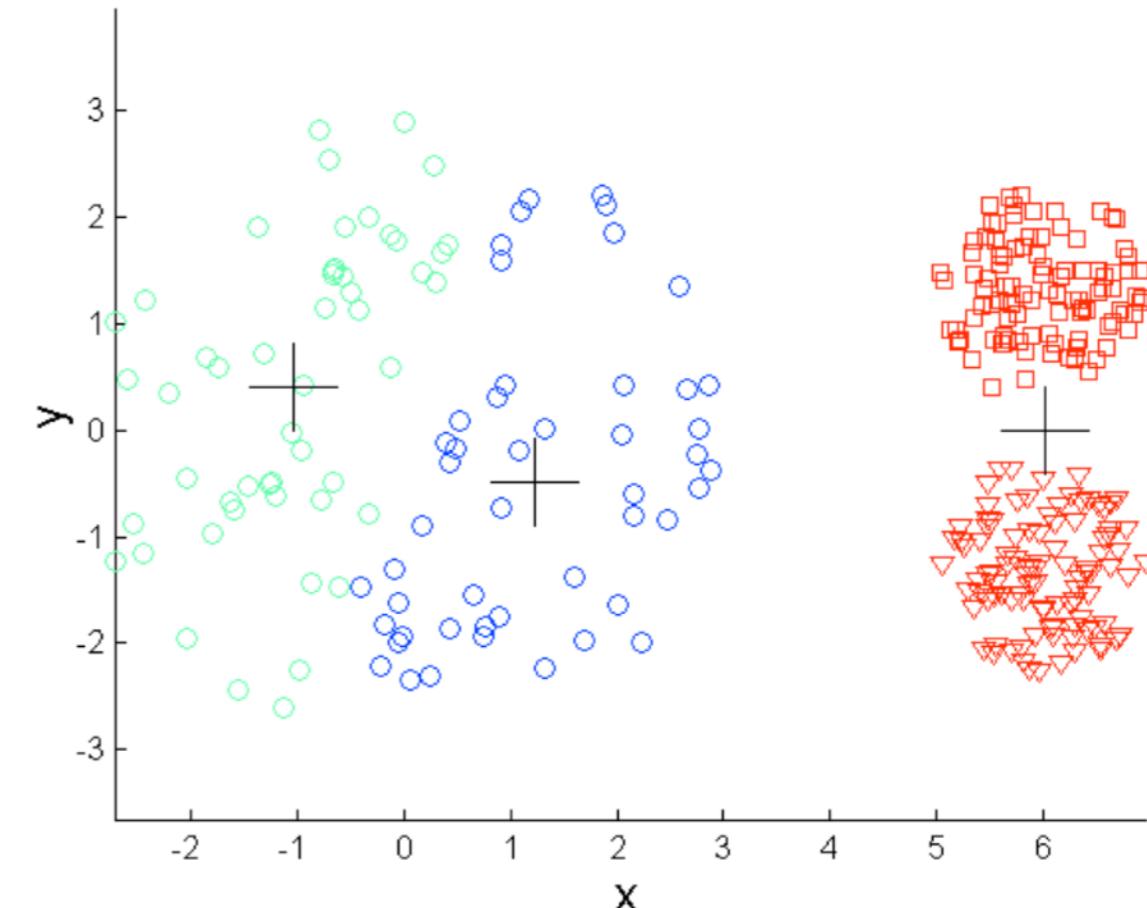
THE UNIVERSITY OF
SYDNEY

K-Means Limitations

Differing Density:



Original Points



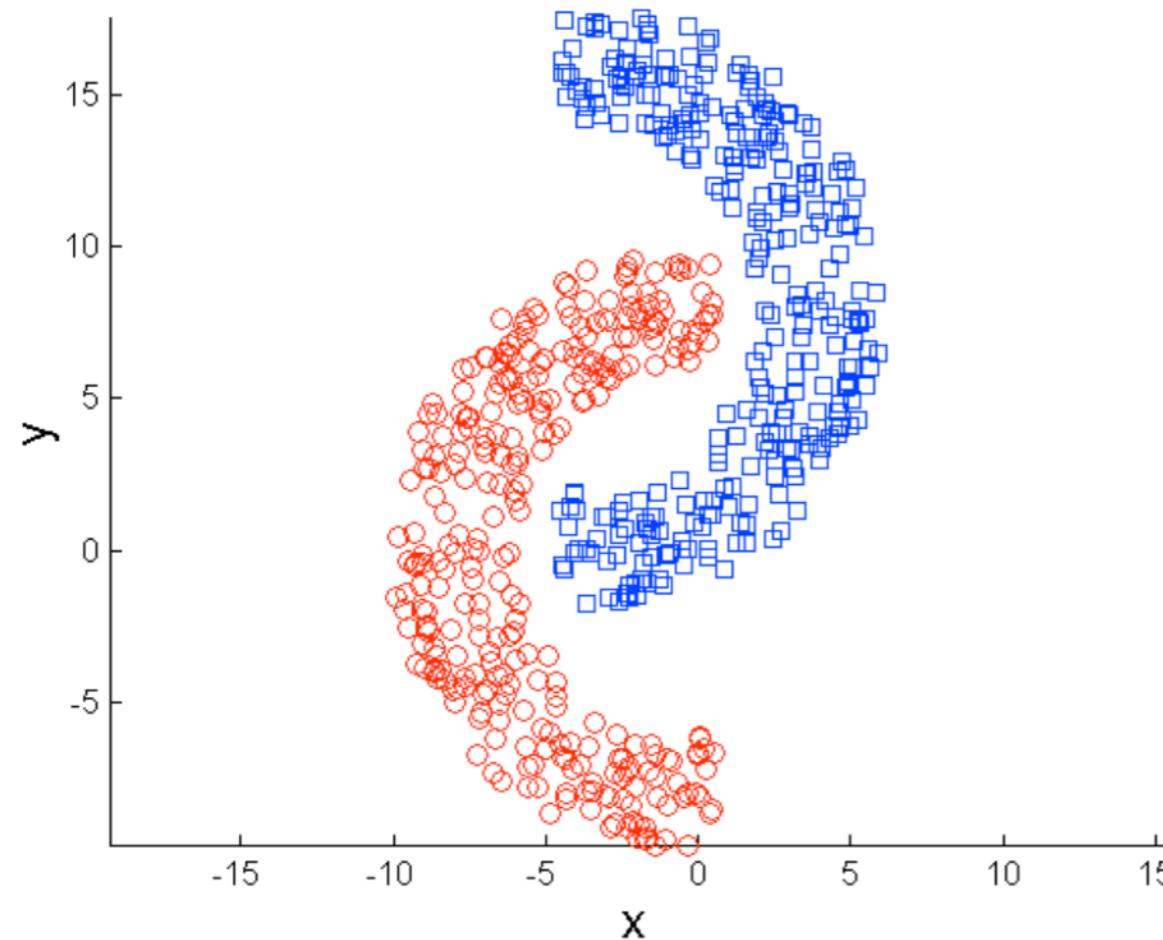
K-Means 3
Clusters



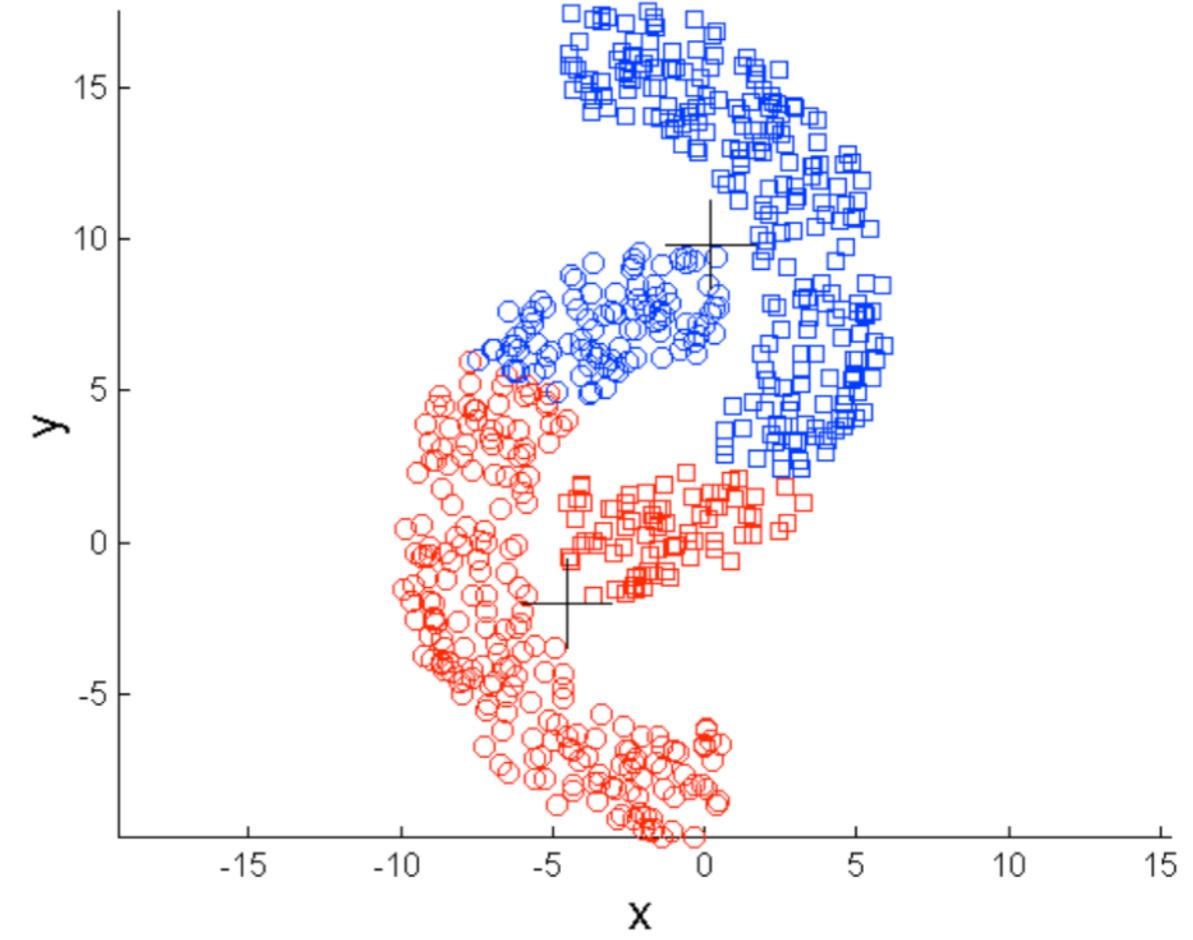
THE UNIVERSITY OF
SYDNEY

K-Means Limitations

Non-Globular Shapes:



Original Points



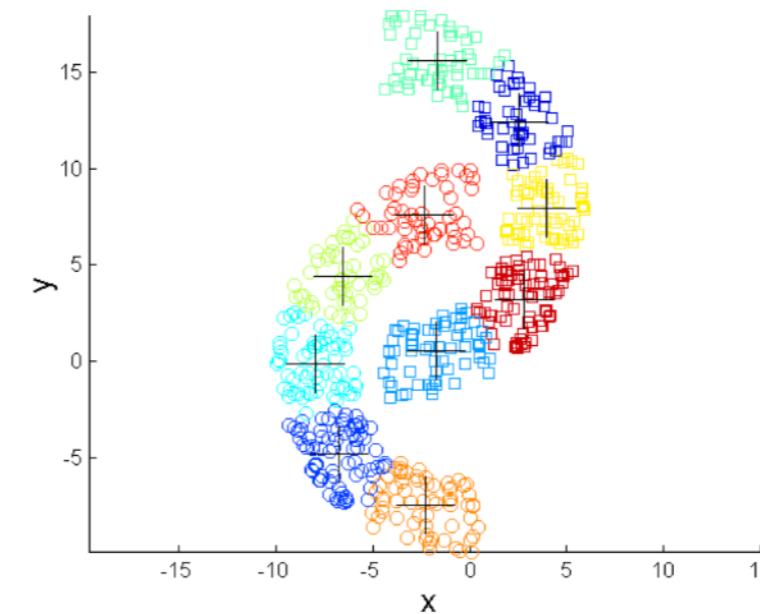
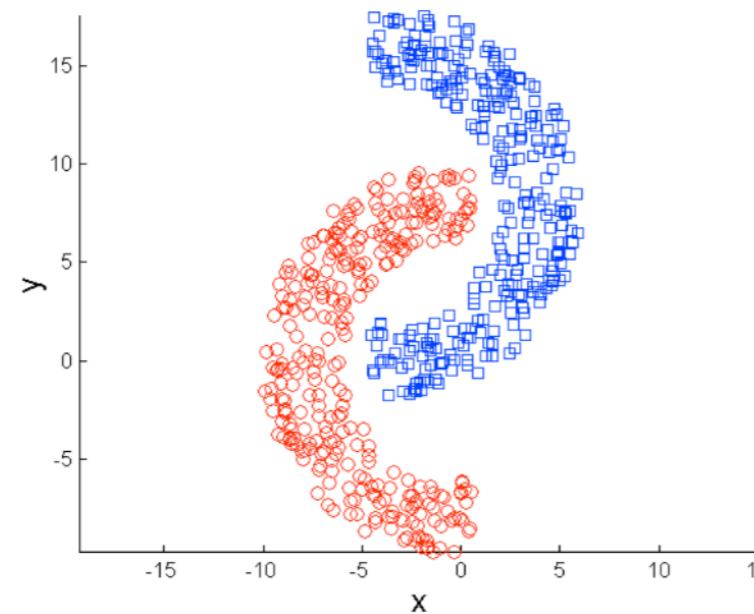
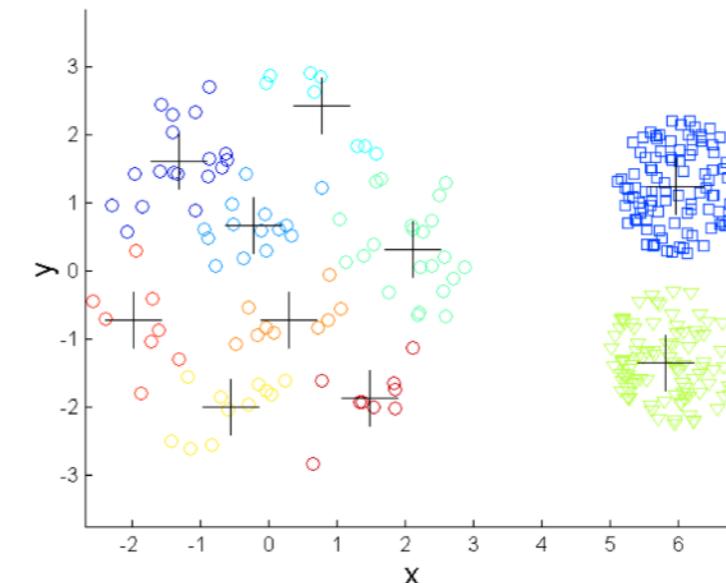
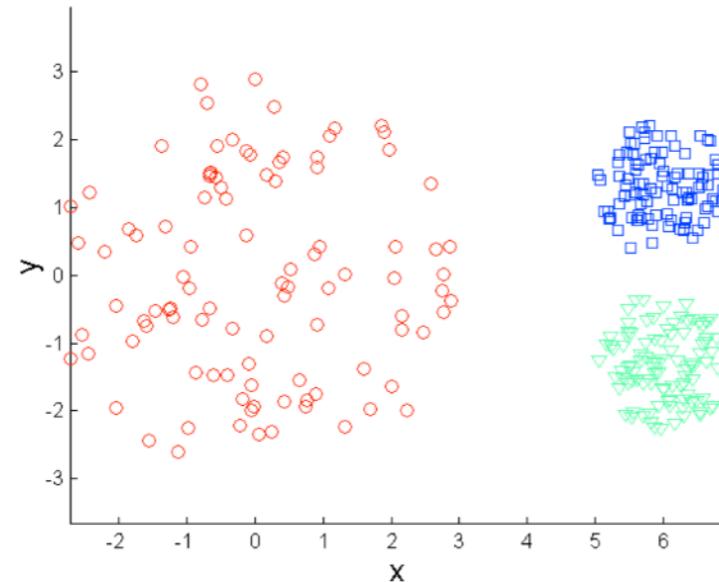
K-Means 2
Clusters



THE UNIVERSITY OF
SYDNEY

Overcome K-Means Limitations

Use large number of clusters.





THE UNIVERSITY OF
SYDNEY

K-Means Enhancements

Generalise distance function:

Allow different data types to be considered (for example, categorical variables).

Robustness to outliers.

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \mathcal{V}(\mathbf{x}_n, \boldsymbol{\mu}_k)$$



THE UNIVERSITY OF
SYDNEY

Hierarchical Clustering



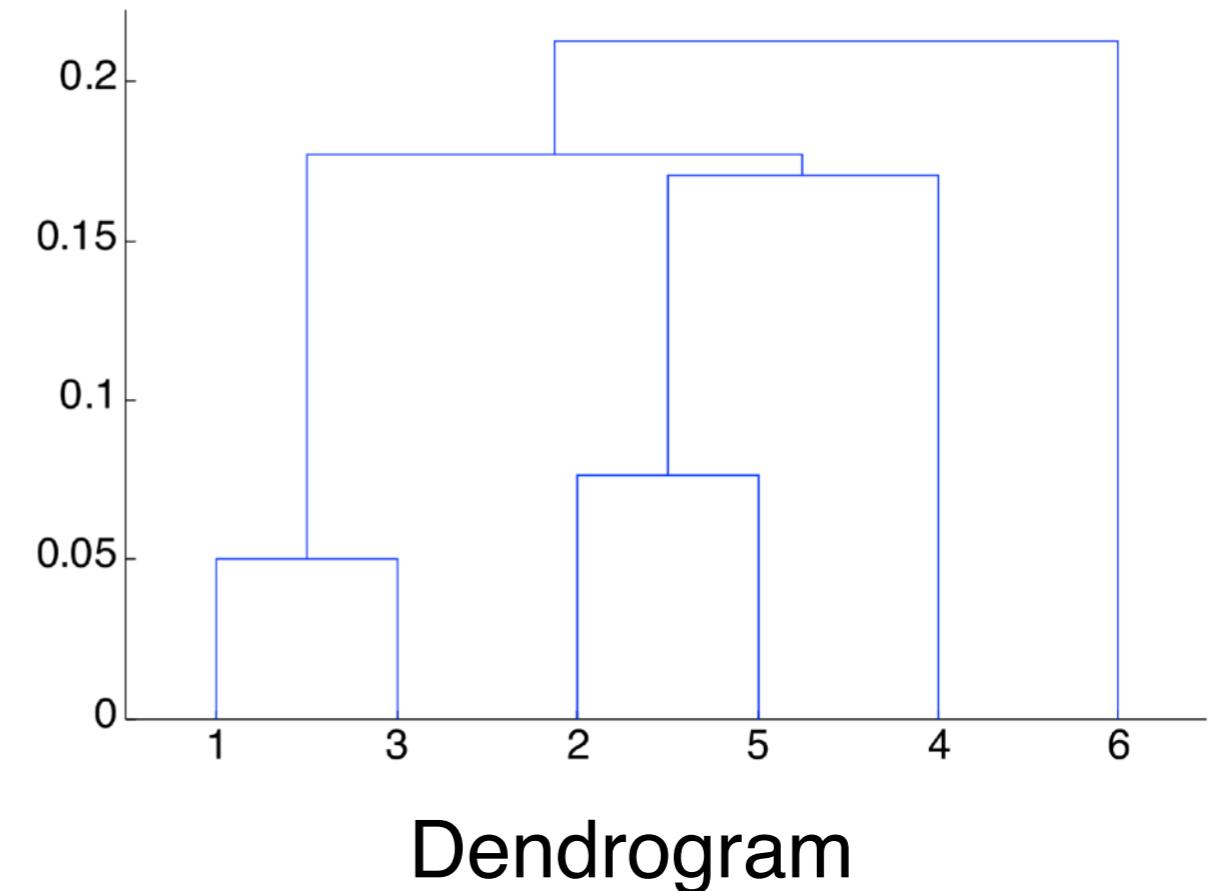
THE UNIVERSITY OF
SYDNEY

Hierarchical Clustering

Nested set of clusters organised as a hierarchical tree.

Any number of clusters can be obtained by ‘cutting’ the dendrogram.

Uses a similarity matrix.





THE UNIVERSITY OF
SYDNEY

Hierarchical Agglomerative Clustering

Simple clustering algorithm.

Uses a inter cluster similarity measure.

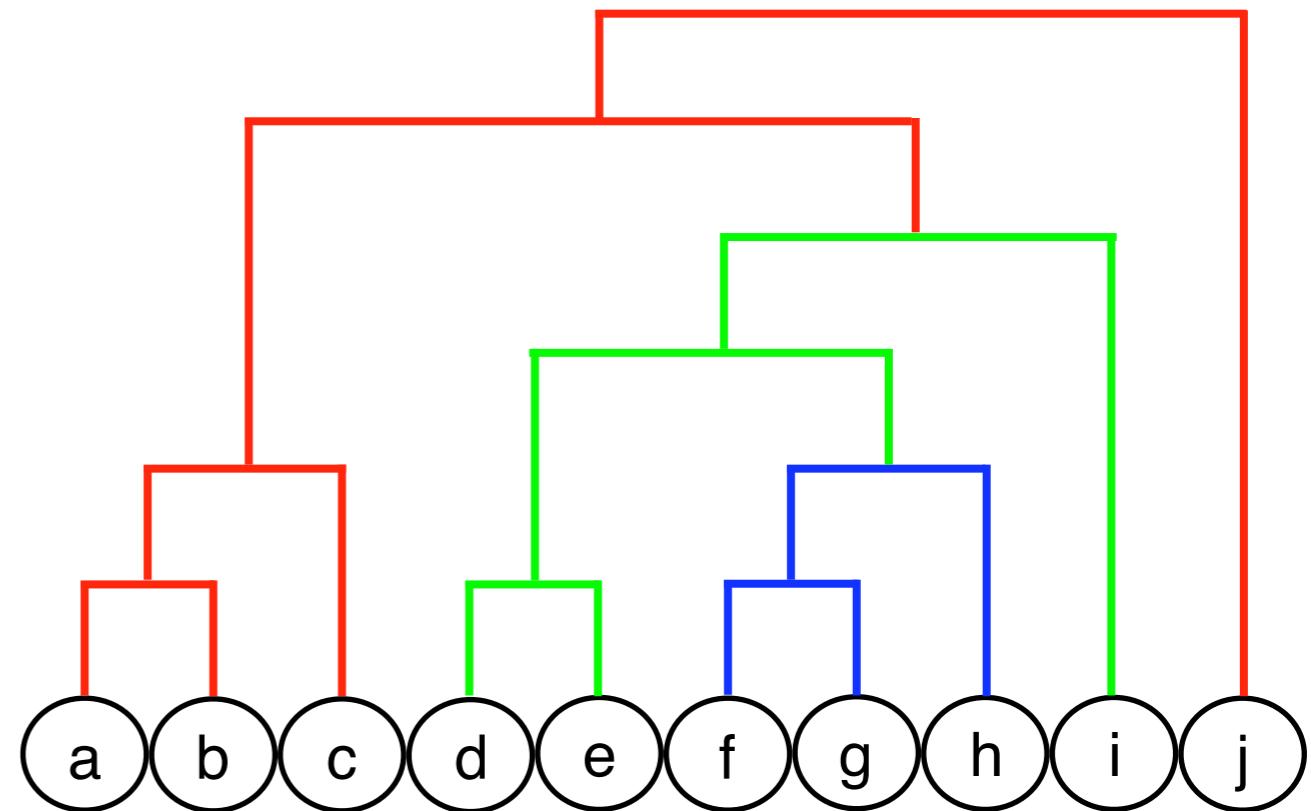
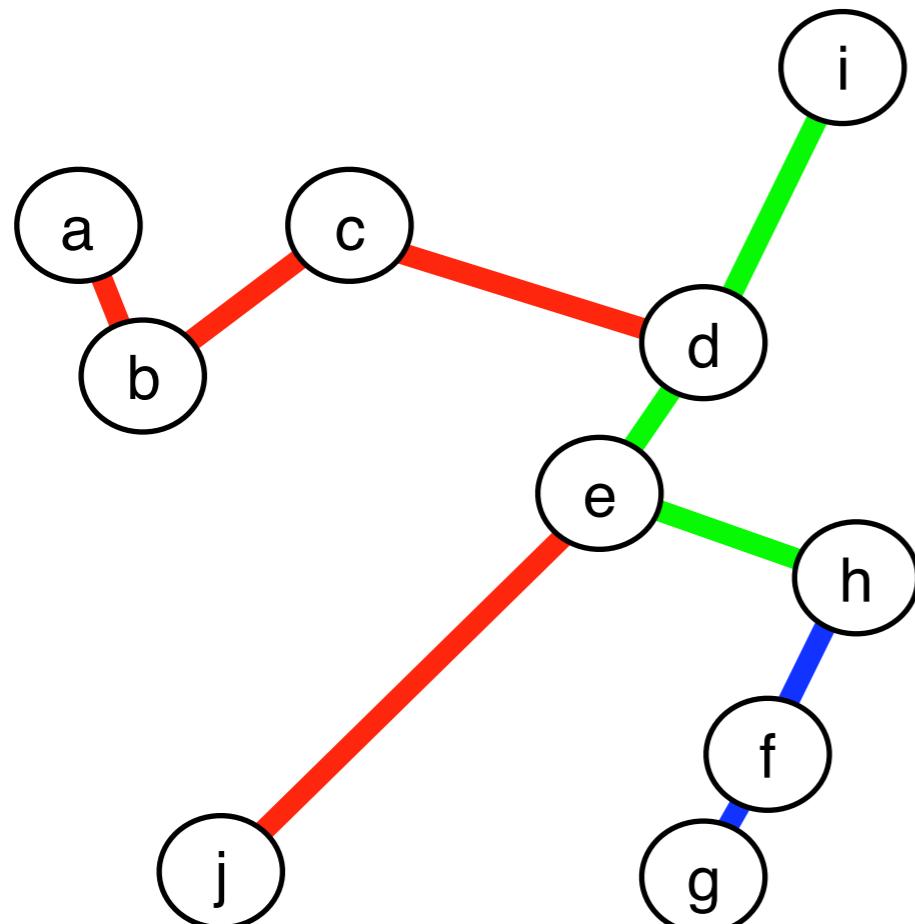
- 1 Initialise: Every data point is a cluster.
- 2 Repeat until one cluster remains.
 - 3 Compute distances between all clusters.
 - 4 Merge closest clusters.
 - 5 Update dendrogram.



THE UNIVERSITY OF
SYDNEY

Hierarchical Agglomerative Clustering

Example:



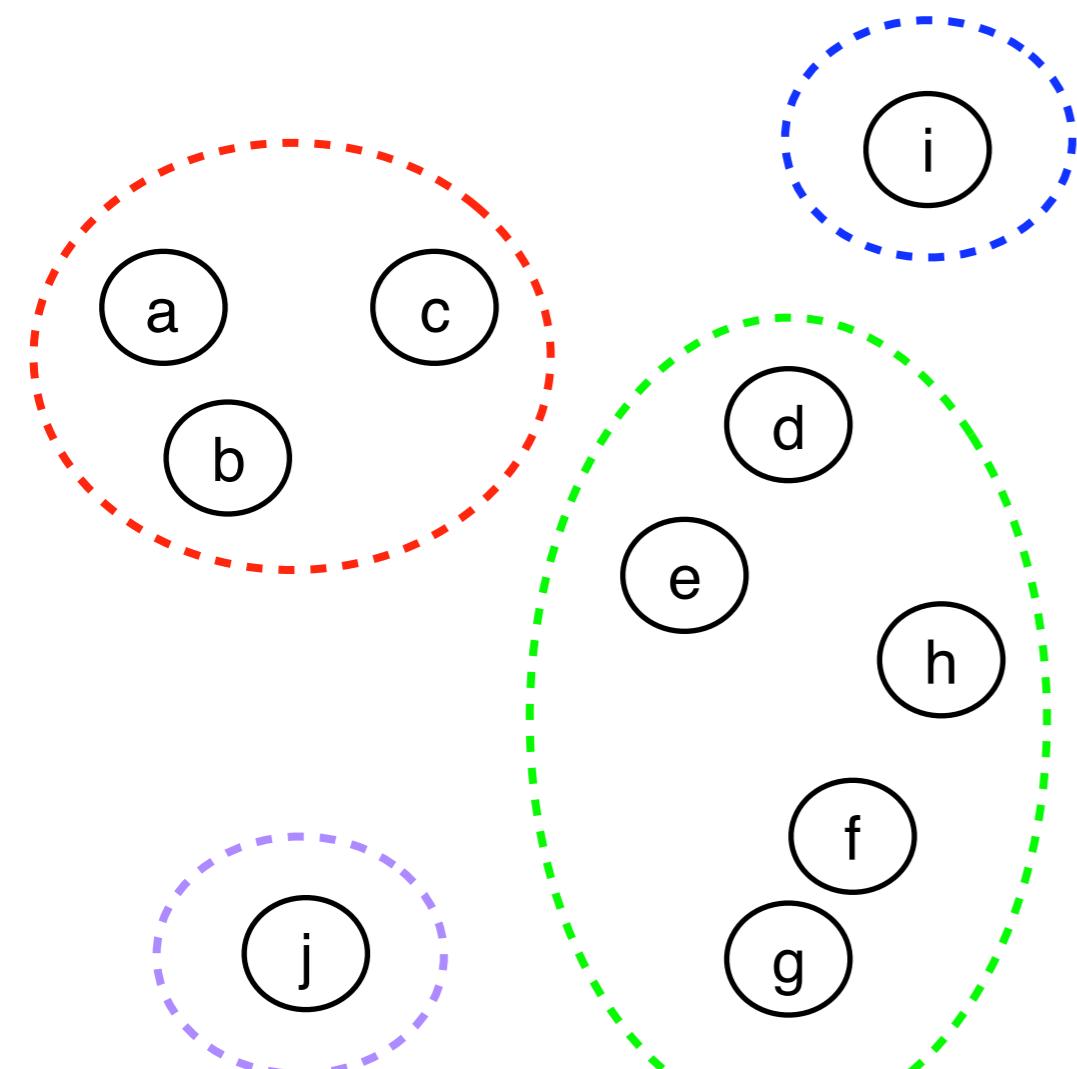
Dendrogram



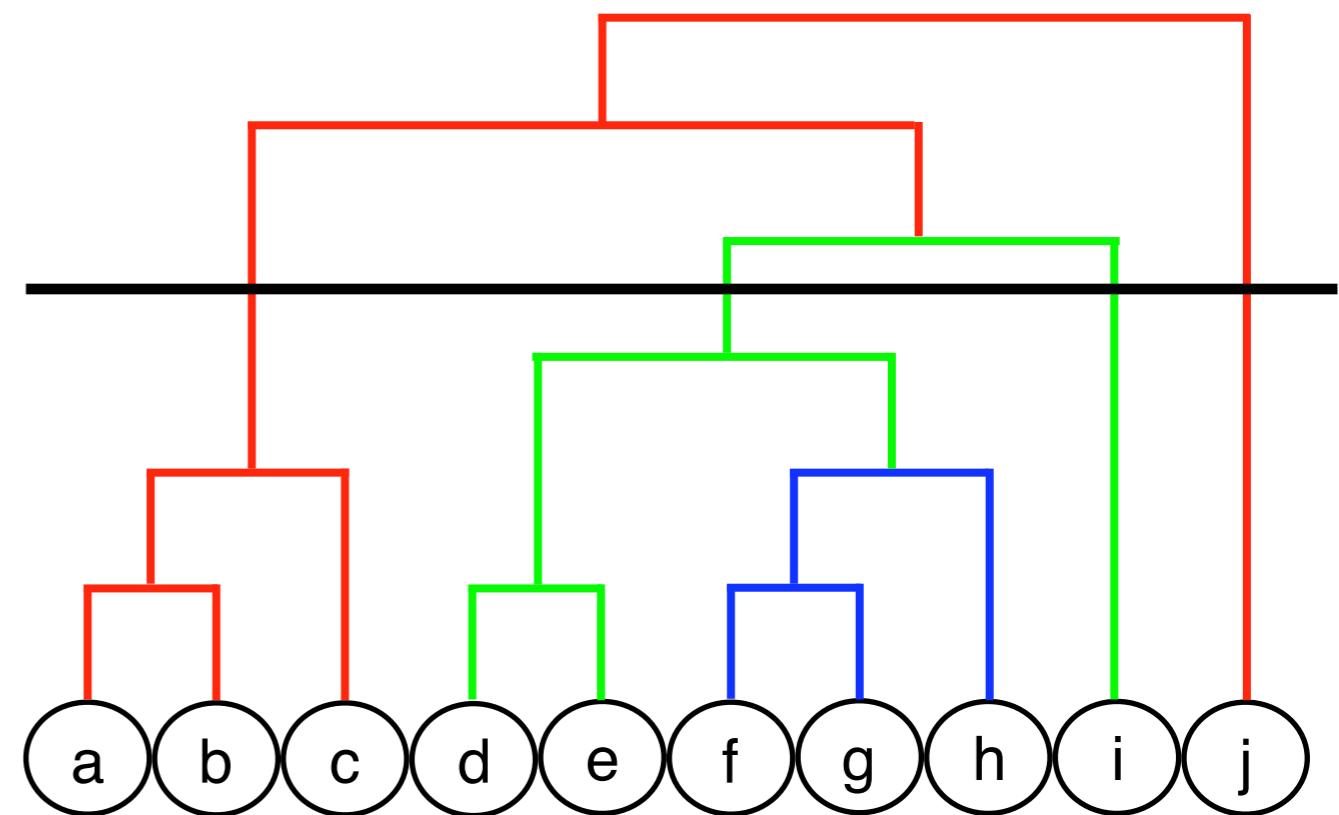
THE UNIVERSITY OF
SYDNEY

Hierarchical Agglomerative Clustering

Example:



'Cut' for K = 4



Dendrogram



Inter Cluster Similarity

Nearest Neighbour

$$D_{\min}(C_i, C_j) = \min_{\mathbf{x} \in C_i, \mathbf{y} \in C_j} \|\mathbf{x} - \mathbf{y}\|^2$$

Furthest Neighbour

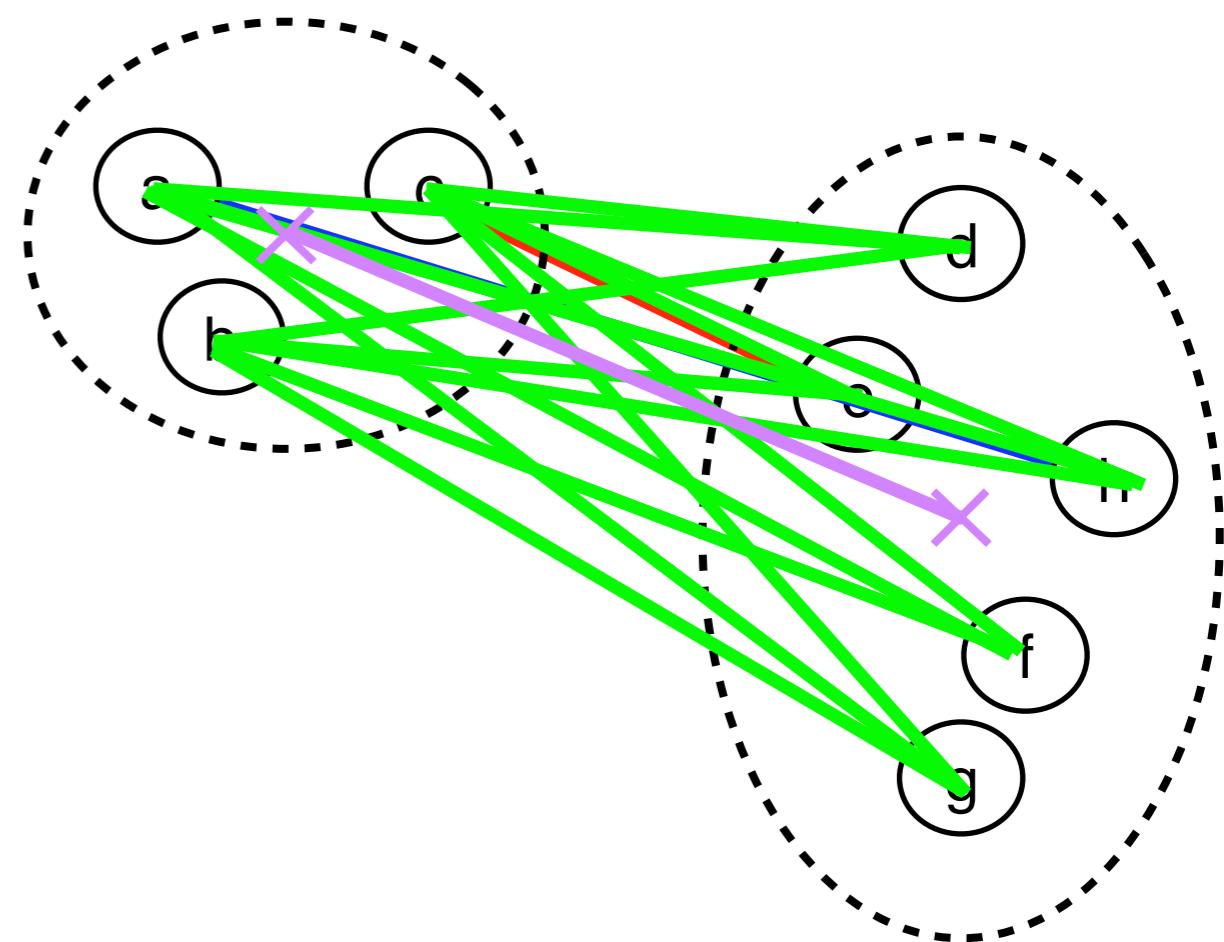
$$D_{\max}(C_i, C_j) = \max_{\mathbf{x} \in C_i, \mathbf{y} \in C_j} \|\mathbf{x} - \mathbf{y}\|^2$$

Group Average

$$D_{\text{avg}}(C_i, C_j) = \frac{1}{|C_i||C_j|} \sum_{\mathbf{x} \in C_i} \sum_{\mathbf{y} \in C_j} \|\mathbf{x} - \mathbf{y}\|^2$$

Centroid Distance

$$D_{\text{means}}(C_i, C_j) = \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|$$



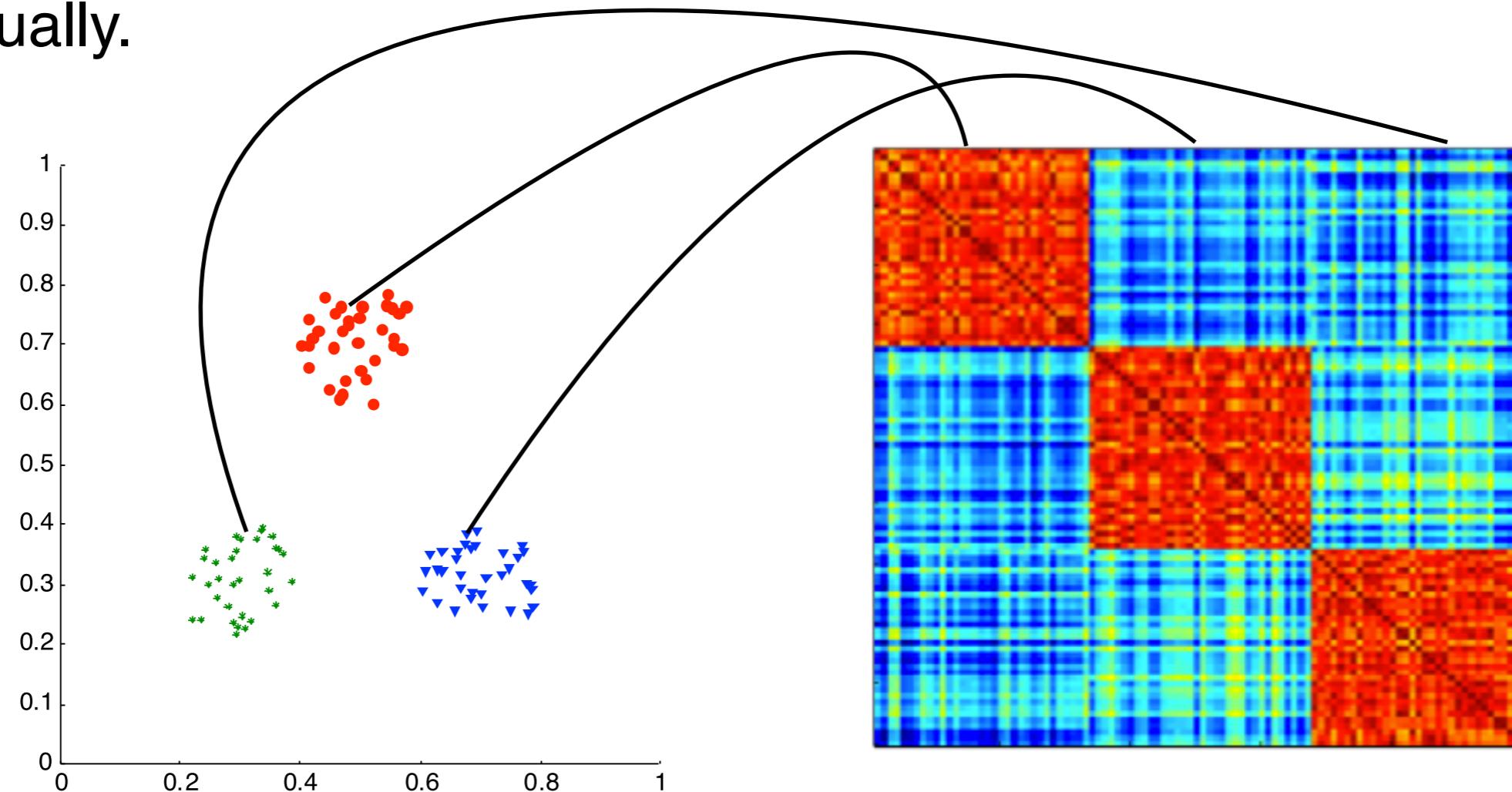
Class centroids will update
as we move through
the iterations.



THE UNIVERSITY OF
SYDNEY

Cluster Validation

Similarity matrix with respect to cluster labels and inspect visually.

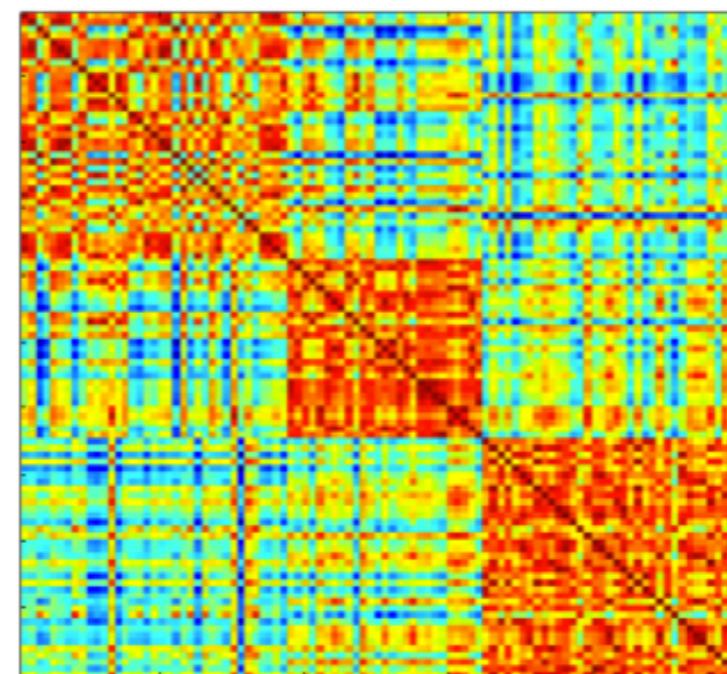
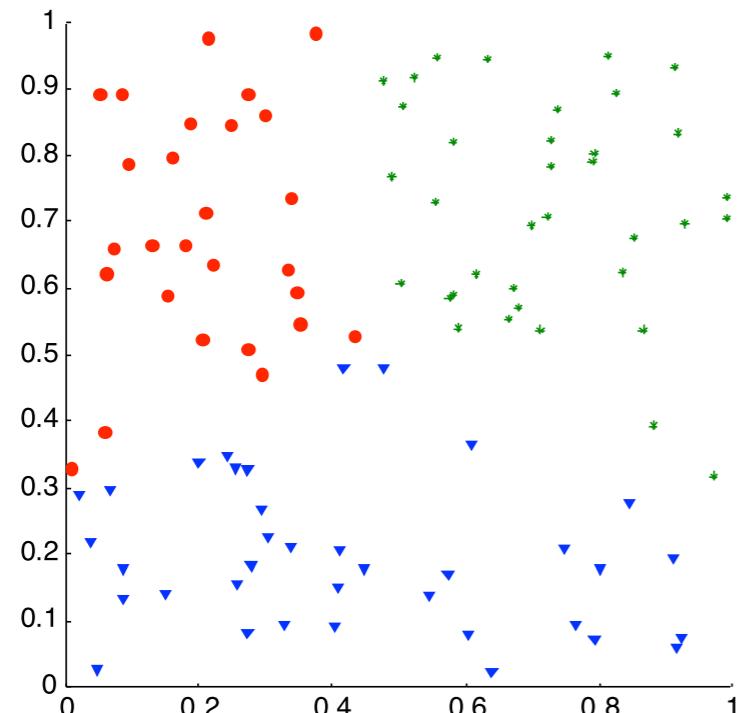




Cluster Validation

THE UNIVERSITY OF
SYDNEY

Random data clusters are not well defined.





THE UNIVERSITY OF
SYDNEY

Probabilistic Approach to Clustering



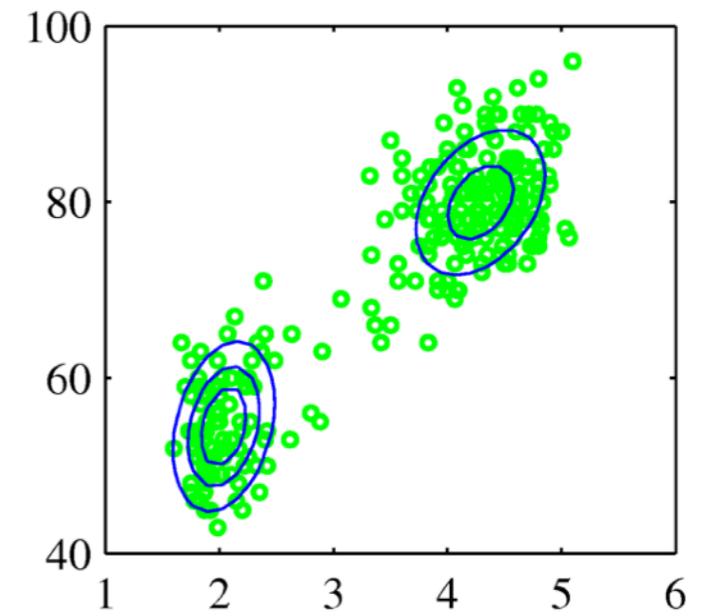
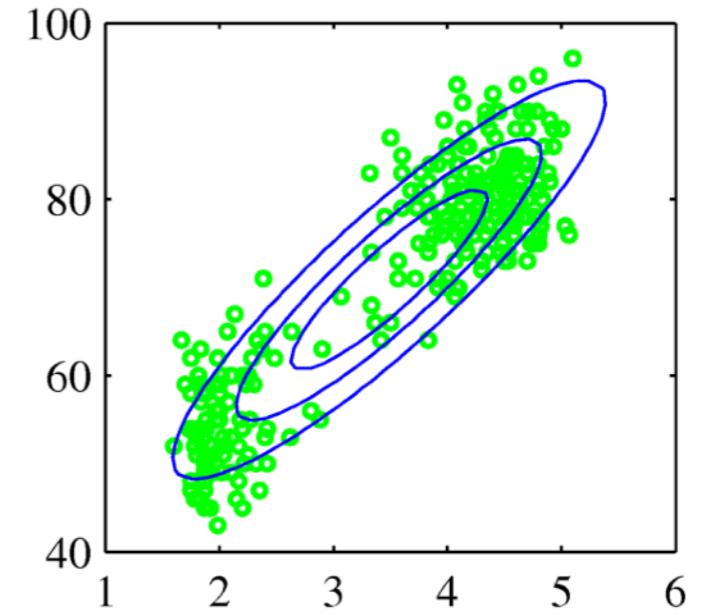
Mixture of Gaussians

Gaussian mixture distribution with K components.

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

$$0 \leq \pi_k \leq 1 \quad \sum_{k=1}^K \pi_k = 1$$

Mixture models provide a probabilistic framework for clustering.
Soft version of K-Means.





THE UNIVERSITY OF
SYDNEY

Mixture of Gaussians

Let us introduce a latent random variable

$$\mathbf{z} = \{z_k\}_{k \in 1, \dots, K} \quad z_k \in \{0, 1\} \quad \sum_{k=1}^K z_k = 1$$

\mathbf{z} has K possible states.

$$p(z_k = 1) = \pi_k \quad p(\mathbf{z}) = \prod_{k=1}^K \pi_k^{z_k}$$
$$p(\mathbf{x}|z_k = 1) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^K \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_k}$$

$$p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{x}|\mathbf{z})p(\mathbf{z}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$



Mixture of Gaussians

Let us apply Bayes theorem and infer the value of the latent variable.

$$p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{p(\mathbf{x})}$$

$$\begin{aligned}\gamma(z_k) \equiv p(z_k = 1|\mathbf{x}) &= \frac{p(z_k = 1)p(\mathbf{x}|z_k = 1)}{p(\mathbf{x})} \\ &= \frac{\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}\end{aligned}$$

$\gamma(z_k)$ is called *responsibility* that component k takes for explaining \mathbf{x} .

π_k is the prior probability of component k.

$\gamma(z_k)$ is the posterior probability after \mathbf{x} is observed.

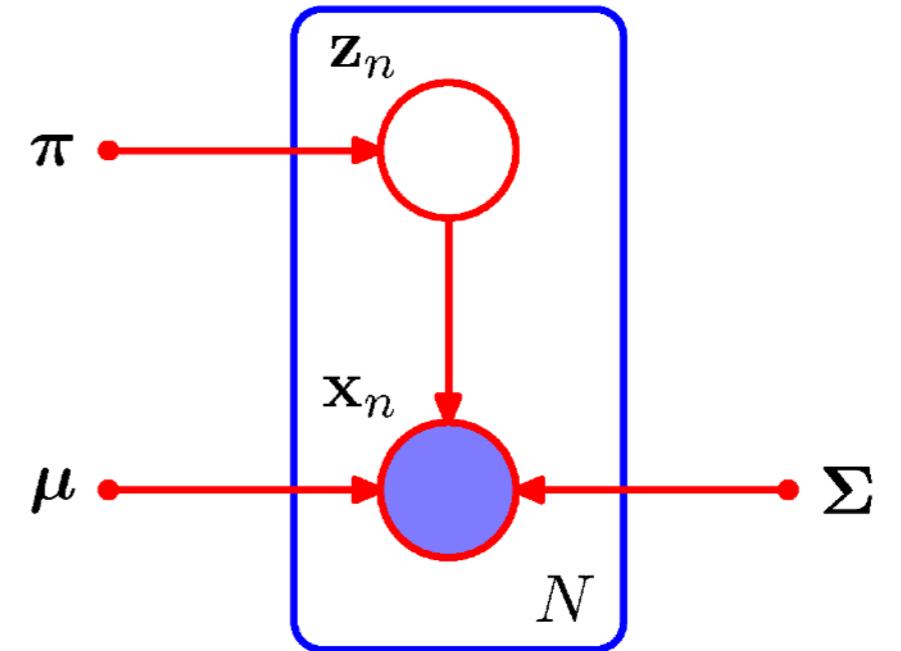


Mixture of Gaussians

N datapoints $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$

induce N latent variables \mathbf{z}_n

$$\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_N\}$$



Likelihood:

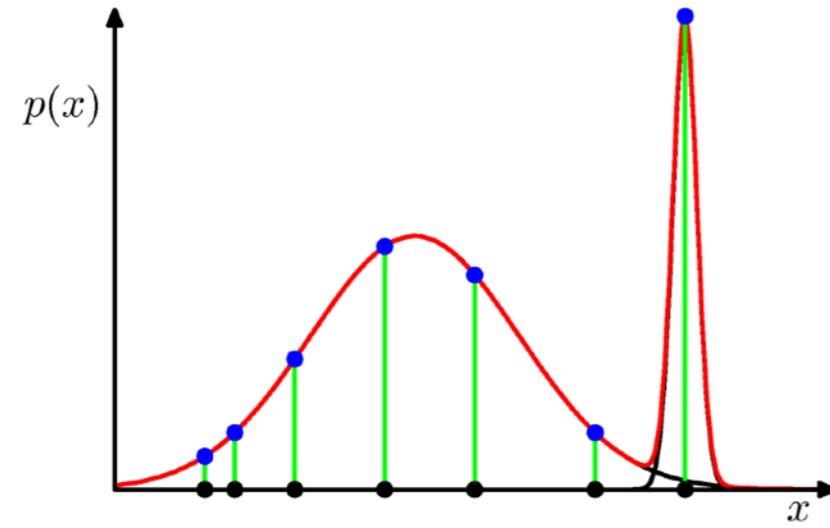
$$\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\}$$

(Assuming samples are drawn independently from the distribution)



THE UNIVERSITY OF
SYDNEY

Maximum Likelihood for Mixture Models



One components fits one data point.

Standard deviation shrinks to zero.

$$\mathcal{N}(\mathbf{x}_n | \mathbf{x}_n, \sigma_j^2 \mathbf{I}) = \frac{1}{(2\pi)^{1/2}} \frac{1}{\sigma_j}$$

Presence of singularities.

Maximum likelihood is not a well posed problem.



THE UNIVERSITY OF
SYDNEY

Expectation Maximisation

EM for Gaussian Mixtures



THE UNIVERSITY OF
SYDNEY

Expectation Maximisation (EM)

Elegant and powerful method for finding MLE or MAP solutions for models with latent variables.

Intuition: If we knew what cluster each point belonged to (i.e. the z variables), we could partition the data and find the MLE for each cluster separately.



EM for Gaussian Mixtures

Likelihood function:

$$\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\}$$

Conditions to be satisfied at maximum likelihood:

$$0 = \frac{\partial \ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}_k}$$

$$0 = - \sum_{n=1}^N \underbrace{\frac{\pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_j \pi_j \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}}_{\gamma(z_{nk})} \boldsymbol{\Sigma}_k (\mathbf{x}_n - \boldsymbol{\mu}_k)$$

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) \mathbf{x}_n$$

$$N_k = \sum_{n=1}^N \gamma(z_{nk})$$

N_k is the effective number of points assigned to cluster k.



THE UNIVERSITY OF
SYDNEY

EM for Gaussian Mixtures

$$0 = \frac{\partial \ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}_k}$$

$$\Rightarrow \boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk})(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^T$$

$$0 = \frac{\partial \ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \pi_k}$$

$$\Rightarrow \pi_k = \frac{N_k}{N}$$



EM Algorithm

1 Initialise means μ_k , covariances Σ_k and mixing coefficients π_k .

2 E-step

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

3 M-step

$$\begin{aligned}\boldsymbol{\mu}_k^{\text{new}} &= \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) \mathbf{x}_n \\ \boldsymbol{\Sigma}_k^{\text{new}} &= \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}}) (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}})^T \\ \pi_k^{\text{new}} &= \frac{N_k}{N}\end{aligned}$$

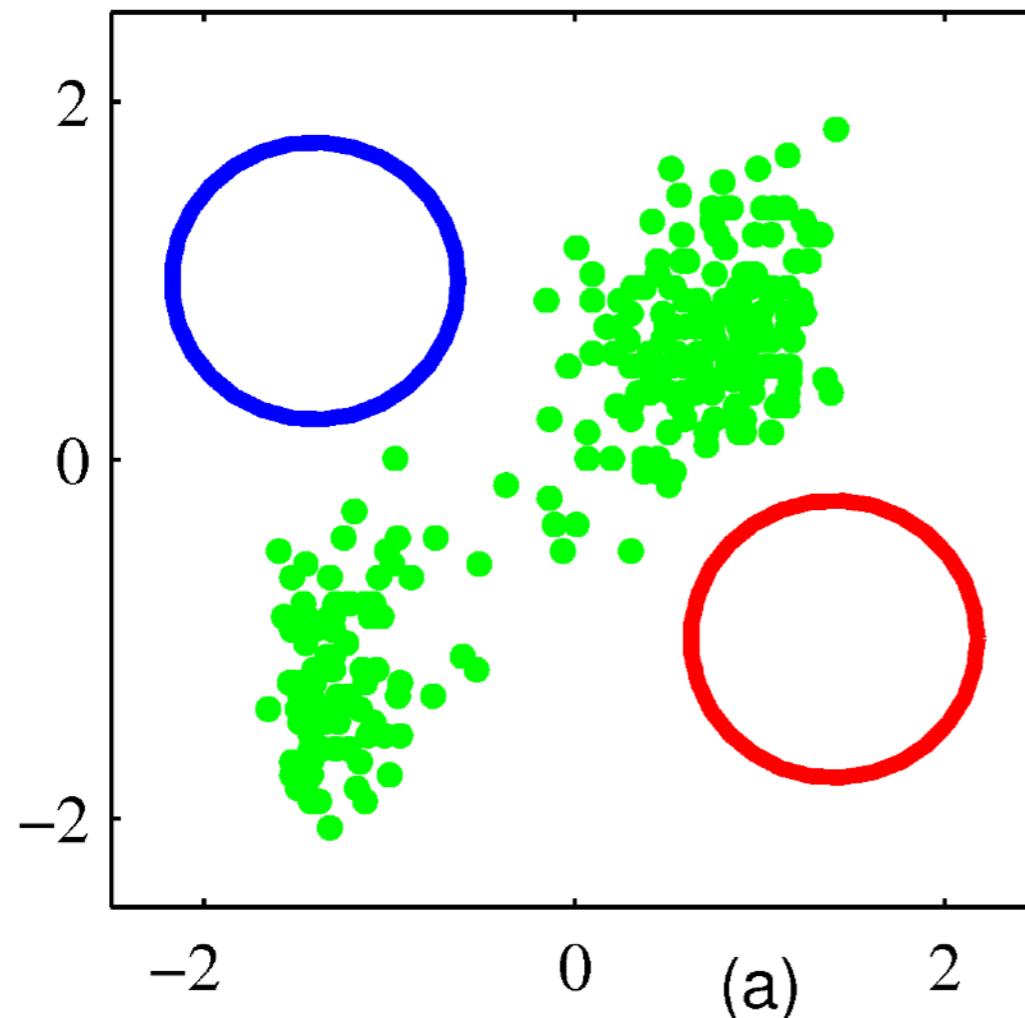
4 Eval Likelihood

$$\ln p(\mathbf{X} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\}$$



THE UNIVERSITY OF
SYDNEY

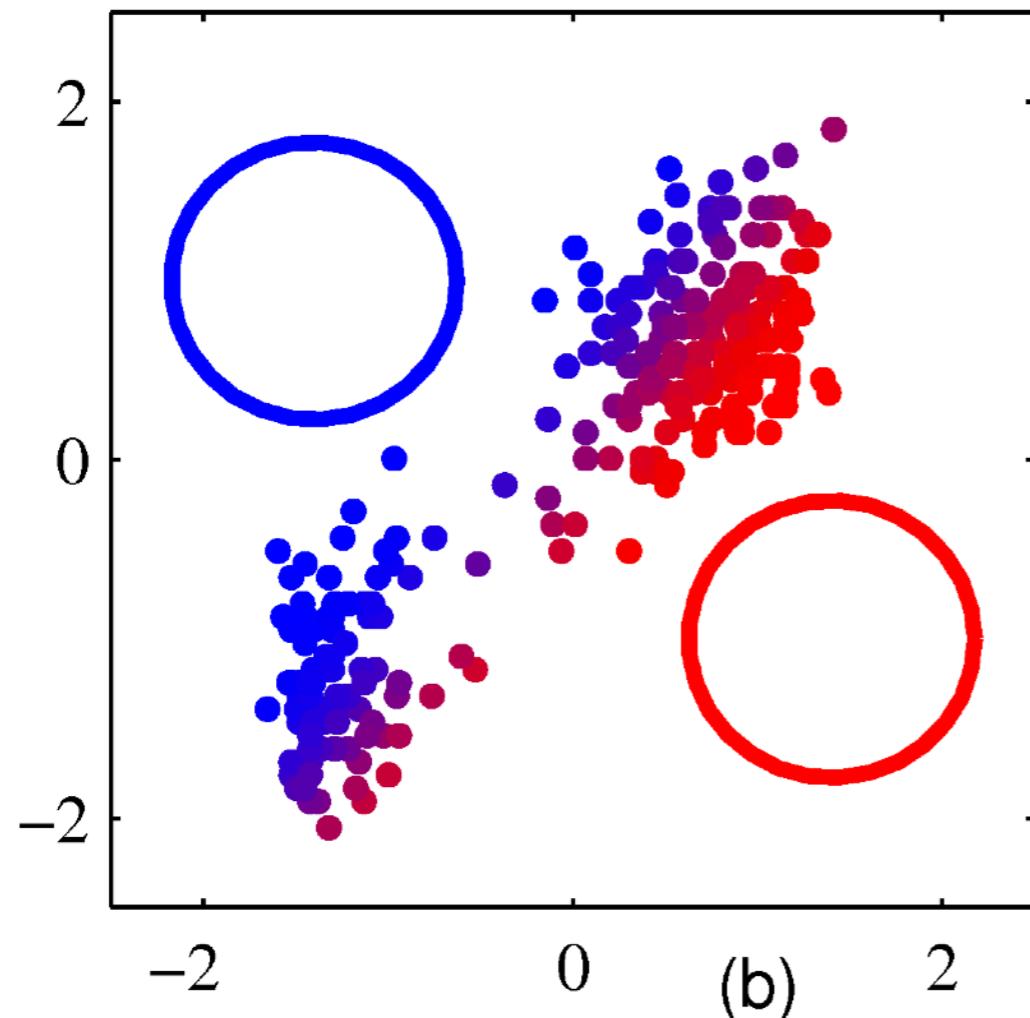
EM Example



Initial values for mean vectors
(same as K-means example).

Diagonal covariance matrices
(showing one std contour).

EM Example



Initial E step.

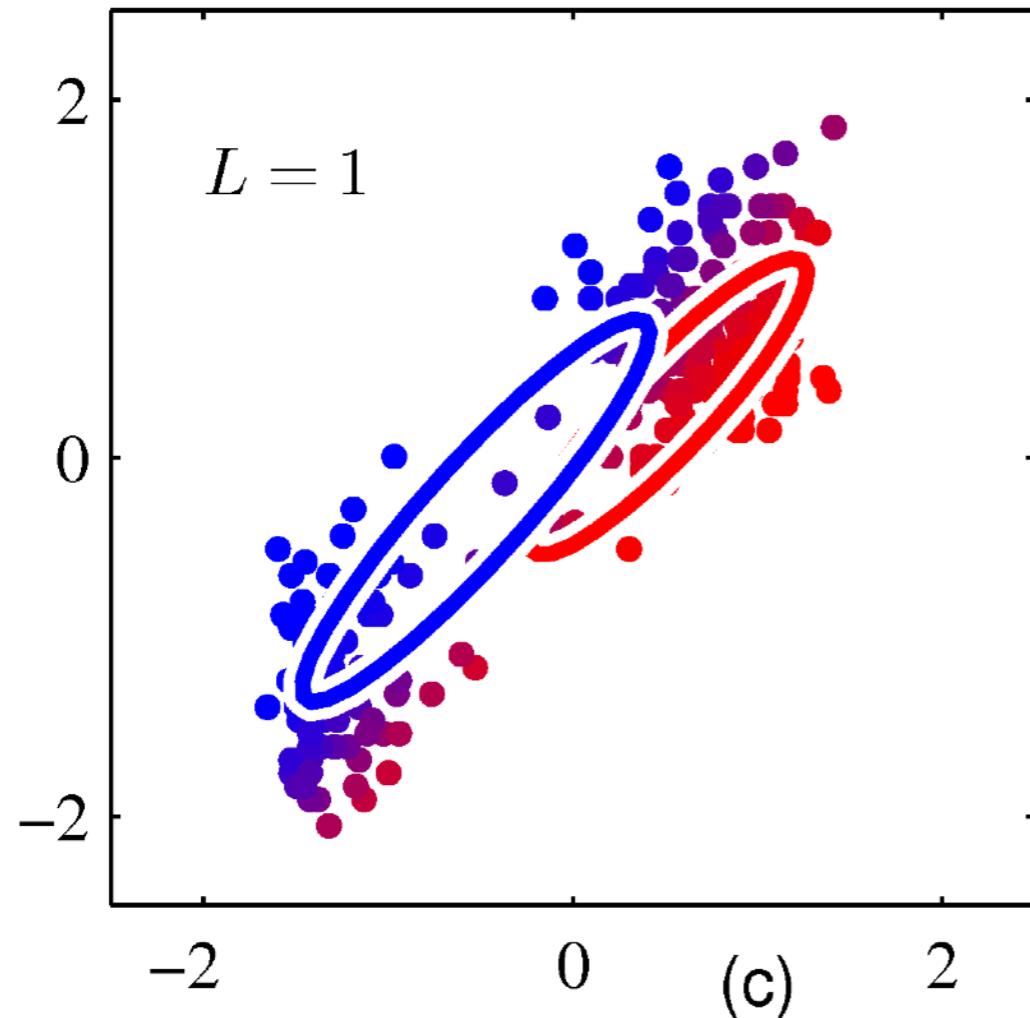
$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}.$$

Colour
proportional to responsibilities.



EM Example

M step:



The means move towards the weighted average of dataset with respective ink colour (responsibilities).

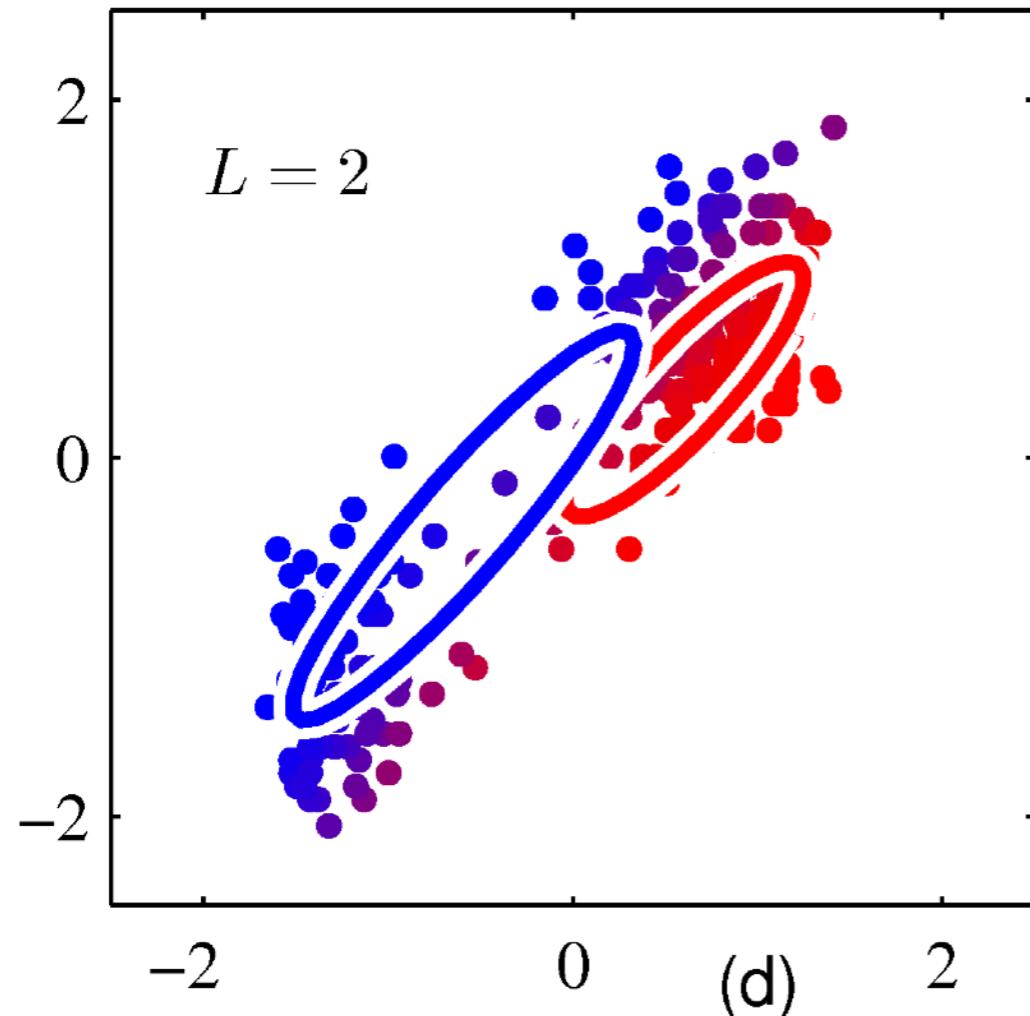
$$\mu_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) \mathbf{x}_n$$

The covariance matrices adapt to the covariance of the respective ink.

$$\Sigma_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (\mathbf{x}_n - \mu_k^{\text{new}}) (\mathbf{x}_n - \mu_k^{\text{new}})^T$$

EM Example

M step:



The means move towards the weighted average of dataset with respective ink colour (responsibilities).

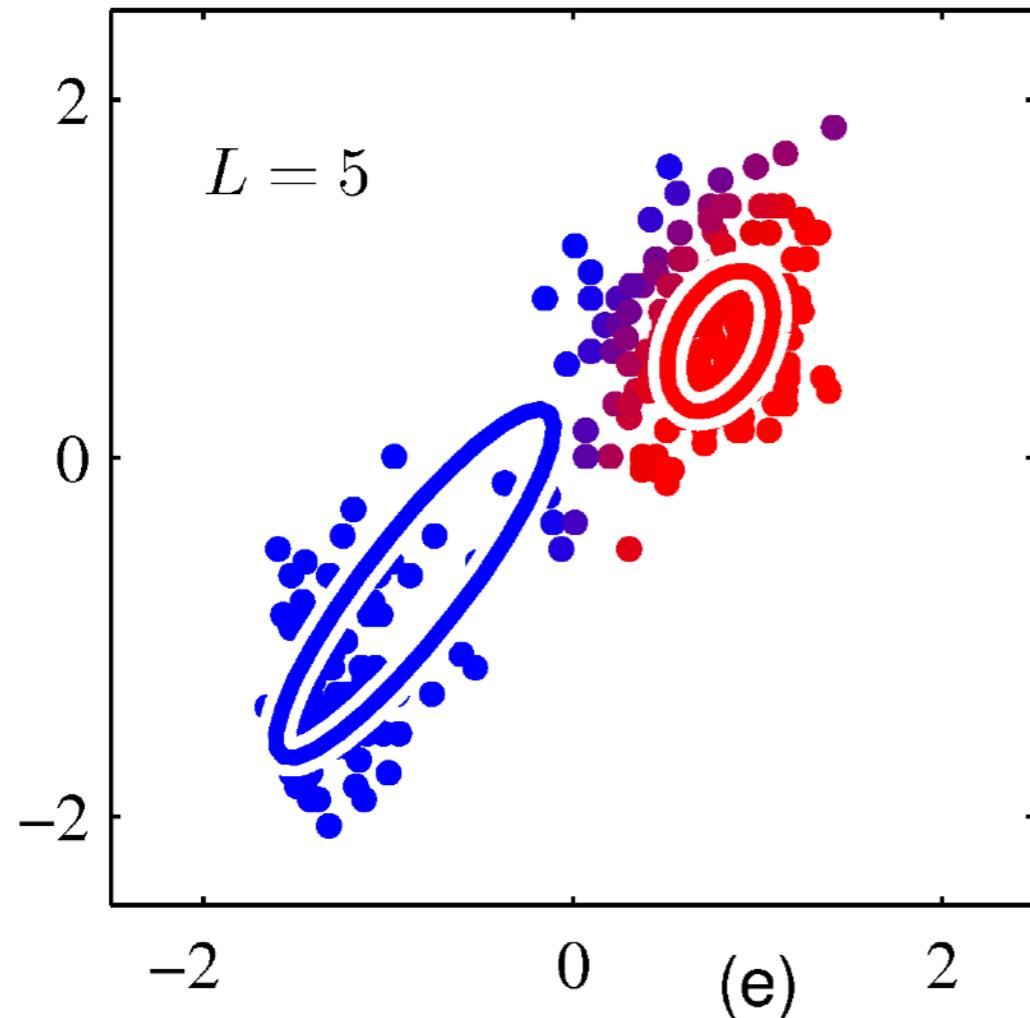
$$\boldsymbol{\mu}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) \mathbf{x}_n$$

The covariance matrices adapt to the covariance of the respective ink.

$$\boldsymbol{\Sigma}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}}) (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}})^T$$

EM Example

M step:



The means move towards the weighted average of dataset with respective ink colour (responsibilities).

$$\boldsymbol{\mu}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) \mathbf{x}_n$$

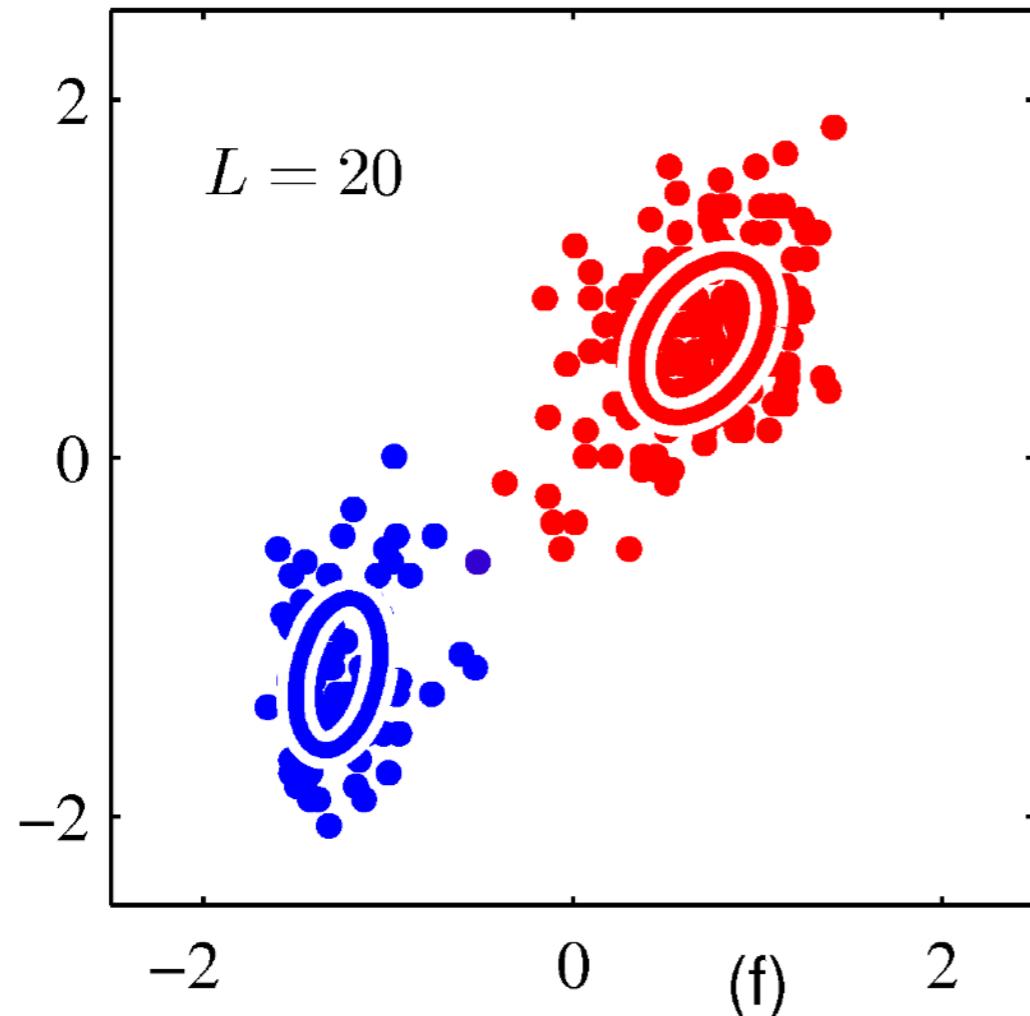
The covariance matrices adapt to the covariance of the respective ink.

$$\boldsymbol{\Sigma}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}}) (\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}})^T$$



EM Example

M step:



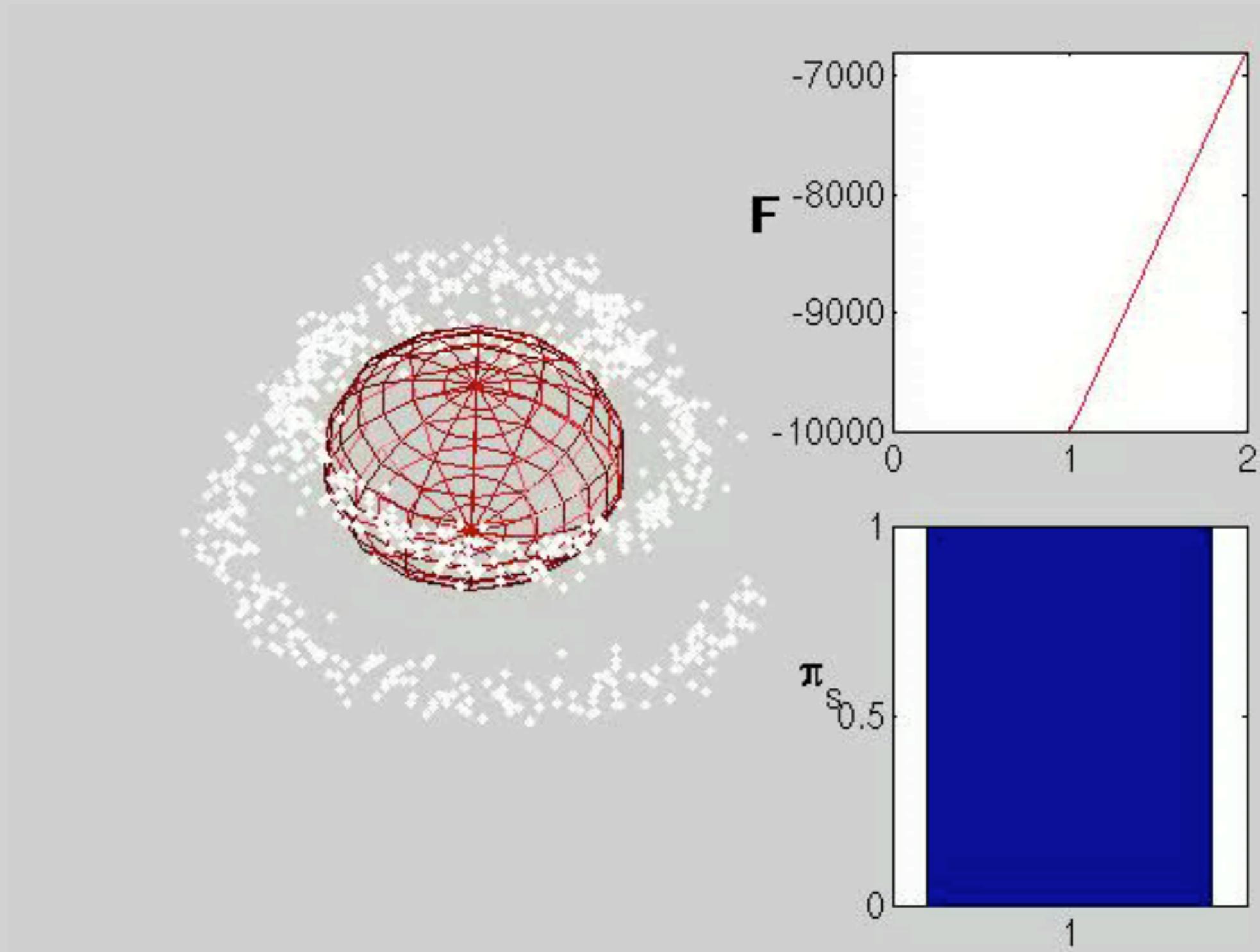
The means move towards the weighted average of dataset with respective ink colour (responsibilities).

$$\mu_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) \mathbf{x}_n$$

The covariance matrices adapt to the covariance of the respective ink.

$$\Sigma_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (\mathbf{x}_n - \mu_k^{\text{new}}) (\mathbf{x}_n - \mu_k^{\text{new}})^T$$

EM Example 2





THE UNIVERSITY OF
SYDNEY

Research

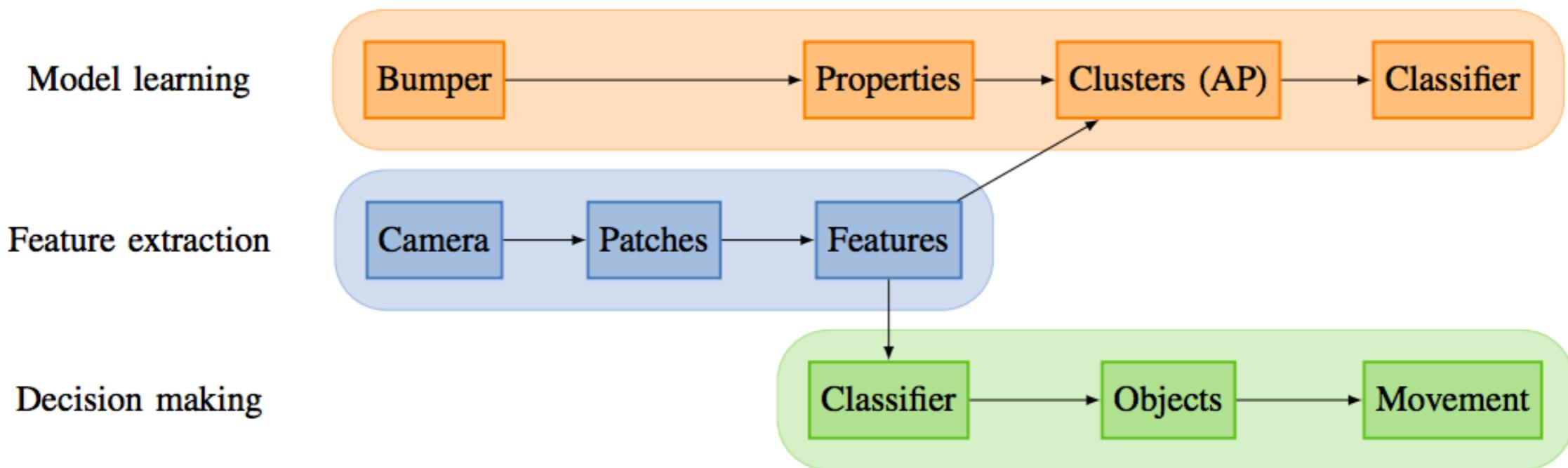


Incremental Learning for Long-Term Autonomy

Goal:

Learn the visual appearance of objects and whether or not they represent an obstacle.

Learn without human supervision.



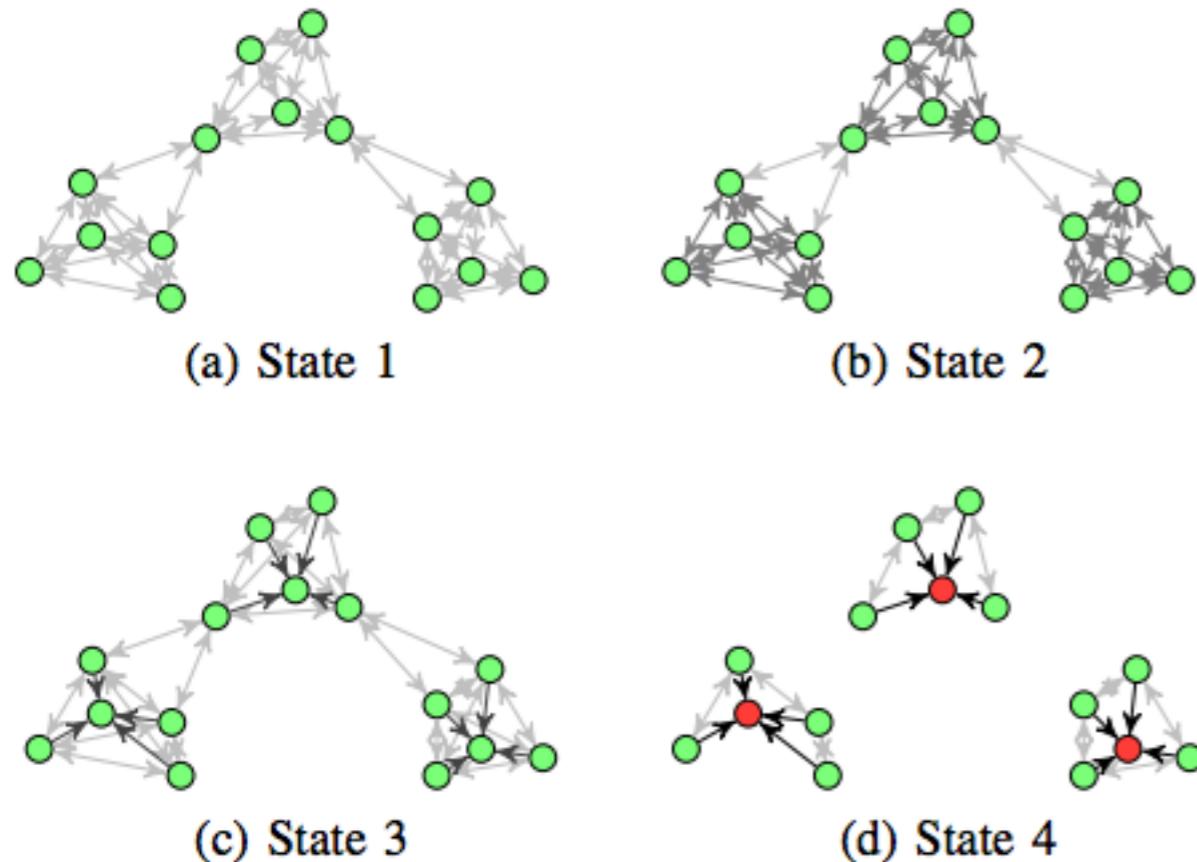


THE UNIVERSITY OF
SYDNEY

Incremental Learning for Long-Term Autonomy

Affinity Propagation:

Does not require a priori number of clusters.





Characterising Criminals

1. What are the important factors that affect the criminality levels of people over time.

- What are drivers for young children for becoming involved in with crime?
- How do life events diminish criminality over criminals?
(Desistance)

2. Is it possible to build a set of characteristics that describe certain type of criminals?

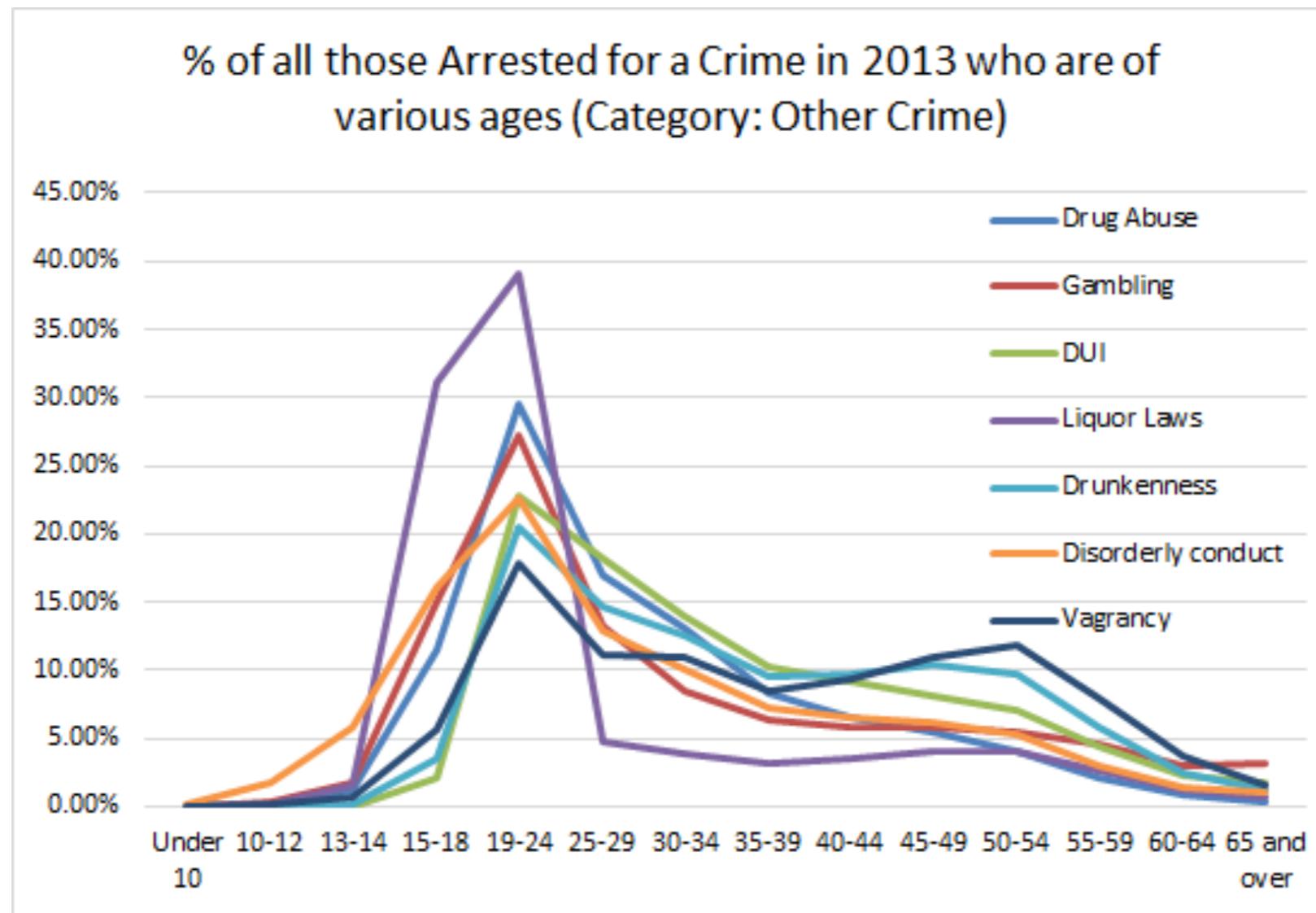
- Can we generate clusters of criminals?





THE UNIVERSITY OF
SYDNEY

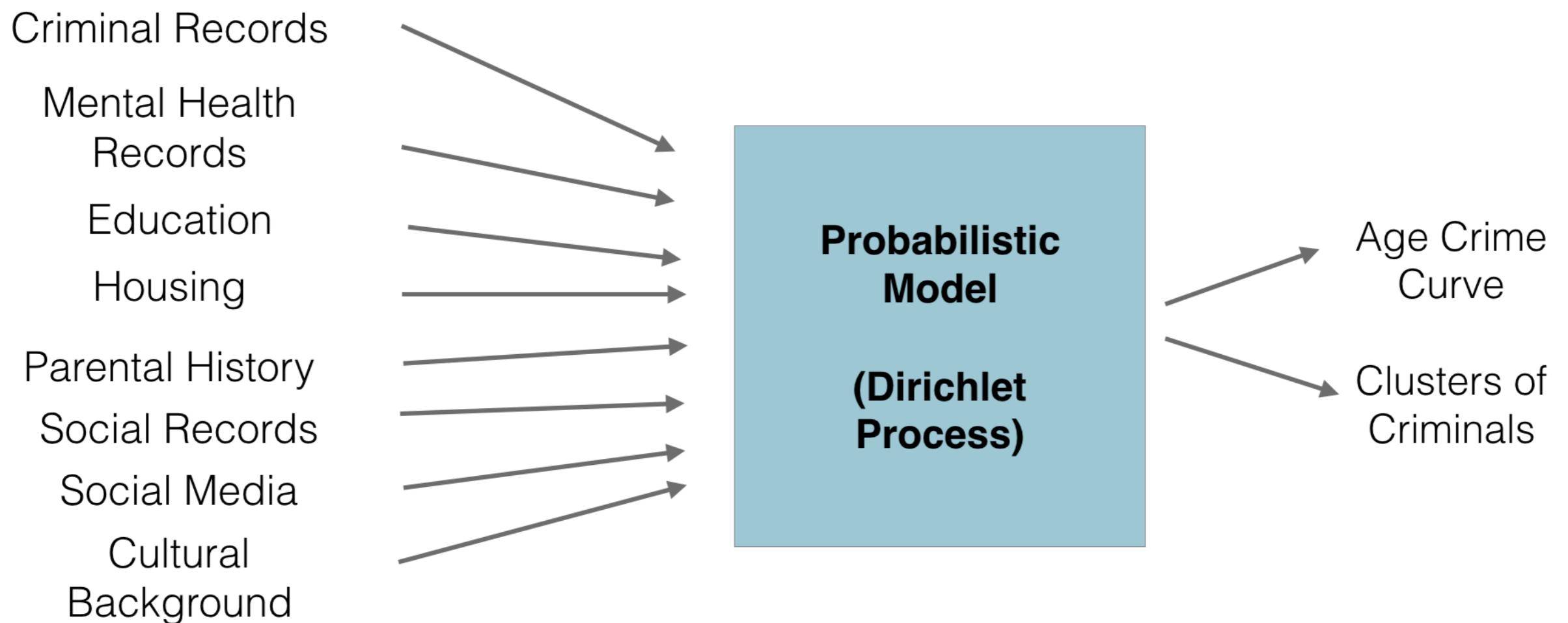
Clustering of Criminals





THE UNIVERSITY OF
SYDNEY

Clustering of Criminals

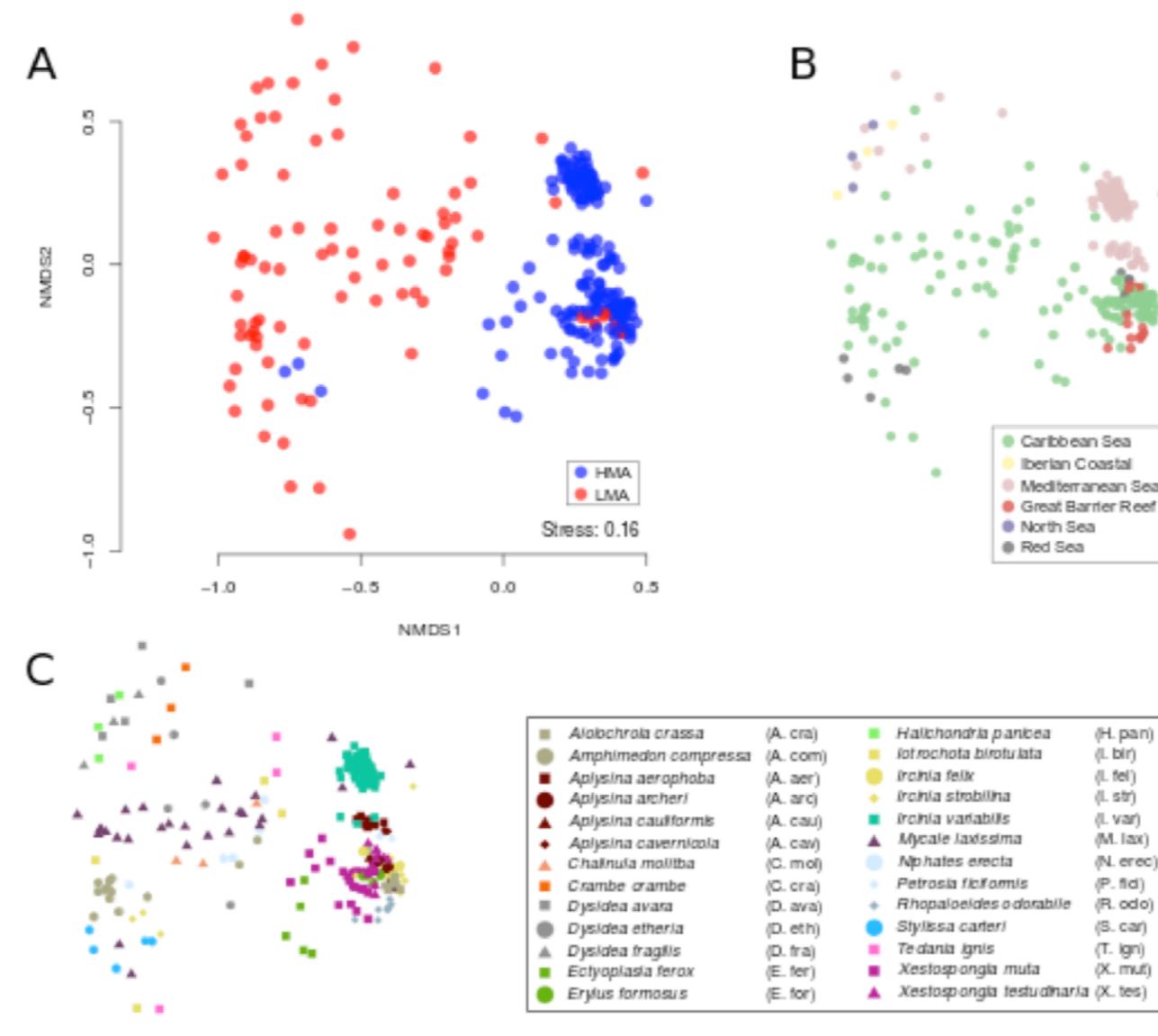


Profiling and Predicting Microbial Community



THE UNIVERSITY OF
SYDNEY

Learn the representation of the microbial community of sponges.



Profiling and Predicting Microbial Community



THE UNIVERSITY OF
SYDNEY

