

Aplicando Ant Colony Optimization y Bee Colony Optimization para Clustering

Adán González Rodríguez
Sara Porto Álvarez

Dataset Utilizado

Hemos decidido usar el Dataset: [Mall Customer Segmentation](#)

El dataset está diseñado para enseñar conceptos relacionados con la segmentación de clientes, también conocida como análisis de canasta de mercado (market basket analysis).

Lo usaremos para hacer un clustering sobre los atributos del mismo, intentando clasificar por grupos de consumo entre los clientes.

Atributos del dataset y cuáles utilizamos

- Gender
- Age
- **Annual Income (k\$)**
- **Spending Score (0-100)**

Ant Colony Optimization

Código: [Link al código de ACO](#)

- **Imports** de las librerías necesarias
- Clase **AntColonyClustering**
 - Método `__init__`
 - Método `run()`
 - Método `_construct_path()`
 - Método `_select_next_node(current_node, unvisited)`
 - Método `_update_pheromones(all_paths)`
 - Método `_form_clusters()`
 - Método `assign_test_data(test_data, train_data):`
 - Método `evaluate_test_data(test_labels, true_test_labels)`
 - Método `plot_clusters(train_data, test_data=None, test_labels=None)`
- Método **main**

Bee Colony Optimization

Código: [Link al código de BCO](#)

- **Imports** de las librerías necesarias
- Clase **BeeColonyOptimization**
 - Método `__init__`
 - Método `fitness(self, centroids)`
 - Método `_bound_solution(solution)`
 - Método `_local_search(current_position)`
 - Método `update_positions()`
 - Método `_finalize_centroids(top_positions)`
 - Método `run()`
 - Método `assign_test_data(test_data)`
 - Método `plot_clusters(train_data)`
- Método **main**

Comparación entre algoritmos ACO y BCO

Métricas de evaluación utilizadas

- Índice de Silhouette

- Evalúa qué tan bien separados están los clústeres y cuán compactos son
- Buen clustering → coeficiente **mayor de 0.5**

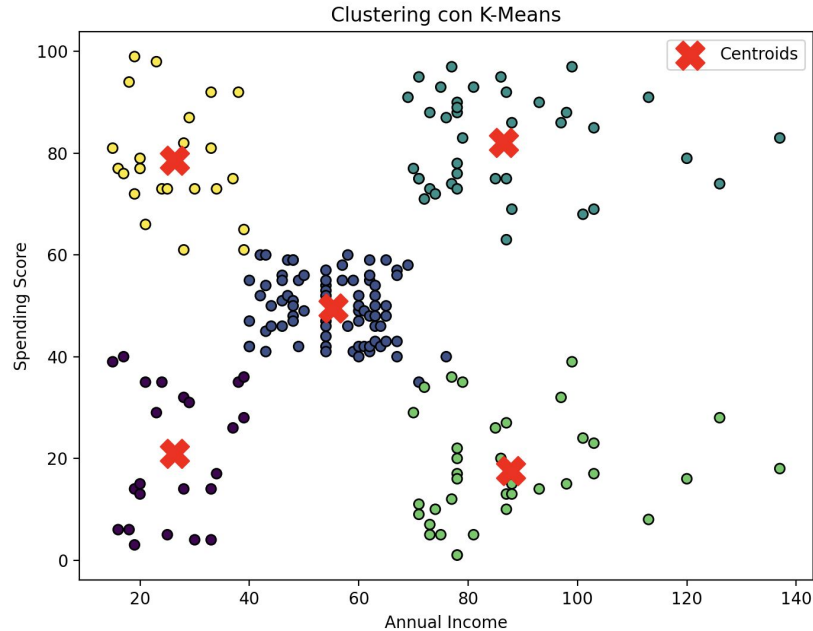
- Índice de Davis-Bouldin

- Mide la compacidad y separación entre clústeres
- Mejor cuanto más **cercano a 0** sea (valores altos indican un mal clustering)

- Índice de Calinski-Harabaz

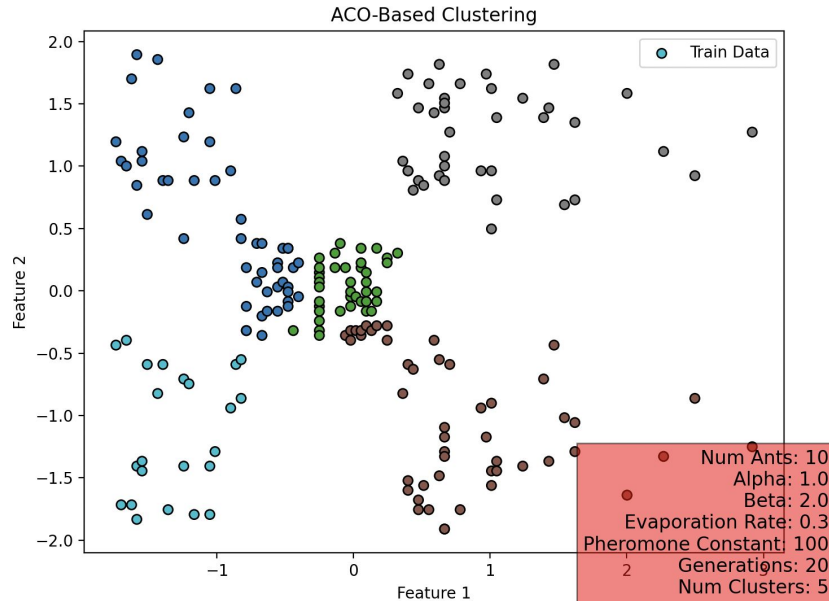
- Mide la dispersión entre y dentro de los clústeres
- **Valor alto** indica buen clustering

K-means: [Link al Código de K-Means](#)



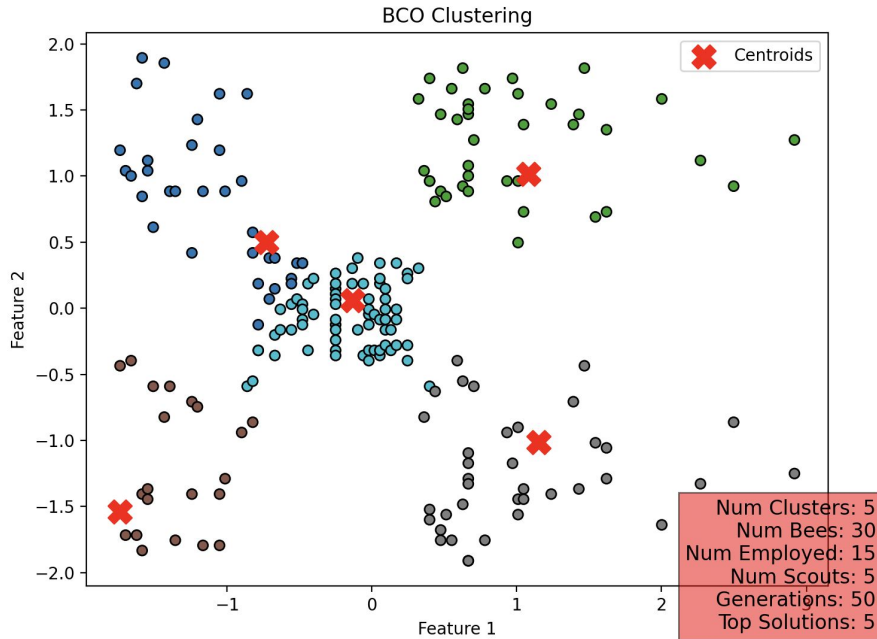
Silhouette Score: 0.44473703994455477
Davies-Bouldin Score: 0.82163717258315
Calinski-Harabasz Score: 151.133658537184

Evaluación ACO: [Link a más pruebas con ACO](#)



Silhouette Score: 0.3386
Davies-Bouldin Score: 0.7935
Calinski-Harabasz Score: 160.1774

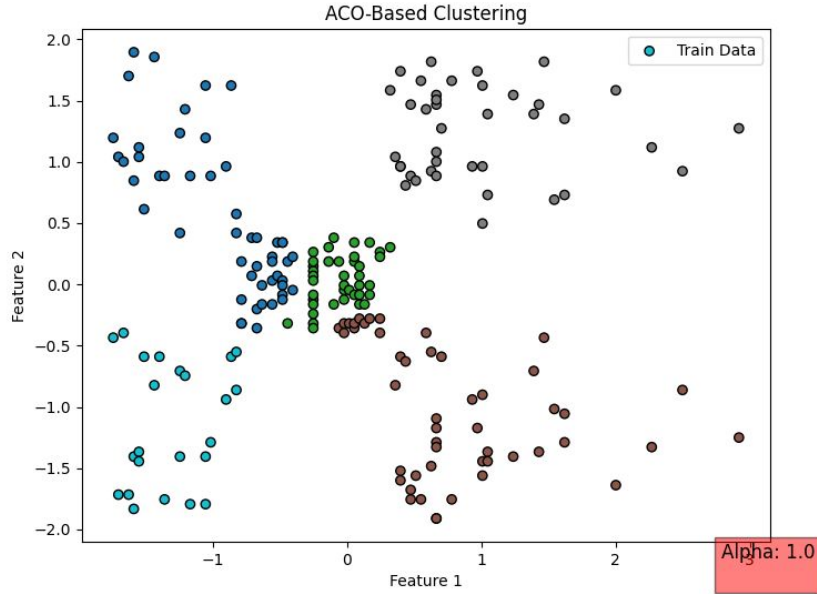
Evaluación BCO: [Link a más pruebas con BCO](#)



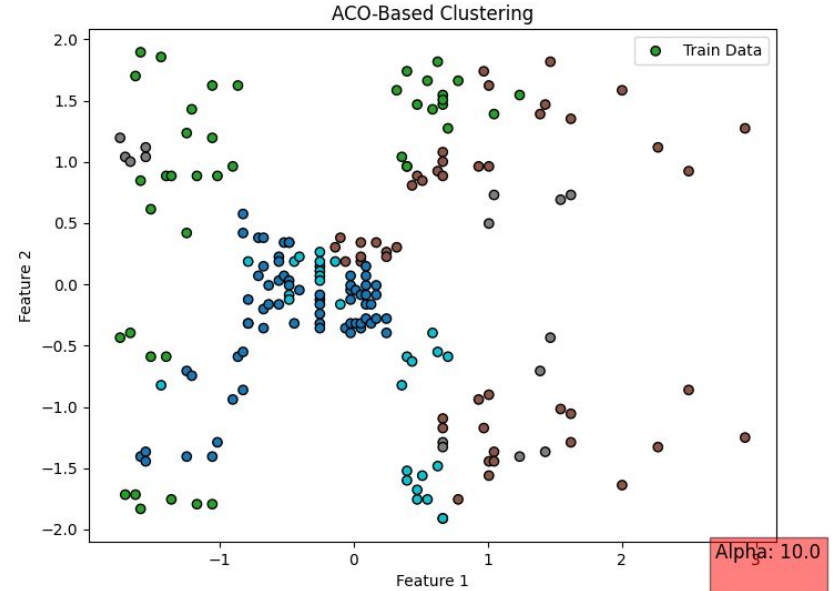
Silhouette Score: 0.4904
Davies-Bouldin Score: 0.6542
Calinski-Harabasz Score: 220.9043

Modificación de Hiperparámetros en el algoritmo ACO

Alpha (Importancia de las Feromonas)

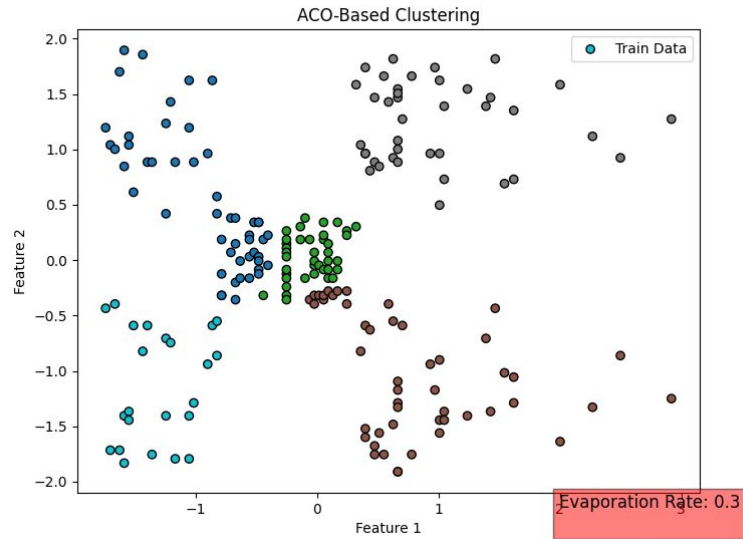


Alpha = 1 (Favorece la exploración)

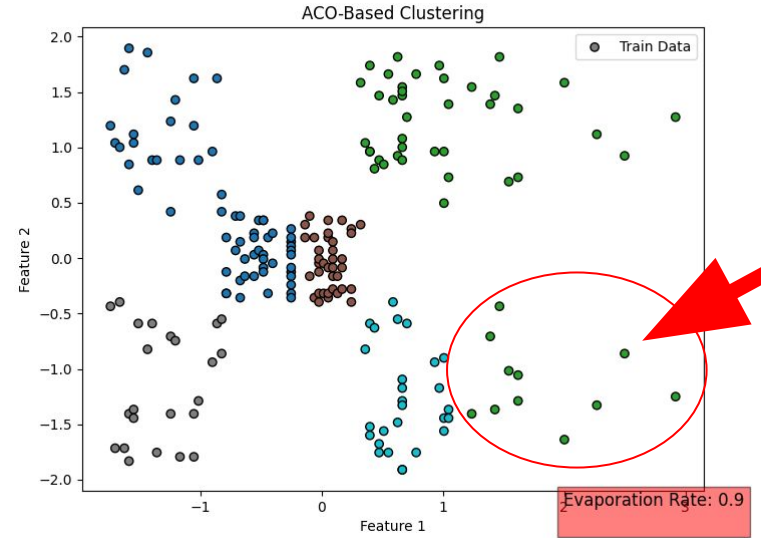


Alpha = 10 (Favorece las Feromonas)

Evaporation Rate



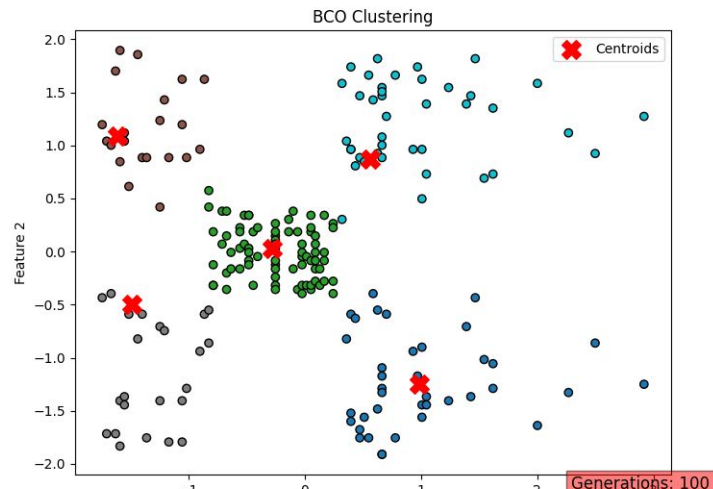
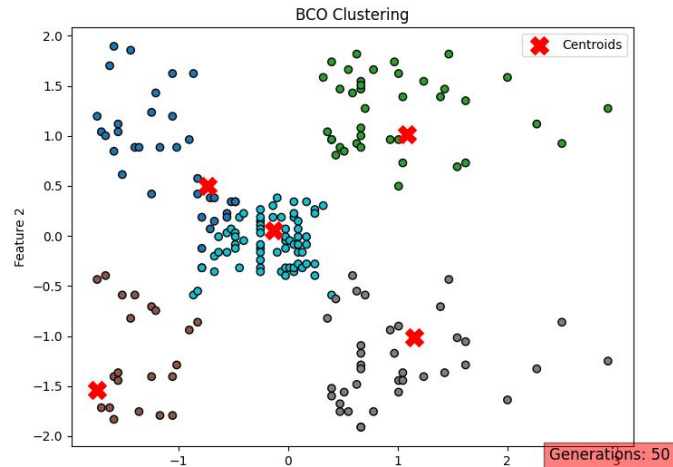
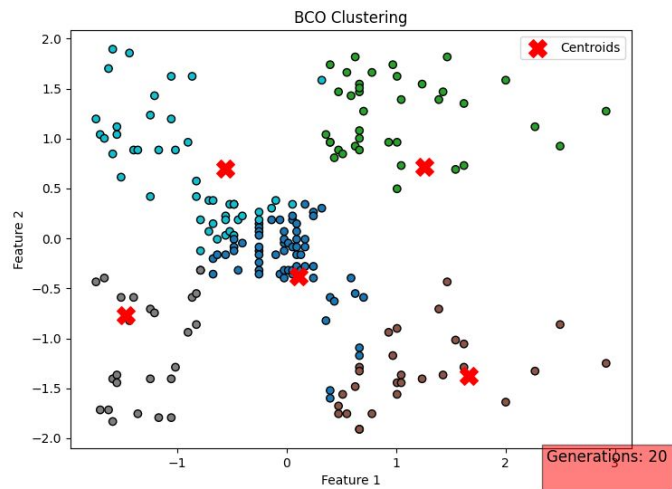
Ev_Rate = 0.3 (Favorece las rutas conocidas)



Ev_Rate = 0.9 (Evita sobreajuste)

Modificación de Hiperparámetros en el algoritmo BCO

Generations



Comparativa final

Comparación

- **¿Los algoritmos identifican correctamente los grupos esperados?**
 - Sí. Ambos tienen un buen desempeño a la hora de identificar los grupos esperados
 - Aún así, los clusters no son exactamente iguales a los identificados por K-Means
- **¿Cuál parece ser más eficiente o robusto?**
 - El BCO presenta mejores resultados
- **¿Existen diferencias en la distribución de los clusters?**
 - Sí, tal y como se vio en las imágenes (aunque no son grandes diferencias)

Ventajas y Desventajas de cada algoritmo

Ant Colony Optimization (ACO)

Buena exploración del espacio de búsqueda

Manejo efectivo de soluciones globales y locales

Puede adaptarse bien a cambios dinámicos

Alto costo computacional en problemas grandes

Requiere ajuste fino de parámetros

Bee Colony Optimization (BCO)

Rápida convergencia hacia soluciones óptimas

Flexibilidad en la asignación de recursos

Balance entre exploración y explotación

Puede quedar atrapado en óptimos locales

Sensible a la elección de parámetros

Fin