

# Application of the Bees Algorithm to Fuzzy Clustering

D.T. Pham, H. AL-Jabbouli, M.  
Mahmuddin, S. Otri and A. Haj  
Darwish

**Carlos Henrique Maciel Sobral Timóteo**

# Roteiro

- Definições
- Algoritmo Fuzzy C-Médias (FCM)
- Abelhas na Natureza
- Algoritmo das Abelhas
- Algoritmo Proposto – Bees Algorithm with FCM
- Resultados Experimentais
- Conclusão

# Definições

- O Algoritmo das Abelhas é usado para otimizar o desempenho do FCM melhorando o resultado da clusterização.
- Resultados Experimentais mostram que o algoritmo proposto fornece uma melhoria significativa ao FCM e ao FCM combinado com o Algoritmo Genético (GA).

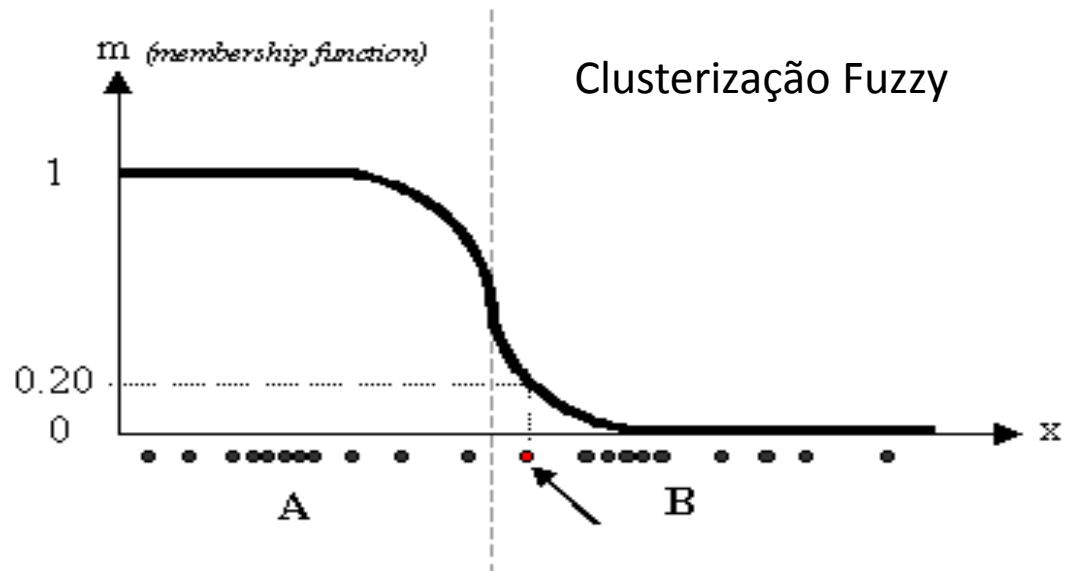
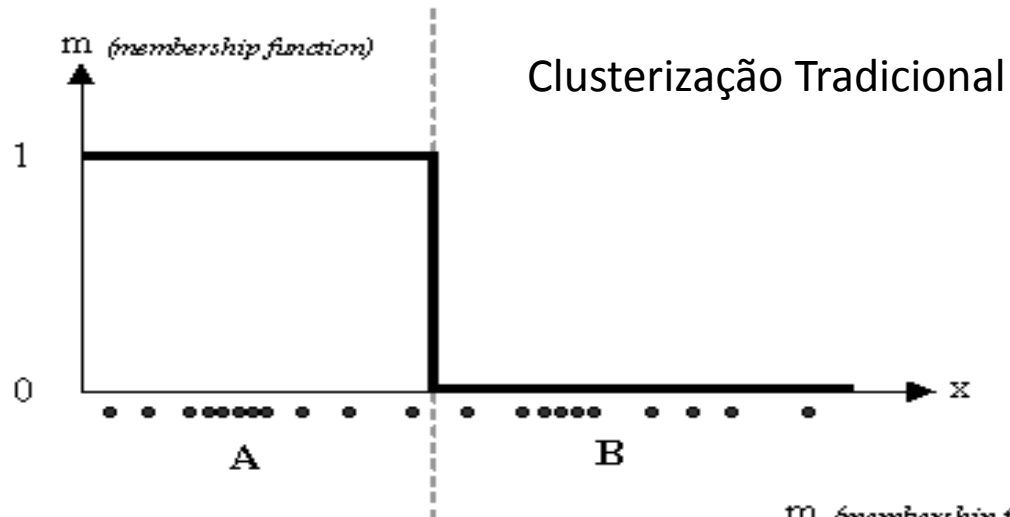
# Definições

- Clusterização está relacionada com o particionamento de um conjunto de dados em grupos homogêneos.
- A clusterização tradicional também chamada de Clusterização Direta utiliza limiares rígidos na separação dos clusters.
- Clusterização Fuzzy usa a lógica fuzzy para criar grupos sobrepostos de dados.

# Clusterização Fuzzy

- Os dados podem pertencer a mais de 1 cluster ao mesmo tempo com graus de possibilidade ou valores da função de pertinência diferentes.
- A Função de Pertinência de um cluster varia de 1 a 0. Na Clusterização Direta os valores podem ser 0 ou 1, somente.

# Clusterização Fuzzy



# Algoritmo Fuzzy C-Médias (FCM)

- O algoritmo mais popular para Clusterização Fuzzy. Introduzido e desenvolvido por Dunn e melhorado por Bezdek.
- Baseado na minimização de uma função objetivo.
- Essa função indica a soma das distâncias de cada centro de cluster para os pontos de dados naquele cluster. Dessa forma, com a minimização dessas distâncias, quanto menor o valor de  $J_m$ , melhor a clusterização.

# Algoritmo Fuzzy C-Médias (FCM)

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2$$

Onde,

$m$  é o grau de fuzzyficação de qualquer número real maior que 1;

$u_{ij}$  é o grau de pertinência de  $x_i$  no cluster  $j$ ;

$c_j$  é o centro do cluster  $j$ ;

$N$  é o número de dados;

$x_i$  é o valor  $d$ -dimensional medido;

$\|\cdot\|^2$  é uma norma expressando a similaridade entre o dado mensurado e o centro.



# Algoritmo Fuzzy C-Médias (FCM)

1- Initialise  $U = [u_{ij}]$

Where:

2- At k-step: calculate the centres vectors

$K$  is the iteration step

$$C^{(k)} = [c_j] \text{ with } U^{(k)}$$

$\sigma$  is a termination criterion between 0 and 1

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m \cdot x_i}{\sum_{i=1}^N u_{ij}^m} \quad \left. \vphantom{\sum_{i=1}^N} \right\} \text{ Média ponderada com todos os pontos}$$

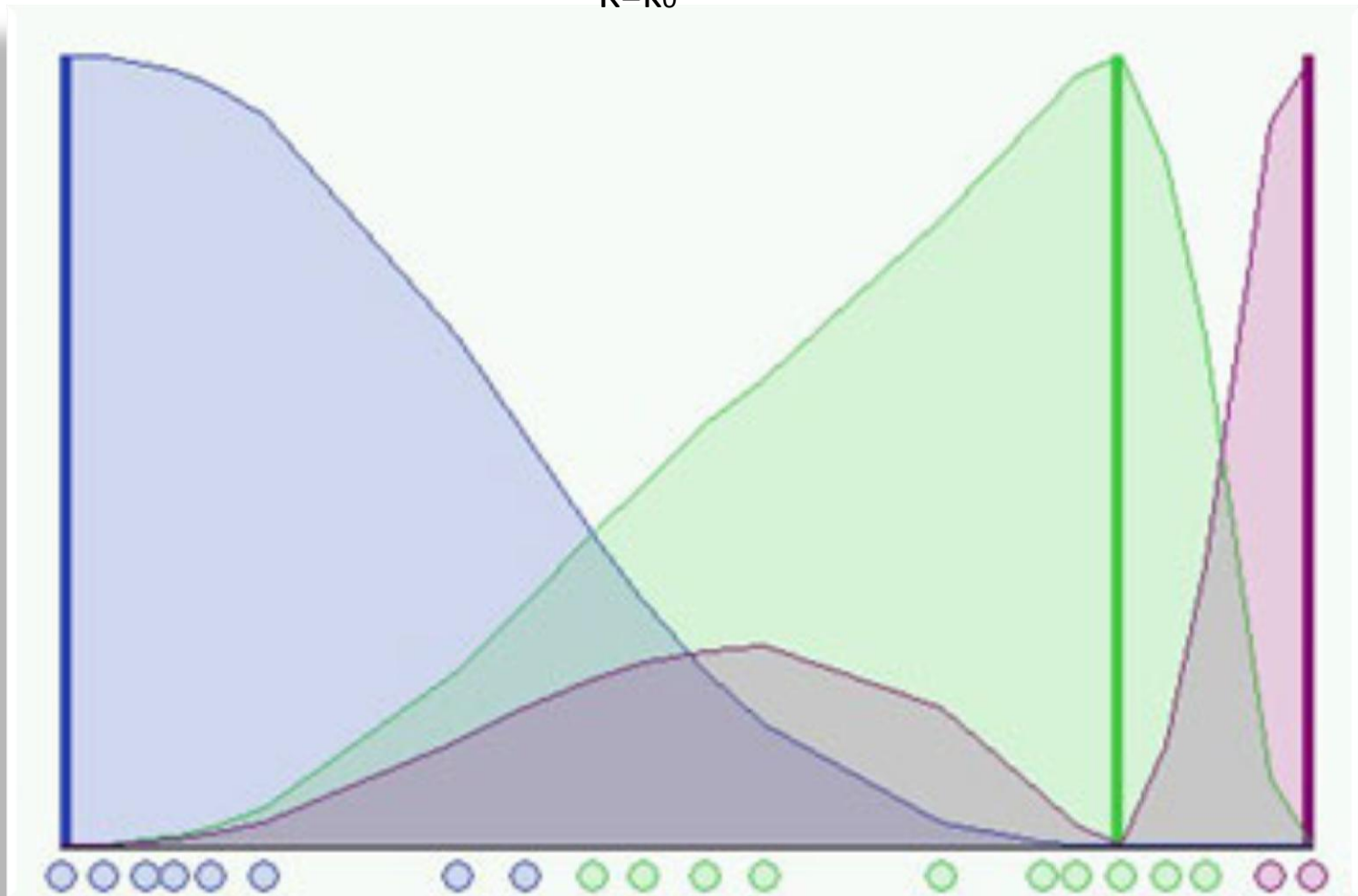
3- Update  $U^{(k)}, U^{(k+1)}$

$$U_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad \left. \vphantom{\sum_{k=1}^c} \right\} \text{ Limites: [0..1]}$$

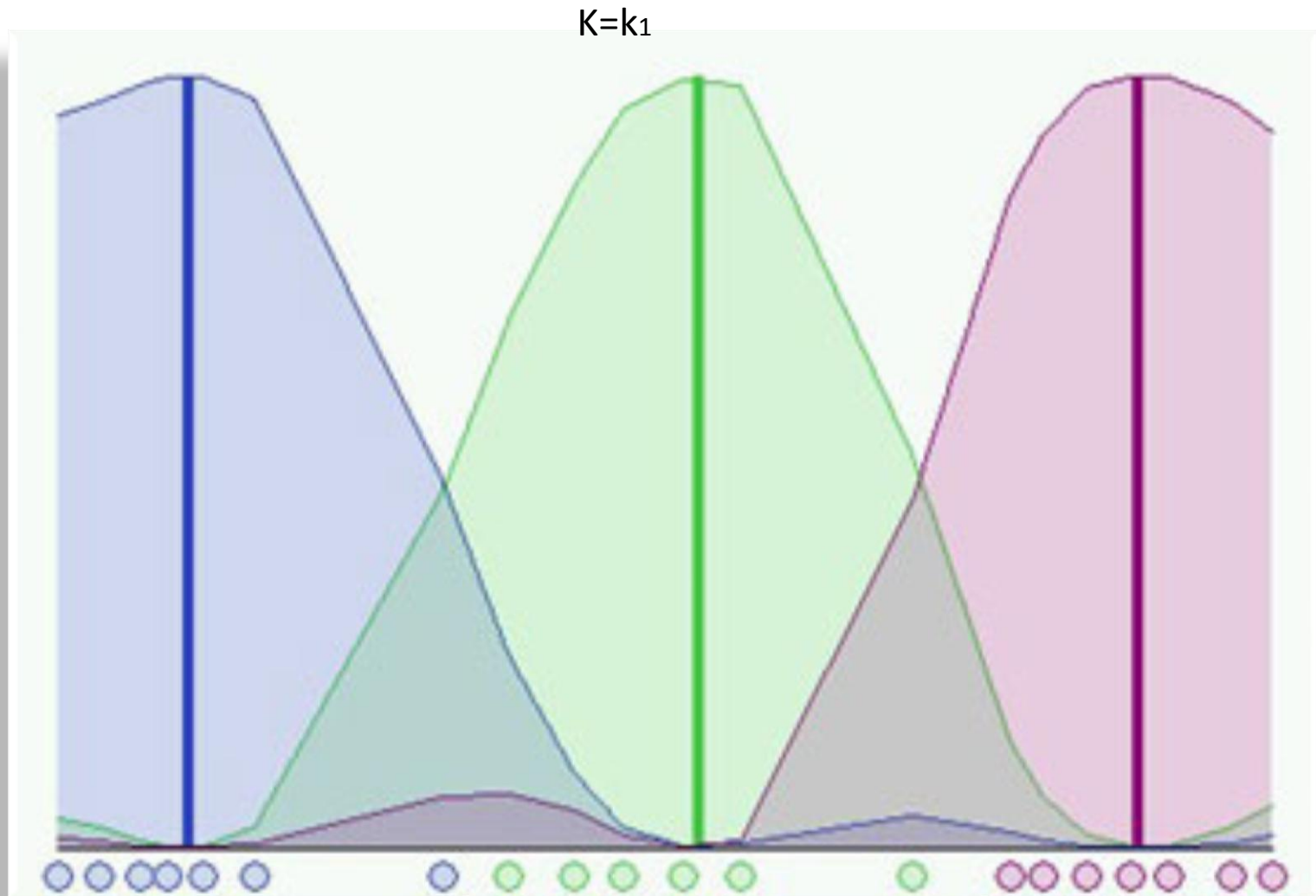
4- If  $\|U^{(k+1)} - U^{(k)}\| < \sigma$  then stop,  
otherwise go to step 2.

# Algoritmo Fuzzy C-Médias (FCM)

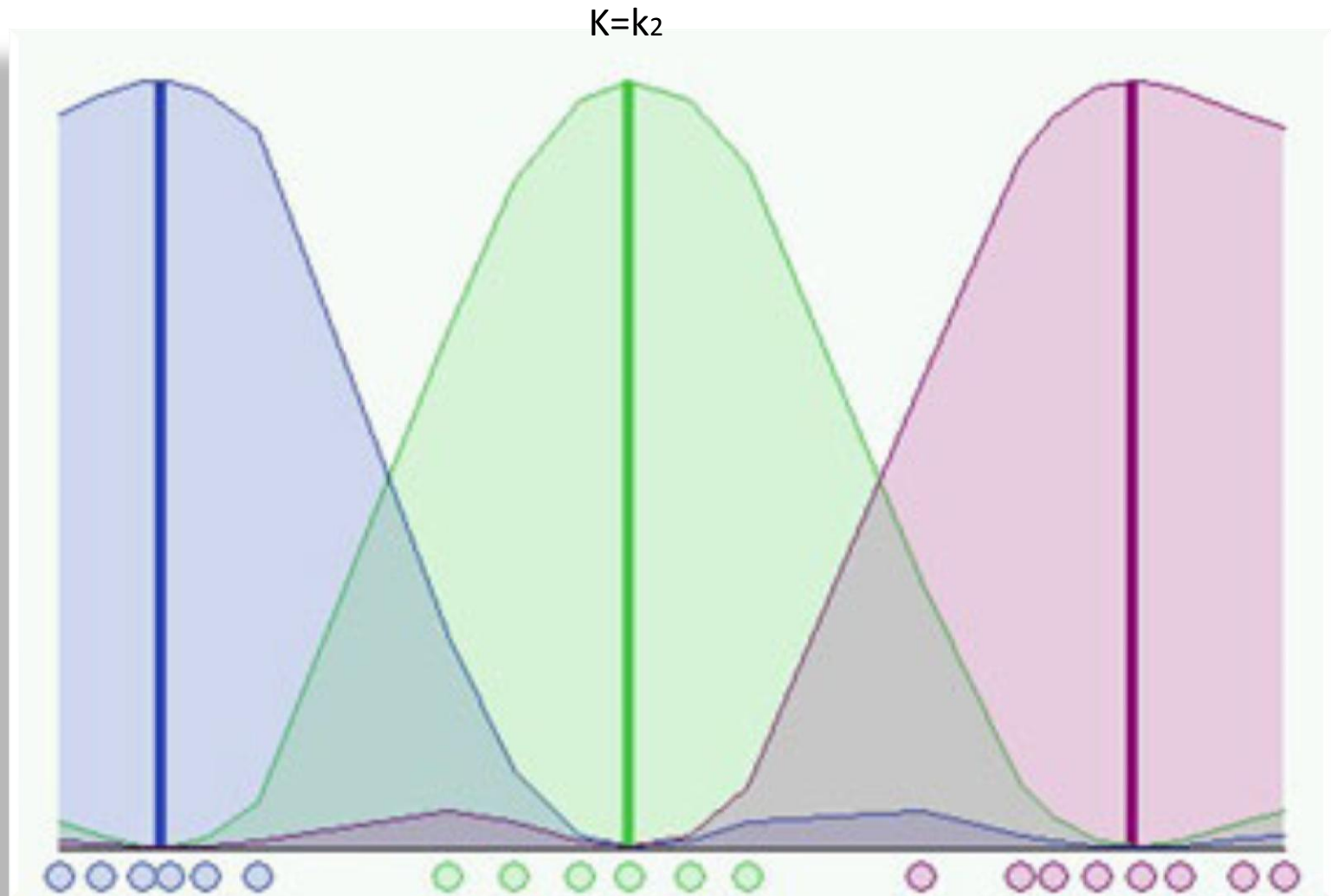
$K=k_0$



# Algoritmo Fuzzy C-Médias (FCM)



# Algoritmo Fuzzy C-Médias (FCM)



# Conclusões do FCM

- O K-Médias é o mais famoso por conta da sua simplicidade e sua baixa complexidade comparada com outros algoritmos.
- Tanto o FCM quanto o K-Médias tendem a convergir para um mínimo local.
- O algoritmo das abelhas é combinado para resolver esse problema da convergência em mínimos locais.

# Abelhas na Natureza

- Baseado no comportamento de procura por alimento das abelhas na natureza.
- A busca têm os seguintes elementos:
  - Scout Bees: Responsável pela busca de caminho de flores promissoras;
  - Waggle Dance: Dança para comunicação entre as abelhas;
  - Dancer: Scout Bee que percorreu um caminho e fornece informações para a colméia;
  - Followers Bees: Abelhas que seguem a scout bee após a dança para coletar o alimento.

# Abelhas na Natureza

- Caminhos com maior quantia de néctar ou pólen que podem ser coletados com menor esforço tendem a ser visitadas por mais abelhas.
- A dança da abelha informa 3 informações: a direção, a distância da colméia e a pontuação de qualidade do caminho. A avaliação dos diferentes caminhos observam a qualidade do alimento que ele fornece e a quantidade de energia necessária para buscar o alimento.

# Algoritmo das Abelhas

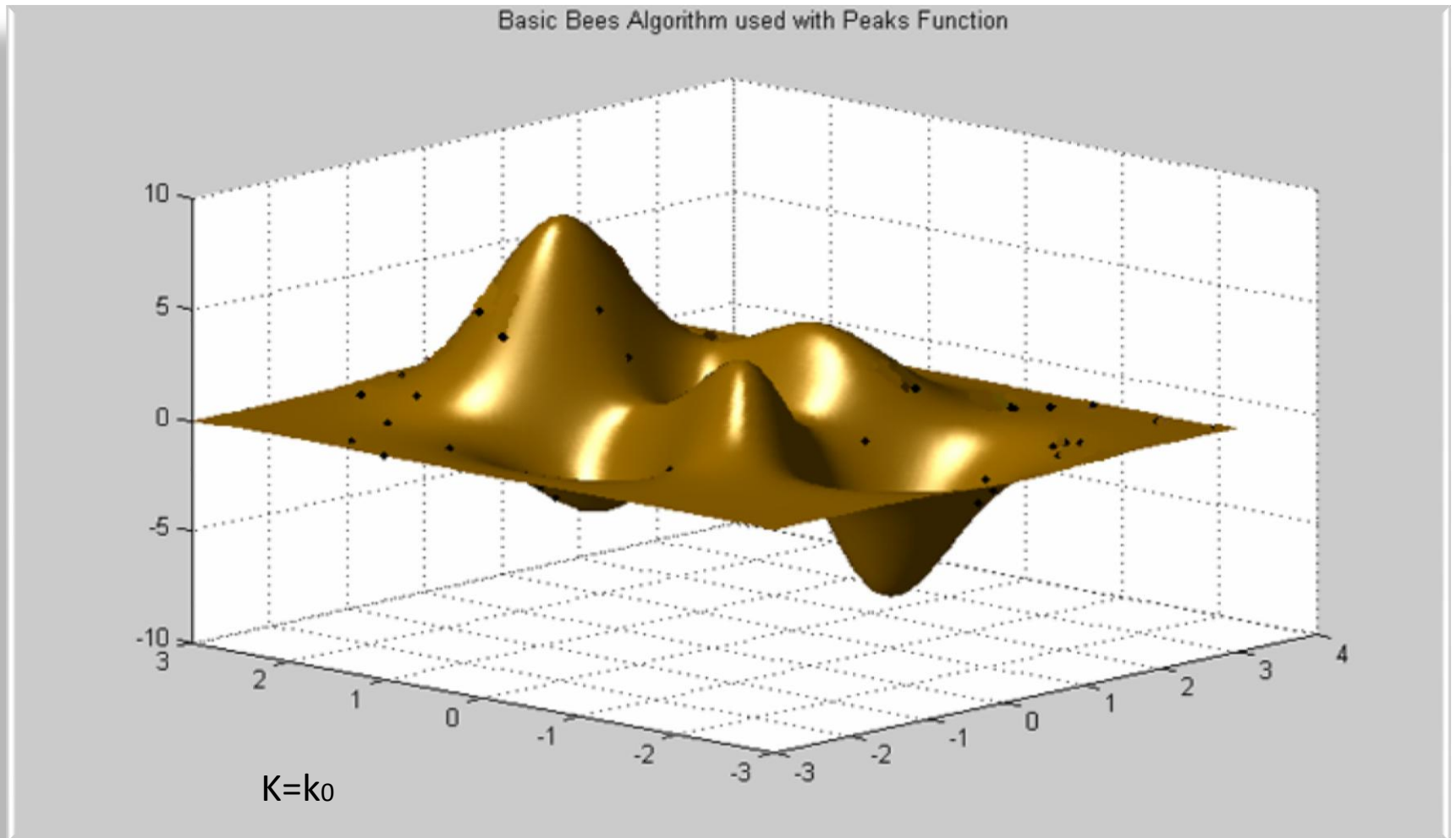
- O algoritmo requer um número de parâmetros a ser configurados como:
    - (n) Número de Scout Bees
    - (m) Número de regiões selecionadas para busca na vizinhança
    - (e) Número de regiões elite fora de m
    - (nep) Número de abelhas recrutadas para as regiões elite
    - (m-e) Número de abelhas recrutadas para as outras regiões
    - (nsp) Lugares selecionados
- Critério de Parada.



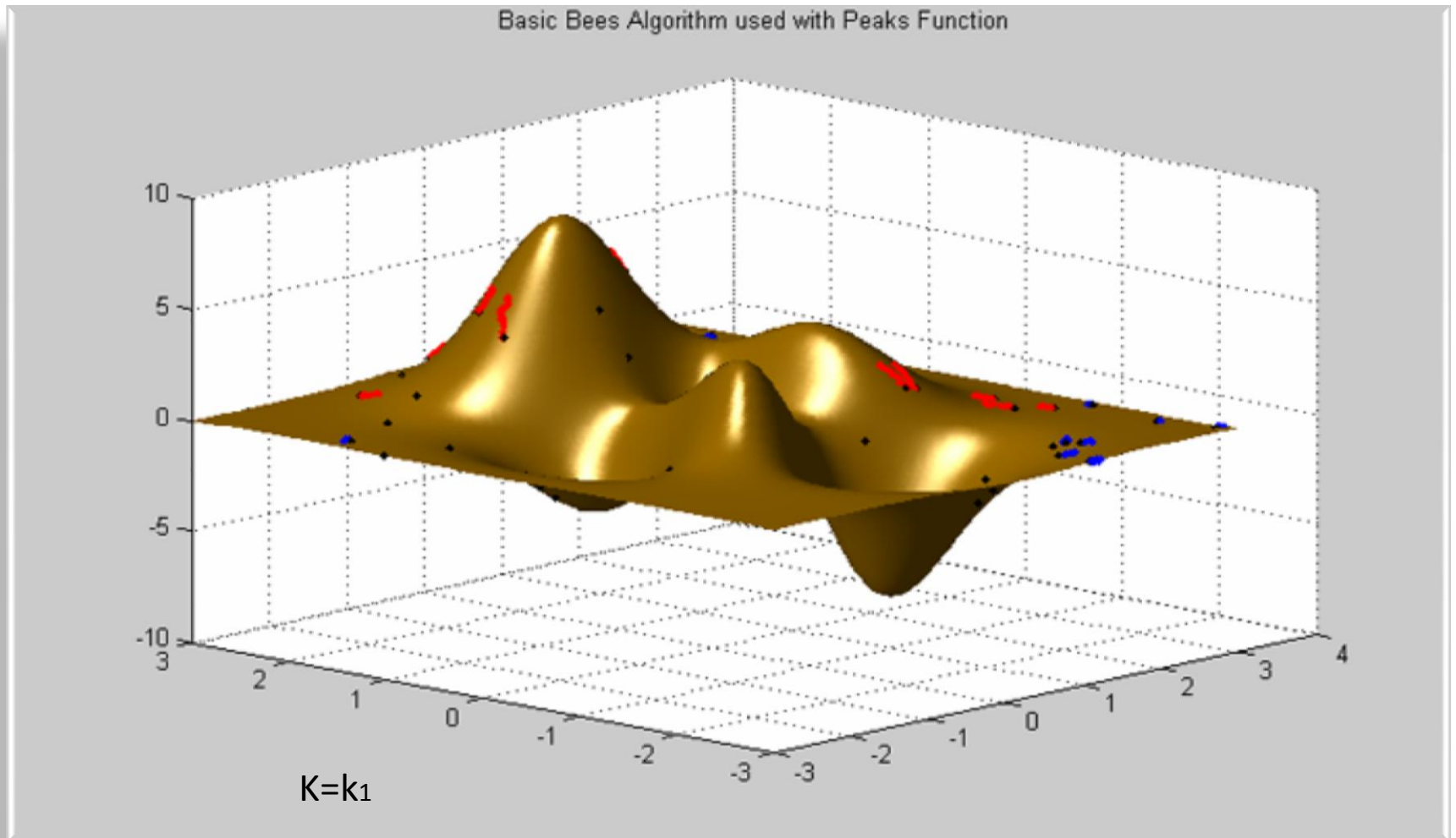
# Algoritmo das Abelhas

1. Initialise the solution population (each initial solution, or 'bee', being a set of randomly placed cluster centres).
2. Evaluate the fitness of the population.
3. While (stopping criterion is not met)  
    // forming new population.
4. Select sites for neighbourhood search.
5. Recruit bees for selected sites (more bees for the best e sites) and evaluate fitnesses.
6. Select the fittest bee from each site.
7. Assign remaining bees to search randomly and evaluate their fitnesses.
8. End While.

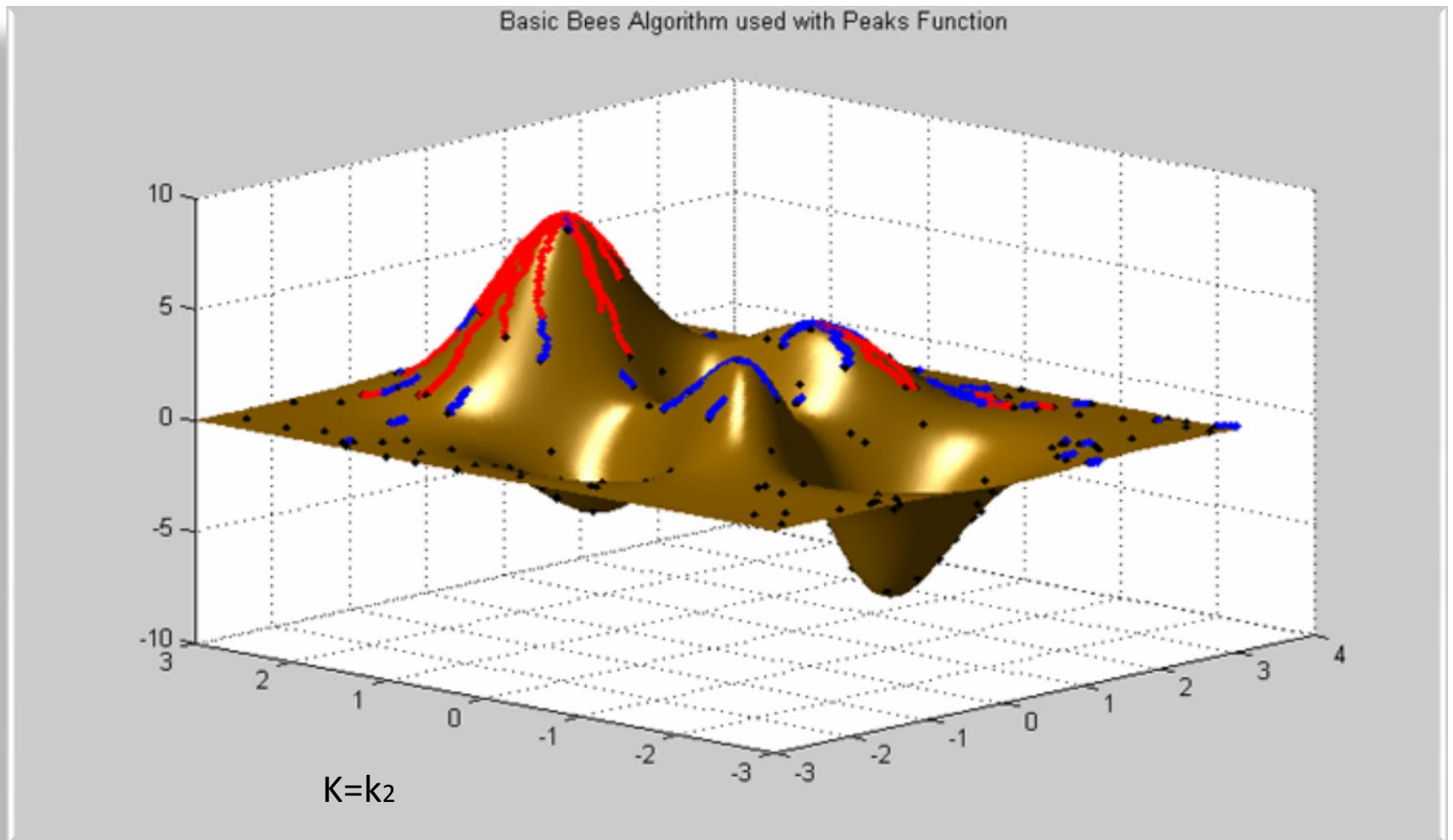
# Algoritmo das Abelhas



# Algoritmo das Abelhas



# Algoritmo das Abelhas



# Conclusão do Algoritmo das Abelhas

- Nós utilizamos o algoritmo das abelhas para encontrar o valor mínimo para  $J_m$ .

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2$$

- O algoritmo das abelhas é capaz de localizar eficientemente soluções próximas da ótima.
- Uma colônia de abelhas pode se espalhar por longas distâncias com o objetivo de *exploitation* (busca em largura) por alimento, ao mesmo tempo.

# Algoritmo Proposto – Bees Algorithm with FCM

1. Read data from the file.
2. Initialise the solution population.  $\longrightarrow$  Bees (1)
3. Fill array  $A$  using formula  $u_{ij}$  and evaluate the fitness of the population using Eq. (1).  $\longrightarrow J_m$
4. While (stopping criterion is not met)
5. Select sites for neighbourhood search.
6. Recruit bees for selected sites (more bees for the best  $e$  sites) and fill array  $A$  using formula  $u_{ij}$  and evaluate the fitness using Eq. (1).  $\longrightarrow J_m$
7. Select the fittest bee from each site.
8. Assign remaining bees to search randomly then fill array  $A$  using formula  $u_{ij}$  and evaluate their fitnesses using Eq. (1).  $\longrightarrow J_m$
9. End While.

$$U_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

# Algoritmo Proposto – Bees Algorithm with FCM

- O array  $A$  contém o valor da função de pertinência para todos os dados.
- A cada iteração a população de abelhas na busca vai aumentando.
- Esse algoritmo otimizado e simples não diminui o custo computacional e a complexidade do FCM, mas evita a queda em mínimos locais, através da busca randômica.

# Resultados Experimentais

Algorithm	Parameters	Value
FCM	Maximum number of iterations	1000
	Crossover probability, $\mu_c$	0.8
GA	Mutation probability, $\mu_m$	0.001
	Population size, P	100
Bees Algorithm	Number of scout bees	21
	Number of sites selected for neighbourhood search	8
	Number of elite bees	2
	Number of bees recruited for the elite sites	5
	Number of bees recruited for the other selected sites	2
	Number of iterations	300



# Resultados Experimentais

- Os algoritmos foram executados 10 vezes, os valores mínimo, máximo e médio de  $J_m$  foram coletados.

Da ta set	Algori thm	Mean	Min	Max
<i>Iris</i>	FCM	61.72	60.58	65.81
	GA	65.54	63.34	69.23
	Bees	<b>60.58</b>	<b>60.58</b>	<b>60.58</b>
<i>Charts</i>	FCM	<b>530.99</b>	<b>530.99</b>	<b>530.99</b>
	GA	802.08	791.23	819.28
	Bees	549.08	546.87	552.15
<i>Wood Defects</i>	FCM	64077395.11	59195094.2 9	66225010.4 9
	GA	168035.20	157508.00	174784.00
	Bees	<b>165132.52</b>	<b>153866.53</b>	<b>173523.00</b>
<i>Crude Oil</i>	FCM	1330.43	1242.30	1560.57
	GA	1252.61	1238.25	1286.04
	Bees	<b>1239.50</b>	<b>1237.24</b>	<b>1241.56</b>

# Resultados Experimentais

Table 3

Data sets used in the experiments

<b>Data Set</b>	<b># of Object</b>	<b># of Feature</b>	<b># of Class</b>
<i>Iris</i>	150	4	3
<i>Control Charts</i>	1500	60	6
<i>Wood Defects</i>	232	17	13
<i>Crude oil</i>	59	5	3

# Conclusão

- Combinando o algoritmo das abelhas com o FCM foi melhorado o resultado da clusterização fuzzy comparado com o FCM na maioria dos casos e com o GA combinado ao FCM.
- Um dos principais conceitos sobre o algoritmo proposto é que ele requer um longo tempo computacional, assim como o FCM.

# FIM

Dúvidas?

Obrigado.