



ugr | Universidad
de **Granada**

TRABAJO FIN DE GRADO
INGENIERÍA INFORMÁTICA

Estudio y Análisis de Metaheurísticas modernas para el problema de Selección de Características

Autor
Miguel García López

Directores
Daniel Molina Cabrera



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

—
Granada, Enero de 2024



Título del proyecto

Subtítulo del proyecto.

Autor

Nombre Apellido1 Apellido2 (alumno)

Directores

Nombre Apellido1 Apellido2 (tutor1)

Nombre Apellido1 Apellido2 (tutor2)

Estudio y Análisis de Metaheurísticas modernas para el problema de Selección de Características

MIGUEL GARCÍA LÓPEZ

Palabras clave: Metaheurística, selección, de características, binario, optimización, población, fitness, aprendizaje automático

Resumen

En el ámbito del *Machine Learning*, algunos algoritmos, como KNN o SVM, muestran excelentes resultados, pero a menudo requieren un procesamiento previo para identificar las características más relevantes. Esto se debe a que, ya sea por la amplitud y complejidad del problema, por falta de conocimiento o contexto al recolectar los datos, es posible aglomerar información innecesaria, dando lugar a conjuntos de datos inmensos. Incluso los algoritmos que internamente realizan esta identificación pueden beneficiarse de un procesamiento adicional. Este preprocesamiento, conocido como selección de características (*feature selection*), se considera un problema complejo de optimización combinatoria.

Existen varios métodos comunes para la selección de características:

- **Filtrado:** Evalúa características individualmente según su relación con la variable objetivo, usando métricas como correlación o chi-cuadrado.
- **Envoltura (*Wrapper*):** Evalúa conjuntos de características mediante el entrenamiento y la evaluación de un modelo, utilizando métodos como búsqueda hacia adelante o hacia atrás.
- **Métodos basados en árboles:** Algoritmos como *Random Forests* proporcionan una estimación de la importancia de cada característica.
- **Análisis de Componentes Principales (PCA):** Técnica de reducción de dimensionalidad que transforma las características originales en componentes principales.
- **Selección L1 (*Lasso Regression*):** Utiliza una penalización L1 en la regresión lineal, eliminando características menos importantes.

Estos métodos pueden utilizarse individualmente o en combinación para mejorar la selección de características y el rendimiento de los modelos de *Machine Learning*. En este documento se hablarán de métodos metaheurísticos para la resolución de este problema.

Las metaheurísticas son algoritmos diseñados para resolver problemas de optimización complejos cuando los recursos son limitados. Aunque inicialmente se desarrollaron para abordar principalmente problemas combinatorios, en la actualidad se están proponiendo y aplicando cada vez más en problemas que implican variables continuas o reales. En respuesta a la creciente demanda en el campo de la selección de características, se han adaptado versiones especializadas de estas metaheurísticas para abordar este tipo de problemas combinatorios. Sin embargo, a pesar del número creciente de propuestas en este campo, las comparaciones objetivas entre ellas son limitadas. Aunque existen revisiones bibliográficas, muchas de ellas carecen de comparaciones adecuadas debido a la importancia y actualidad del problema de selección de características. Por lo tanto, hay una necesidad de estudios que proporcionen una evaluación comparativa más rigurosa y exhaustiva de las diferentes propuestas en este ámbito.

En este trabajo de carácter científico, se llevará a cabo una revisión bibliográfica de diversas metaheurísticas recientes para abordar el problema de selección de características. Se estudiarán e implementarán aquellas consideradas más prometedoras, con el objetivo de construir un repertorio amplio y variado de propuestas. Posteriormente, se realizará un estudio comparativo exhaustivo utilizando diversos algoritmos de machine learning y conjuntos de datos representativos. Finalmente, se llevará a cabo un análisis crítico utilizando diversas métricas y valoraciones, como la tasa de acierto y el tiempo de ejecución, entre otras.

Study and Analysis of Modern Metaheuristics for the Feature Selection Problem

MIGUEL GARCÍA LÓPEZ

Keywords: Metaheuristic, feature selection, binary, optimization, population, fitness, machine learning

Abstract

In the field of Machine Learning, some algorithms, such as KNN or SVM, show excellent results, but often require preprocessing to identify the most relevant features. This is because, whether due to the broad and complex nature of the problem, lack of knowledge, or lack of context when collecting data, it is possible to aggregate unnecessary information, resulting in immense datasets. Even algorithms that internally perform this identification can benefit from additional processing. This preprocessing, known as feature selection, is considered a complex combinatorial optimization problem.

There are several common methods for feature selection:

- **Filtering:** Evaluates features individually based on their relationship with the target variable, using metrics such as correlation or chi-square.
- **Wrapper:** Evaluates sets of features by training and evaluating a model, using methods such as forward or backward search.
- **Tree-based methods:** Algorithms like Random Forests provide an estimate of the importance of each feature.
- **Principal Component Analysis (PCA):** A dimensionality reduction technique that transforms the original features into principal components.
- **L1 Selection (Lasso Regression):** Uses an L1 penalty in linear regression, eliminating less important features.

These methods can be used individually or in combination to improve feature selection and the performance of *Machine Learning* models. This document will discuss metaheuristic methods for solving this problem.

Metaheuristics are algorithms designed to solve complex optimization problems when resources are limited. Initially developed mainly for combinatorial problems, they are increasingly proposed and applied to problems involving continuous or real variables. In response to the growing demand in feature

selection, specialized versions of these metaheuristics have been adapted to tackle combinatorial problems. However, despite the increasing number of proposals in this field, objective comparisons between them are limited. Although there are literature reviews, many lack adequate comparisons due to the importance and timeliness of the feature selection problem. Therefore, there is a need for studies providing a more rigorous and exhaustive comparative evaluation of different proposals in this area.

In this scientific work, a literature review of various recent metaheuristics for feature selection will be conducted. The most promising ones will be studied and implemented to build a broad and varied repertoire of proposals. Subsequently, a comprehensive comparative study will be conducted using various machine learning algorithms and representative datasets. Finally, a critical analysis will be carried out using various metrics and assessments, such as accuracy rate and execution time, among others.

Yo, **Miguel García López**, alumno de la titulación Ingeniería Informática de la **Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada**, con DNI 77159865E, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Grado en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

Fdo: Miguel García López

Granada a 29 de Enero de 2024.

D. **Daniel Molina Cabrera**, Profesor del Departamento Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada.

Informan:

Que el presente trabajo, titulado *Estudio y Análisis de Metaheurísticas modernas para el problema de Selección de Características*, ha sido realizado bajo su supervisión por **Miguel García López**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a 29 de Enero de 2024.

El director:

Daniel Molina Cabrera

Agradecimientos

Poner aquí agradecimientos...

Índice general

1. Introducción	7
1.1. Motivación	7
1.2. Objetivos	8
2. Planificación	11
3. Fundamentos Teóricos	17
3.1. Optimización	17
3.1.1. Definición general de un problema de optimización . .	17
3.1.2. Función objetivo y función fitness	18
3.1.3. Óptimos globales o locales	18
3.2. Selección de características	19
3.2.1. Necesidad y motivo	19
3.2.2. Concepto	21
3.2.3. Maldición de la dimensionalidad	22
3.3. Metaheurísticas	22
3.3.1. Exploración vs explotación	23
3.4. Teorema No Free Lunch	24
3.5. Aprendizaje automático	25
3.5.1. Aprendizaje supervisado	25
3.5.2. Aprendizaje no supervisado	26
3.5.3. SVM	26
3.5.4. K-nearest neighbors	27
4. Revisión de la literatura	29
4.1. Introducción a los primeros algoritmos poblacionales	32
4.2. Propuestas metaheurísticas modernas	33
4.3. Versiones binarias	36
5. Descripción de los algoritmos	39
5.1. Algoritmos genéticos	39
5.1.1. Introducción	39
5.1.2. Funcionamiento y Operadores	40
5.2. Algoritmos de colonias de hormigas	42

5.2.1. Introducción	42
5.2.2. Funcionamiento y Operadores	43
5.3. Optimización por enjambre de partículas	44
5.3.1. Introducción	44
5.3.2. Funcionamiento y Operadores	44
5.4. Evolución diferencial	45
5.4.1. Funcionamiento y Operadores	46
5.5. Optimización por colonia de abejas artificial	48
5.5.1. Introducción	48
5.5.2. Funcionamiento y Operadores	49
5.6. Algoritmo de optimización del saltamontes	49
5.6.1. Introducción	49
5.6.2. Funcionamiento y Operadores	50
5.7. Algoritmo de la libélula	51
5.7.1. Introducción	51
5.7.2. Funcionamiento y Operadores	52
5.8. Algoritmo de optimización de ballenas	54
5.8.1. Introducción	54
5.8.2. Funcionamiento y Operadores	54
5.9. Algoritmo del murciélagos	56
5.9.1. Introducción	56
5.9.2. Funcionamiento y Operadores	57
5.10. Optimización de los lobos grises	58
5.10.1. Introducción	58
5.10.2. Funcionamiento y Operadores	59
5.11. Algoritmo de la luciérnaga	60
5.11.1. Introducción	60
5.11.2. Funcionamiento y Operadores	61
5.12. Búsqueda Cuckoo	62
5.12.1. Introducción	62
5.12.2. Funcionamiento y Operadores	62
6. Diseño experimental	65
7. Análisis y resultados	71
7.1. Binario	72
7.1.1. General	72
7.1.2. Fitness	74
7.1.3. Clasificación y reducción	88
8. Bibliografía	97

Índice de figuras

2.1. Diagrama de Gantt inicial	13
2.2. Diagrama de Gantt final	14
3.1. Puntos globales y locales	19
3.2. Correlación entre error del modelo y N	20
3.3. Tendencia de incremento de dimensionalidad en varios espacios de características	22
3.4. Clasificación metaheurísticas	23
3.5. Aprendizaje supervisado	25
3.6. Aprendizaje no supervisado	26
3.7. Margen en SVM	27
4.1. Popularidad de feature selection sobre los años	30
4.2. Popularidad de feature selection + metaheuristics sobre los años	31
4.3. PBT en Redes Neuronales	34
4.4. Popularidad metaheurísticas modernas	35
4.5. Funciones de transferencia	37
5.1. Funcionamiento de un algoritmo genético	40
5.2. One point crossover	41
5.3. Blend crossover	41
5.4. Mutación en DE	47
5.5. Crossover en DE	48
5.6. Convergencia en GOA	51
5.7. Vecindario de libélulas	53
5.8. Mecanismo de encogimiento del cerco	55
5.9. Mecanismo de actualización de posición en espiral	55
5.10. Rodeo de la presa en 2D y 3D	59
7.1. Boxplot ionosphere en kNN	72
7.2. Boxplot wdbc en kNN	73
7.3. Boxplot diabetes en kNN	74
7.4. Boxplot parkinsons en kNN	75
7.5. Boxplot ionosphere en SVC	76

7.6. Boxplot wdbc en SVC	77
7.7. Boxplot diabetes en SVC	78
7.8. Boxplot parkinsons en SVC	79
7.9. Rankings para fitness	82
7.10. Convergencia con SVC parte 1	83
7.11. Convergencia con SVC parte 2	84
7.12. Rankings para reducción de características	92

Índice de tablas

2.1.	Tabla de duración de cada tarea	13
2.2.	Costos servidor Google Cloud - Part 1	14
2.3.	Costos servidor Google Cloud - Part 2	15
2.4.	Costo estimado del proyecto	15
4.1.	Algoritmos modernos	31
4.2.	Algoritmos clásicos	31
6.1.	Información de los conjuntos de datos	65
6.2.	Parámetros de GOA [56]	68
6.3.	Parámetros de WOA. En el paper original [36] y en el de la versión binaria [54] no se da un valor. El valor por defecto ha sido sacado de la implementación de [77]	68
6.4.	Parámetros de ABCO [44]	68
6.5.	Parámetros de BA [57]	68
6.6.	Parámetros de PSO [60]	68
6.7.	Parámetros de FA [59]	69
6.8.	Parámetros de GA. Dada la cantidad de variantes y cantidad de valores de parámetros propuestos, estos han sido escogidos de manera empírica y basándose en recomendaciones del tutor del proyecto.	69
6.9.	Parámetros de ACO [71]	69
6.10.	Parámetros de CS [58]	69
6.11.	Parámetros de DE [43]	69
7.1.	P-valores del mejor vs el resto	80
7.2.	Fitness medio con clasificador SVC	85
7.3.	Fitness medio con clasificador kNN	86
7.4.	Ranking de los algoritmos en <i>fitness</i>	87
7.5.	P-valores de ACO vs el resto de algoritmos en reducción de características	91
7.6.	Ranking de los algoritmos en <i>accuracy</i>	93
7.7.	P-valores para todos los algoritmos en <i>accuracy</i>	94

7.8. P-valores para todos los algoritmos en <i>reducción de características</i>	95
7.9. Porcentajes de características seleccionadas y precisión en claseficación para cada algoritmo binario	96

Capítulo 1

Introducción

1.1. Motivación

El problema de la selección de características se define como el proceso de seleccionar un subconjunto de características relevantes. Una característica es una propiedad individual medible de un fenómeno concreto. Este problema es considerado un problema **NP duro** [1, 2]. La reducción de dimensionalidad, y con ello de características, suele ser necesario a la hora de crear un modelo predictivo por medio del aprendizaje automático ya que muchas de las características dentro de un conjunto de datos pueden no llegar a ser relevantes para solucionar aquellos problemas que se intentan solucionar, ya sea por que no aporta información, porque puede ser agrupada junto a otras tantas en una sola propiedad o incluso porque hay ruido en los datos, lo cual puede ser inevitable [3].

Gracias a la reducción de características es posible mejorar tanto la capacidad de generalización como la precisión del modelo predictivo gracias a la reducción de *ruido*.

Además de la simplificación del modelo, que conduce a una reducción del ruido, la selección de características es un preprocesamiento necesario por varias razones:

1. Interpretabilidad: La presencia de características irrelevantes puede complicar innecesariamente la interpretación y el rendimiento de los modelos de aprendizaje automático. La selección de un subconjunto relevante de características puede simplificar el modelo resultante, haciéndolo más comprensible y fácilmente interpretable.
2. Mejora de la eficiencia computacional: La reducción de la dimensionalidad

dad puede conducir a un ahorro significativo en términos de tiempo y recursos computacionales necesarios para el entrenamiento y la evaluación de modelos. Al eliminar características irrelevantes, se reduce la complejidad del problema y se acelera el proceso de aprendizaje. Cabe mencionar que el coste computacional de pre-procesamiento que constituye la selección de características es sí extenso, pero merece la pena de cara a posteriores usos en la construcción de modelos. Se hace una vez y se usa siempre.

3. Reduce la maldición de la dimensionalidad: Cuando la dimensionalidad se incrementa en un problema, el volumen del espacio también lo hace, y esto ocurre de manera exponencial. De forma que para obtener un resultado seguro/fiable, la cantidad de datos necesitados debe verse incrementada de manera también exponencial [4].

El reciente interés del problema de la selección de características en el ámbito de las metaheurísticas en los últimos años es más que evidente. Puede comprobarse como en los últimos años hay una tendencia en la publicación de artículos presentando nuevos métodos metaheurísticos, mejores con respecto a los clásicos o incluso comparativas y análisis entre distintos algoritmos.

Este crecimiento viene acompañado, sin embargo, de comparaciones que distan de ser objetivas por varios motivos [5]. Entre varios artículos se comparan algoritmos del mismo tipo con soluciones y resultados muy variables entre sí a pesar de mismas configuraciones a la hora de experimentar, artículos sin código referenciado, de forma que sea más fácil interpretar los resultados o duplicarlos, y algoritmos novedosos presentados por su autor o autores que superaban al resto en alguna métrica concreta sin llegar a la rigurosidad adecuada.

Por ello, la motivación principal de este trabajo es la de proveer información todo lo objetiva posible por medio de un análisis comparativo entre los algoritmos optimizadores metaheurísticos más populares y más citados junto con los algoritmos más robustos y clásicos en el campo de la optimización pseudo estocástica.

1.2. Objetivos

Objetivo General:

Realizar una comparación exhaustiva y objetiva de diversas metaheurísticas utilizadas en la selección de características, con el propósito de proporcionar

una visión integral y evaluativa sobre su eficacia y aplicabilidad en diferentes contextos de análisis de datos.

Objetivos Específicos:

1. Evaluar el desempeño de las metaheurísticas más relevantes en el ámbito de la selección de características, analizando métricas clave como precisión, estabilidad de las soluciones y eficiencia computacional. Se emplearán conjuntos de datos de referencia y metodologías de validación cruzada para garantizar la robustez de los resultados.
2. Investigar la transferibilidad de las técnicas diseñadas para dominios continuos y binarios en el contexto de la selección de características. Se analizará si las metaheurísticas efectivas en un dominio son igualmente eficaces cuando se aplican a otro, identificando posibles ventajas y limitaciones de cada enfoque.
3. Identificar las fortalezas y debilidades de cada metaheurística según el tipo de representación de las características. Se realizará un análisis detallado del comportamiento de las técnicas en problemas de selección de características con diferentes tipos de datos, destacando su rendimiento relativo y sus áreas de aplicación más adecuadas.
4. Proporcionar recomendaciones prácticas basadas en los resultados obtenidos, con el objetivo de orientar a practicantes y académicos en la selección y aplicación de metaheurísticas en problemas reales de selección de características.
5. Evaluar los resultados de las metaheurísticas en problemas de selección de característica usando distintos como algoritmos de aprendizaje los métodos *kNN* y *SVM*. Se realizará una comparativa a nivel de eficiencia en tiempo, estabilidad y calidad de los resultados.

Capítulo 2

Planificación

El proyecto se ha llevado a cabo durante un plazo de 8 meses, con una carga de trabajo adecuada a la larga duración del proyecto, siempre teniendo en cuenta el tiempo de referencia con respecto a los créditos para intentar encajar la planificación en ese intervalo temporal. Un trabajo de fin de grado consta de 12 créditos ECTS, donde se estima que cada crédito debe valer unas 25 horas de trabajo aproximadamente. Teniendo en cuenta estos datos, se calcula que la duración del TFG no debería ser superior a 300 horas, teóricamente hablando. En la práctica se han visto sobrepasadas, no por mucho, para poder realizar un proyecto de calidad. Ha de tenerse en cuenta también que el alumno trabaja 25 horas semanales y debe superar algunas asignaturas además de su proyecto final para terminar la carrera. Por lo tanto, el proyecto se planifica con una duración extendida en el tiempo, pero con una carga de trabajo semanal menos intensa.

Se planifica una duración de 8 meses aproximadamente, como ya se ha mencionado. Se utilizará un diagrama de Gantt [6] para describir la planificación del proyecto, de manera que se realizarán tareas en un orden cronológico, pero con superposición entre algunas. Varias tareas probablemente requerirán iteraciones posteriores, ya que es probable que se mejore y perfeccione el proyecto a lo largo de su ciclo de vida. Además hay otras tareas, como la investigación de metaheurísticas y la implementación de sus versiones binarias, que claramente se superponen, ya que tiene más sentido ir desarrollando cada una según se termina de estudiar para después comenzar con el siguiente algoritmo.

Las fases del ciclo de vida son:

- **Investigación inicial** Esto incluye investigar sobre conceptos básicos ya aprendidos, en forma de repaso sobre conceptos generales de aprendizaje automático, tipos de metaheurísticas, tipos de codificación, optimización de funciones, test estadísticos y conceptos básicos, código

Python y librerías asociadas, instalación de estas a partir de un entorno virtual, configuración del entorno de trabajo e investigación sobre el problema de selección de características.

- **Diseño del software:** Planificación de la estructura general del código, uso de patrones de diseño que puedan ser de utilidad de cara a al mantenimiento del software a lo largo del desarrollo, concepto de modularización inicial del código (estructura del proyecto), uso de entornos virtuales.
- **Investigación metaheurísticas:** Realización de un estudio más exhaustivo acerca de las metaheurísticas a implementar y sus diferentes versiones binarias. Esto incluye un listado de 12 metaheurísticas, cuya elección será justificada en posteriores secciones, siendo estas:
 - Firefly Algorithm
 - Whale Optimization Algorithm
 - Bat Swarm Optimizer
 - Grey Wolf Optimizer
 - Dragonfly Algorithm
 - Grasshopper Algorithm
 - Cuckoo Search
 - Differential Algorithm
 - Ant Colony Optimization
 - Artificial Bee Colony Optimization
 - Particle Swarm Optimization
 - Genetic Algorithm

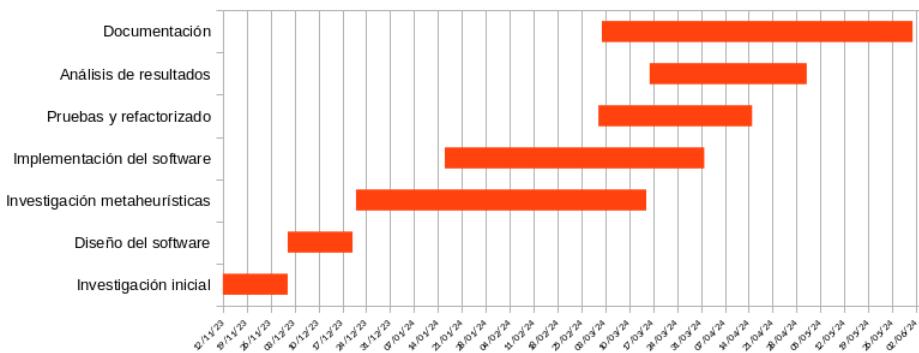
De cada una de ellas se investigará su inspiración, funcionamiento, implementación y versiones binarias, normalmente asociadas al problema de selección de características.

- **Implementación del software:** Una vez claros los requisitos programáticos quedan establecidos, se implementará el software base. Esto incluye código en Python para la generación de gráficas, manejo de datasets en formato *arff*, codificación de los algoritmos metaheurísticos en versión binaria, implementación de función objetivo (*fitness*) y parametrización del programa para distintas pruebas.
- **Pruebas y refactorizado:** En esta etapa se llevarán a cabo pruebas exhaustivas para verificar la robustez y eficacia de los diferentes algoritmos implementados. Además, se considerará la refactorización del código si es necesario, con el fin de mejorar su estructura, claridad y mantenibilidad.

- **Análisis de resultados:** En esta fase se recopilarán datos de la ejecución de los algoritmos en sus diferentes versiones, así como entre ellos, utilizando los conjuntos de datos seleccionados para el proyecto. Esta recopilación de métricas permitirá una evaluación del rendimiento y la eficacia de cada algoritmo en comparación con los demás, así como su comportamiento en diferentes conjuntos de datos.
- **Documentación:** En esta etapa final se generará una documentación del proyecto que incluirá de forma general la descripción del problema, los objetivos, planificación, implementación, resultados y pruebas.

Tarea	Duración (horas)
Investigación inicial	20
Diseño del software	19
Investigación metaheurísticas	85
Implementación del software	76
Pruebas y refactorizado	25
Ánalisis de resultados	30
Documentación	91

Tabla 2.1: Tabla de duración de cada tarea



han incidido en la duración prevista del proyecto. Esto ha llevado a una re-evaluación de la estrategia de planificación original. Entre otras cosas, la implementación del software se ha visto alargada debido a numerosos parches.

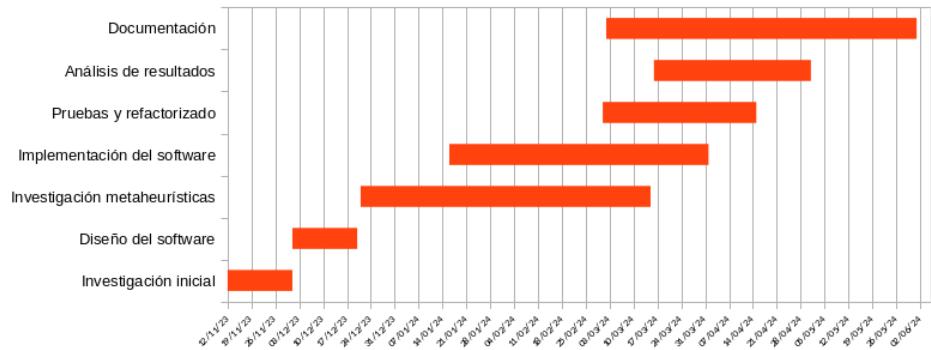


Figura 2.2: Diagrama de Gantt final

El coste estimado del proyecto se divide en varios sub-costes:

- **Sueldo:** Considerando un precio de proyecto que deba cubrir gastos y compensar, se estima un salario por investigador de 25€/hora. El proyecto ha durado aproximadamente 8 meses. Pese a ello, las horas trabajadas equivalen a un total de aproximadamente 400 horas. Por tanto, el salario final total es de 10.000€.
- **Ordenador portátil:** El estudiante consta con un portátil HP Pavilion Laptop 14-ec0xxx como herramienta principal de trabajo, con un precio de 1000€. Ha sido usado durante unos 8 meses aproximadamente. Si se tiene en cuenta que la vida media de un portátil es de aproximadamente 4 – 5 años [7] entonces el costo del ordenador sería de 133.3€.
- **Servidor** Para la experimentación del proyecto se ha hecho uso del servidor de investigación ofrecido por el tutor [8], por lo que el coste real ha sido de cero. Pese a ello, se estima cuento costaría un servicio parecido en un servidor *Cloud*. Se estima el precio por mes y total dada una estimación de los servicios de *Google Cloud - Compute Engine*.

Name	Quantity	Region	Service ID
Compute optimized Core running in London	150.0	europe-west2	6F81-5844-456A
Compute optimized Ram running in London	600.0	europe-west2	6F81-5844-456A
Storage PD Capacity in London	0.2	europe-west2	6F81-5844-456A

Tabla 2.2: Costos servidor Google Cloud - Part 1

SKU	Total Price (USD)
271A-7F2A-C5C5	6.56775
0BD6-E233-D705	3.5202
BF1A-6647-009D	0.0096
Total Price:	10.09755

Tabla 2.3: Costos servidor Google Cloud - Part 2

En la primera tabla, se detallan tres servicios distintos implementados en la región de Londres. Cada servicio se identifica por su nombre, cantidad (indicando la cantidad de recursos utilizados), región de implementación y un identificador único del servicio. Lo importante es que se ha calculado un servicio con 30 vCPUs para paralelizar los trabajos.

La segunda tabla complementa la primera proporcionando detalles adicionales sobre los servicios en términos de SKU (Stock Keeping Unit, una identificación única para un producto) y el precio total en dólares estadounidenses asociado con cada uno de ellos. El total, hecho el cambio de divisa a euros, es de 9.48€ al mes, es decir, 75.84€ en total (calculado para 8 meses). Dado este desglose, se calcula el coste final del proyecto:

Item	Costo (€)
Salario	10.000
Ordenador portátil	133.3
Servidor CPU - GC Compute Engine	75.84
Total	10.209,14

Tabla 2.4: Costo estimado del proyecto

Capítulo 3

Fundamentos Teóricos

En este capítulo se describirán aquellos conceptos teóricos fundamentales para comprender el trabajo realizado en este proyecto.

3.1. Optimización

La optimización es un campo de estudio que trata, mediante el uso de las adecuadas herramientas matemáticas, de maximizar o minimizar una función objetivo. Esto significa, obtener la mejor solución posible para un problema dado dentro de un conjunto de alternativas y normalmente sujeto a una serie de restricciones que hacen de una solución satisfacible. Es un área interdisciplinaria que aborda desde campos tales como la Economía, Ingeniería, Biología y muchas otras tantas disciplinas.

La optimización es una disciplina arraigada en la naturaleza humana. Este impulso innato hacia la optimización ha llevado al desarrollo de diversas metodologías y técnicas a lo largo de la historia, desde los rudimentarios métodos de prueba y error hasta los sofisticados algoritmos de optimización computacional utilizados en la actualidad.

3.1.1. Definición general de un problema de optimización

Para poder convertir un problema abstracto en un problema de optimización concreto, con el que se pueda trabajar, es necesario establecer ciertos elementos fundamentales que lo definan de manera precisa y clara. En general, un problema de optimización se expresa de la siguiente forma [9]:

Función objetivo a minimizar $f(x)$ (3.1a)

Sujeto a

s restricciones de desigualdad $g_i(x) \leq 0, \quad j = 1, 2, \dots, s$ (3.1b)

w restricciones de igualdad $h_j(x) = 0, \quad j = 1, 2, \dots, w$ (3.1c)

Donde el número de variables es dado por $x_i, \quad i = 1, 2, \dots, n$

La definición general de un problema de optimización proporciona una estructura sólida para abordar el problema abstracto. Establece la función objetivo que se busca minimizar o maximizar, junto con las restricciones que deben cumplirse. Estas restricciones pueden ser tanto desigualdades como igualdades, y todas juntas definen el **espacio de búsqueda** del problema.

3.1.2. Función objetivo y función fitness

Ambos términos, aunque a menudo se utilizan como sinónimos, desempeñan roles distintos en el ámbito de la optimización. La función *objetivo*, como su nombre indica, establece el objetivo a alcanzar en la resolución del problema. Esta función cuantifica el rendimiento de las soluciones encontradas en relación con el objetivo específico del problema, que puede ser maximizar ganancias, minimizar distancia, entre otros objetivos. Por otro lado, la función de *fitness* solo tiene que evaluar la idoneidad de una solución dentro de una población de soluciones. Es decir, determina la calidad relativa de la solución respecto a otras alternativas.

Por otro lado, la función objetivo puede ser positiva o negativa, dependiendo de si se busca maximizar o minimizar el objetivo. Además, la función *fitness* también puede llegar a ser una aproximación de la función objetivo, pero no necesariamente coinciden exactamente.

Resumiendo, la función *fitness* es un tipo particular de función objetivo que se utiliza como métrica de rendimiento [10].

3.1.3. Óptimos globales o locales

Se conocen como punto de óptimo global (mínimo o máximo) la solución (o vector hablando en términos matemáticos) cuyo valor para la función objetivo es el más grande en todo el espacio de soluciones posible, es decir, el espacio de búsqueda. Los óptimos locales en cambio son varios, no solo uno como es el global. Son soluciones máximas o mínimas dentro de una región de soluciones.

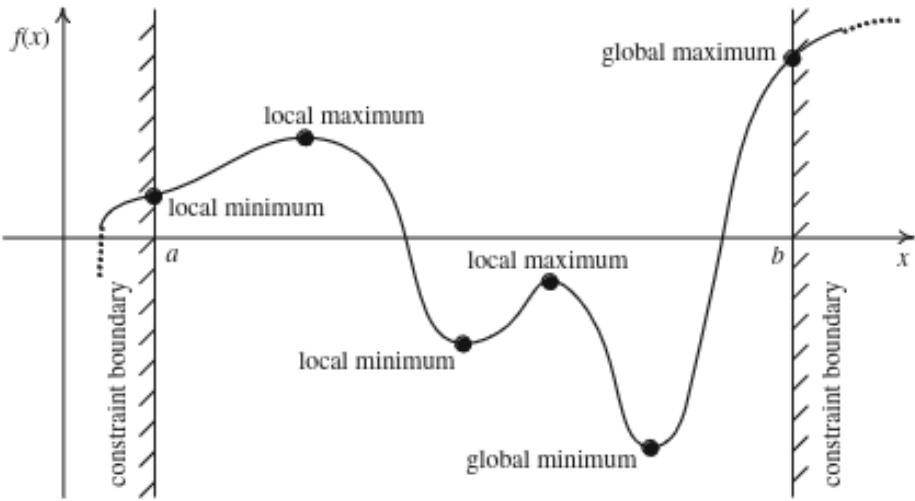


Figura 3.1: En esta figura extraída de [9] puede observarse de manera intuitiva la diferencia entre punto global y puntos locales. El máximo global es el valor más alto para $f(x)$ en el espacio, mientras que un máximo global solo representa el valor más alto dentro de un “vecindario”

Sea $f(x)$ una función a maximizar y x^* una solución óptima. Un objetivo $G(x)$ está en su máximo global sí y solo si [9]:

$$f(x^*) \geq f(x) \quad \forall x \quad (3.2)$$

En cambio el objetivo está en un máximo local en el punto x^* si:

$$\begin{aligned} f(x^*) &\geq f(x) \quad \forall x \\ &\text{dentro de un vecindario de } x^* \end{aligned} \quad [9] \quad (3.3)$$

3.2. Selección de características

La selección de características es un ejemplo de problema *NP-Hard* y uno de los problemas más importantes en el mundo de la inteligencia artificial, más concretamente en el **machine learning** o aprendizaje automático.

3.2.1. Necesidad y motivo

El aprendizaje automático se basa en los datos como fuente de aprendizaje. Es indispensable tener un buen conjunto de datos para poder obtener un modelo robusto y preciso. Por norma general o como se diría en inglés “as a

rule of thumb", a más grande el conjunto de datos, mayor calidad del modelo. De hecho hay una correlación inmediata con esta sentencia.

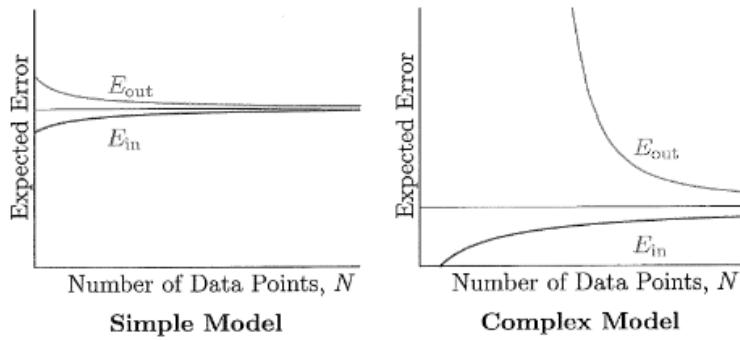


Figura 3.2: Esta figura extraída de [3] visualiza la relación directa entre calidad del modelo (menos error esperado) y número de datos que se le proporciona al algoritmo de aprendizaje.

Es inmediato pensar que, a mayor cantidad de datos, mejor calidad del modelo, y así suele ser. Sin embargo, también existe una “normal general” que relaciona la complejidad del modelo y su capacidad de generalización. Es cierto que un modelo muy complejo (muchos parámetros) es capaz de ajustar mejor funciones más complejas, pero dentro de un conjunto con una serie de modelos suficientemente complejos, suele ser mejor idea elegir el más simple. Hay una serie clara de ventajas para ello:

1. Menor sensibilidad al sobre-ajuste.
2. Mayor interpretabilidad del modelo y sus resultados.
3. Mayor eficiencia y por tanto menor tiempo de ejecución y menor peso en memoria.

De hecho, como puede observarse en la figura 3.2, el modelo más complejo tiene un error esperado mucho mayor que el simple. El E_{out} o error fuera de la muestra (error de generalización) es mucho más abrupto, es decir, generaliza peor.

Por supuesto, con suficientes datos, con un N (número de datos) suficientemente grande, la tendencia del error es a la baja [3, 11]. Ocurre, sin embargo, que la recolección, limpieza y transformación de datos es una tarea compleja, por ello es mejor ceñirse, de nuevo, a el modelo más pequeño que obtenga una solución con suficiente calidad.

Como se ha mencionado anteriormente, la selección de características ayuda a la reducción del ruido. Siendo f la función objetivo a predecir, H^n el conjunto de hipótesis o conjunto de modelos de dimensión n posibles, $h^*(x)$ el mejor modelo aprendido y x una variable de entrada. El ruido conocido como ruido estocástico es aquel que atiende a una variación aleatoria que puede surgir de diversos factores, como mediciones imprecisas de señales o la falta de precisión en sensores. Por otro lado, el ruido determinista está directamente relacionado con la complejidad de un modelo. Su presencia aumenta la probabilidad de sobre-ajuste. El ruido determinista puede explicarse como la parte de la función f que el conjunto de hipótesis H^n no puede capturar, es decir, $f(x) - h^*(x)$. Este tipo de ruido se considera así porque la función (modelo) no es lo suficientemente compleja como para comprender esa parte. Este ruido depende de H^n y permanece constante para un valor dado de x [3].

La reducción de características ayuda a manejar ambos tipos de ruido [3, 12] al simplificar el modelo, lo que puede reducir el impacto del ruido estocástico y disminuir la complejidad del modelo, lo que a su vez puede ayudar a mitigar el ruido determinista al mejorar la capacidad del modelo para capturar las características relevantes y descartar las irrelevantes. Esto puede conducir a una mejor capacidad de generalización y a una reducción del sobre-ajuste.

3.2.2. Concepto

El funcionamiento y motivo es explicado ya en la sección de motivación en 1.1. Por evitar redundancias se procede a explicar el concepto de manera abreviada y puntuizando en los apartados más importantes.

La selección de características es una conocida y necesaria técnica de preprocesamiento de datos para la construcción de un modelo de aprendizaje. Su función es escoger un subconjunto óptimo de características dentro del conjunto inicial, de forma que la complejidad del modelo se reduzca.

Este problema es del tipo *NP-Hard*, como ya se ha mencionado previamente en 1.1 y explicado en ???. La selección de características ayuda a una serie de factores como son la interpretabilidad, mejora de la eficiencia temporal y espacial del modelo y sobre todo con la **maldición de la dimensionalidad** [13, 14].

3.2.3. Maldición de la dimensionalidad

El término fue acuñado por Bellman en 1961 [15]. Este fenómeno ocurre cuando la dimensionalidad de los datos es muy grande. En un espacio de características de alta dimensión, es común que los datos estén muy dispersos, lo que significa que hay muy pocos datos en comparación con la cantidad posible de características. Esto dificulta que los modelos representen correctamente todo el espacio de características [16].

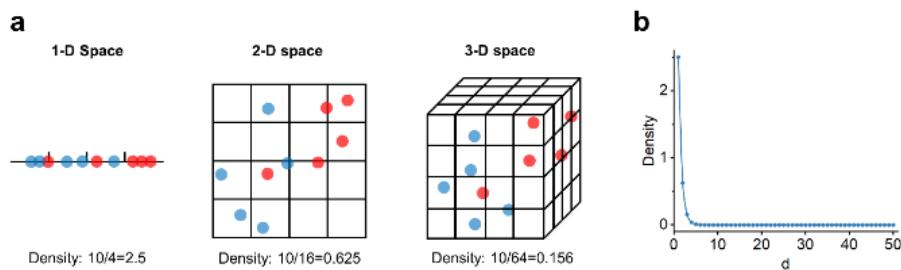


Figura 3.3: Esta figura extraída de [16] muestra la tendencia en la densidad a medida que se incrementa la dimensionalidad del espacio. En (a) se muestra la densidad con diez puntos de datos en un espacio 1D, en (b) se muestra la tendencia al incrementar las dimensiones.

Además, en espacios de alta dimensión, la medición de distancias puede volverse menos precisa. Esto se debe a que las distancias entre puntos de datos diferentes tienden a converger a un mismo valor a medida que aumenta la dimensionalidad. Esto significa que las medidas de distancia ya no son tan útiles para medir la similitud entre datos [13, 16]. En el trabajo de Beyer y colaboradores [17] se observó como esto ocurría y por tanto un escaneo lineal (recorrer los puntos uno a uno) resultaba en altas dimensiones más práctico que otras técnicas complejas.

3.3. Metaheurísticas

Dentro de la optimización hay muchos tipos de métodos, dentro de los pseudoaleatorios pueden encontrarse las metaheurísticas. Estas son algoritmos basados en una abstracción de mayor nivel de la **heurísticas**. Mientras que las heurísticas se apoyan en el conocimiento específico del campo en el que se encuentra el problema, y están restringidas a su dominio, las metaheurísticas son aplicables a todo tipo de problemas, independientemente de su área de optimización [18]. Es cierto que hay algoritmos que son más convenientes para ciertos problemas que otros, pero su aplicación es generalizada. Normalmente,

las metaheurísticas son diseñadas a siguiendo una inspiración en la naturaleza, ya sea en fenómenos físicos o en el comportamiento animal. Ejemplo de estos son algoritmos como el *Búsqueda Cuckoo*, *Enfriamiento Simulado* o incluso *Algoritmos Genéticos*.

Las metaheurísticas son especialmente útiles en problemas cuya resolución no es factible debido a altos costos computacionales, ya sea porque es posible analíticamente pero computacionalmente costoso, o porque el problema no es abordable mediante algoritmos convencionales. Son capaces de encontrar óptimos locales lo suficientemente aceptables, soluciones no óptimas, pero si muy buenas [18].

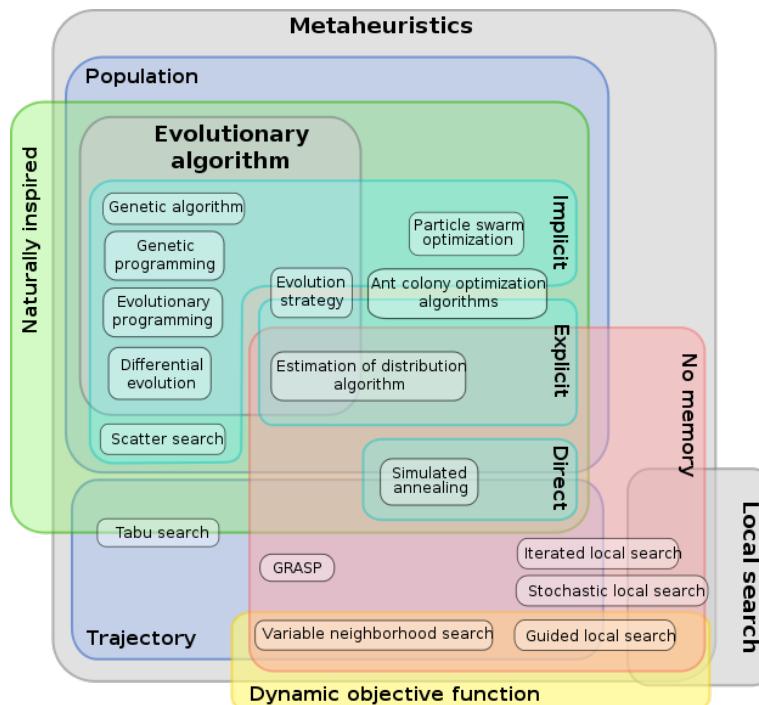


Figura 3.4: En esta figura de los autores Johann “nojhan” Dréo y Caner Candan, se clasifican las distintas metaheurísticas por inspiración.

3.3.1. Exploración vs explotación

Las metaheurísticas, y algoritmos basados en técnicas pseudoaleatorias, deben tener un balance entre sus factores explorativos y explotativos. De no ser así, los algoritmos tendrían tendencias a la convergencia temprana, dejando de lado mucho espacio por explorar y por ende soluciones posiblemente mejores (estancamiento en óptimos locales) o serían muy lentos en la convergencia hacia una solución.

La **exploración** se refiere a la habilidad del algoritmo de buscar nuevas y diversas regiones en el espacio de búsqueda/solución. Es una característica de la búsqueda global, que también puede ser llamada *diversificación*. En cambio, la **explotación** es la habilidad de la búsqueda de explotar las mejores soluciones encontradas hasta el momento y mejorárlas localmente, dentro de un “vecindario” [19].

No existe un equilibrio *de facto* entre exploración y explotación en las metaheurísticas. Aún no se ha alcanzado una respuesta definitiva a esta cuestión. Desde una perspectiva de sistemas, una metaheurística puede entenderse como un sistema dinámico compuesto por numerosos “individuos” que interactúan entre sí. Estos individuos representan las distintas soluciones o posiciones en el espacio de búsqueda que la metaheurística explora [20]. Las interacciones entre estos individuos son las que conforman comportamientos explorativos o explotativos y dependen del problema a solucionar y del propio algoritmo su equilibrio.

3.4. Teorema No Free Lunch

El Teorema de “No Free Lunch” (NFL) establece que, en promedio, ningún algoritmo de búsqueda puede superar a otros algoritmos en la búsqueda de todas las funciones objetivo posibles. En otras palabras, no existe un algoritmo universalmente óptimo que pueda dominar en todos los problemas de búsqueda. Esto implica que, si un algoritmo es efectivo para un conjunto particular de problemas, es probable que no lo sea para otros [21].

Relacionar este teorema con las metaheurísticas implica reconocer que no hay una única metaheurística que sea la mejor para todos los problemas de optimización. Cada problema puede tener características únicas que lo hacen más o menos adecuado para ciertas metaheurísticas. Por lo tanto, en lugar de buscar una solución universal, las metaheurísticas se centran en explorar y explotar diferentes áreas del espacio de búsqueda para encontrar soluciones aceptables o incluso óptimas para problemas específicos.

Pese a ello, las suposiciones de este tipo de teoremas, como conjuntos de datos extraídos de una distribución uniforme sobre todos los conjuntos de datos posibles, están completamente desalineadas con el mundo real, donde los datos suelen ser altamente estructurados y no están uniformemente muestreados [22].

De esta forma y faltando evidencia concluyente al respecto, es interesante mencionar el **NFL**, pero sin llegar a negar la posibilidad de nuevas y más precisas interpretaciones.

3.5. Aprendizaje automático

El aprendizaje automático es una sub-rama de estudio de la inteligencia artificial o **IA**, la cual aglomera una serie de métodos que pueden, de manera automática, detectar patrones en conjuntos masivos de datos para predecir datos futuros [23]. El aprendizaje automático o *machine learning* en inglés, es usado en una amplia variedad de campos por su utilidad trasversal. Algunos de ellos son la agricultura, marketing, videojuegos, meteorología, física, etc.

Los algoritmos de aprendizaje automático son capaces de encontrar patrones en los datos, como ya se ha mencionado. Por ello, es necesario nutrir a estos algoritmos con datos de calidad. Un modelo de aprendizaje automático será tan bueno como los datos que se le puedan proveer, no más. De esta forma la recolección de datos y su procesamiento se convierten en una prioridad a la hora de crear modelos.

Hay muchas formas de estructurar los datos y muchos tipos de algoritmos acorde a estos “inputs”. En este documento se tratará con información en formato tabular, es decir, información en tablas con filas y columnas, donde cada fila representa un registro de información y cada columna una característica asociada. Esta última es la que determinará la complejidad del modelo.

3.5.1. Aprendizaje supervisado

Este subtipo de aprendizaje automático se caracteriza por tener un conjunto de datos sobre los que se entrena y una salida esperada (etiquetas) para cada punto de los datos [24]. Es bastante costoso porque etiquetar cada dato es una tarea laboriosa y poco escalable.

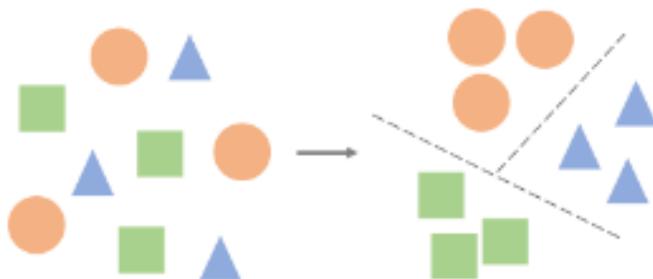


Figura 3.5: Figura extraída de [24] que muestra un ejemplo de aprendizaje supervisado.

3.5.2. Aprendizaje no supervisado

Cuando los datos solo vienen en forma de entrada y no se tiene ninguna salida (no hay etiquetas) entonces se trata de un aprendizaje no supervisado. Los algoritmos de este tipo se basan en la diferenciación de los datos mediante los patrones subyacentes que puedan encontrar. Son comunes en el aprendizaje no supervisado los algoritmos de *clustering* [24].

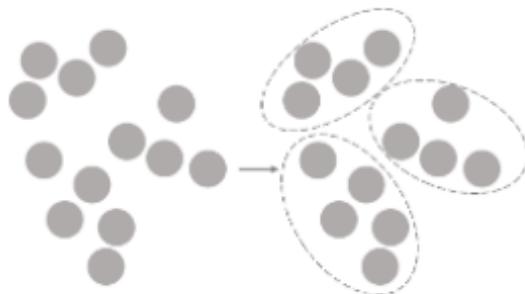


Figura 3.6: Figura extraída de [24] que muestra un ejemplo de aprendizaje no supervisado. Se trata de dividir en clústers sin información de etiquetas.

3.5.3. SVM

Las máquinas de vectores de soporte o *SVM* fueron introducidas por primera vez en [25]. En este documento será uno de los algoritmos de clasificación usados en la función *fitness* para cuantificar la calidad de los pesos aprendidos por los algoritmos optimizadores.

Las máquinas de vectores de soporte son algoritmos cuya principal característica se basa en la creación de un hiperplano o conjunto de hiperplanos en un espacio n -dimensional [26]. Este hiperplano o hiperplanos es elegido de manera que maximice el margen entre los puntos de datos de todas las clases a predecir. El margen es la distancia entre el hiperplano y los puntos de datos más cercanos de cada clase, estos puntos se denominan vectores de soporte. De esta manera se consigue minimizar el error de generalización [27].

La idea más intuitiva es, que al haber más espacio entre los puntos de distintas clases, hay más posibilidad de que los puntos no vistos, los puntos a predecir, caigan en zonas correctas.

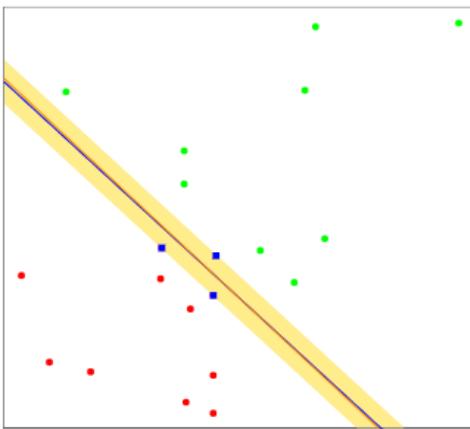


Figura 3.7: Figura extraída de [27] en la que se muestra de forma gráfica en un espacio 2-dimensional el margen óptimo de separación de dos clases.

3.5.4. K-nearest neighbors

El algoritmo de k vecinos más cercanos fue presentado en [28] por Evelyn Fix y Joseph Hodges, y más tarde ampliado en [29] por Thomas Cover.

Puede ser utilizado para regresión o clasificación, pero su uso suele darse más en este último problema. Este algoritmo ofrece una premisa sencilla, la clase de un punto p dependerá de cuál sea la clase mayoritaria dentro del subconjunto de k vecinos más cercanos [30].

Valores mayores de k incrementan la varianza, pero decrementan el sesgo (equilibrio sesgo-varianza [3]). De forma inversa, a menor k mayor sesgo y menor varianza.

Otra forma de verlo es que si $k = 1$ (el valor mínimo), el modelo entrenado será muy complejo, pues cada punto puede variar de clase por mínimo que sea el cambio en su entorno. Es un modelo que tiende a sobre-ajustar. Con $k = n$ el modelo es lo más simple posible, al tener en cuenta absolutamente todos los puntos, la clasificación será siempre la de la clase más repetida.

Capítulo 4

Revisión de la literatura

La selección de características es un problema cuya popularidad ha ido en aumento con el paso de los años. No es algo casual, pues la cantidad de datos recogidos para tareas de aprendizaje automático y áreas derivadas, como el aprendizaje profundo, ha ido en aumento de manera casi exponencial.

En este trabajo, se lleva a cabo una investigación y análisis comparativo entre varios métodos de la familia **wrapper** o métodos de envoltura. Existen multitud de estrategias [12] que intentan dar solución a este problema. Los métodos de búsqueda más famosos son los de filtrado (**filter**), los cuáles seleccionan las características más discriminativas según la naturaleza de los datos [12]. Por lo general, estos métodos realizan la selección de características antes de las tareas de clasificación y agrupamiento. Ejemplos de algoritmos de filtrado son *reliefF* [31] o F-statistic [32].

Los métodos **wrapper**, en cambio, utilizan el algoritmo de aprendizaje usado post-procesamiento para evaluar las características y seleccionar así las más útiles [12].

Los algoritmos clasificatorios de aprendizaje utilizados en este trabajo son *SVM* [25] y *kNN* [28, 29], siendo las máquinas de vectores de soporte un método robusto y eficiente y los vecinos más cercanos un método simple, interpretable y muy eficaz. Se analizará el resultado entre ambos clasificadores entre otros muchos análisis comparativos.

Se puede observar en 4.1 la creciente popularidad de los métodos de selección de características en el buscador de **Scopus**. La tendencia parece indicar que la tendencia científica puede incluso crecer aún más alrededor del problema de selección de características. Y es que pese a los grandes avances en *hardware*, parte del techo con el que se topan los algoritmos de aprendizaje sigue siendo los enormes tiempos de entrenamiento. También ha de mencionarse la tendencia, prácticamente igual en cuanto a escala de crecimiento, en la popularidad sobre los años del problema de selección de características, pero

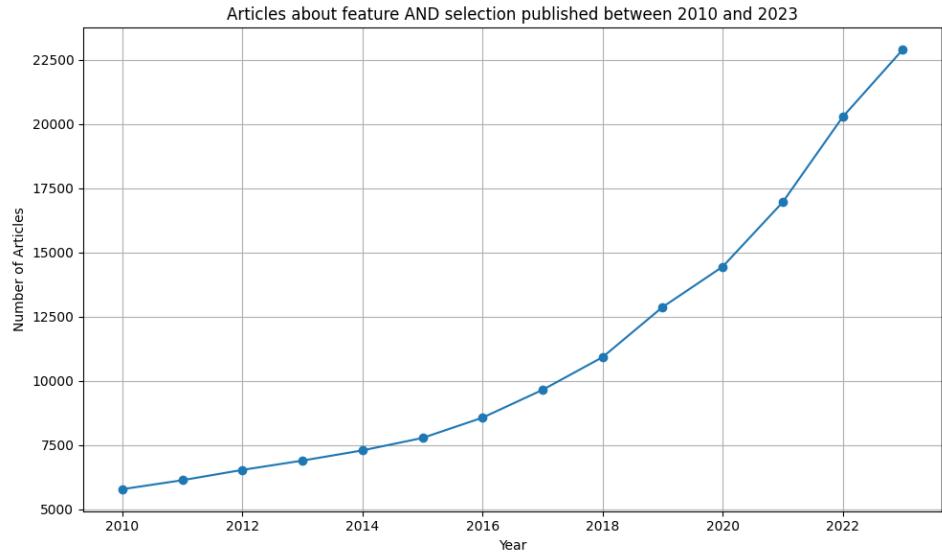


Figura 4.1: En esta figura se muestra el número de artículos publicados relacionados con la selección de características. Se ha usado el buscador Scopus para estos resultados.

abordado con técnicas metaheurísticas 4.2.

Y es que las metaheurísticas son algoritmos muy convenientes en la resolución de este tipo de problemas. Los algoritmos analíticos pueden enfrentar dificultades cuando el conjunto de datos tiene un número muy elevado de características, debido a que la cantidad de combinaciones posibles puede hacer que el problema sea inabordable en términos computacionales. Debido a ello, se suelen utilizar algoritmos más rápidos cuya solución no es óptima en términos globales, pero si suficientemente buena.

Muchos algoritmos metaheurísticos han sido propuestos a lo largo de los años. Se han escogido los más destacables en cuanto a resultados y más mencionados entre todos ellos, dando una selección de algoritmos metaheurísticos modernos de, en principio, muy alta calidad. No solo son seleccionados algoritmos modernos, sino que también se han escogido una serie de algoritmos más “clásicos”, pero cuya aplicación es más extendida y con resultados que, de forma empírica, han demostrado ser más que buenos a lo largo de décadas de uso.

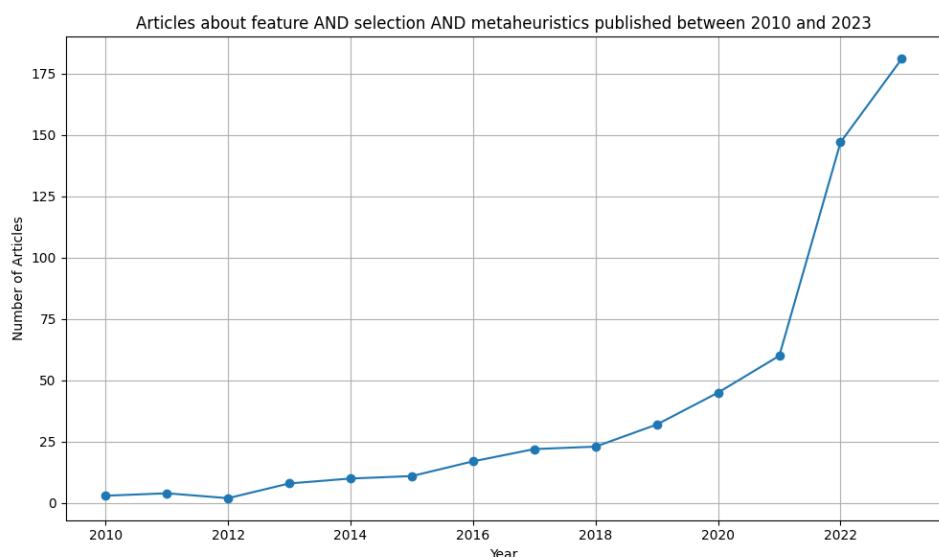


Figura 4.2: De igual forma que en la figura 4.1 las mateheuristicas han ido muy ligadas a la resolución de este problema. Se ha usado el buscador Scopus para estos resultados.

Algoritmos
GWO [33]
FA [34]
GOA [35]
WOA [36]
DA [37]
CS [38]
BA [39]

Algoritmos
PSO [40]
GA [41]
ACO [42]
DE [43]
ABCO [44]

Tabla 4.2: Algoritmos clásicos

Tabla 4.1: Algoritmos modernos

Los algoritmos usados en este proyecto pertenecen todos a la sub-categoría conocida como algoritmos poblacionales, esto es, algoritmos que parten de un conjunto de soluciones inicializadas de forma aleatoria y representadas en forma de vectores llamadas población. Este tipo de algoritmos “evolucionan” durante un número de generaciones hasta alcanzar un máximo número de generación o alcanzar un miembro de la población cuya solución sea lo suficientemente buena [45].

4.1. Introducción a los primeros algoritmos poblacionales

Uno de los primeros algoritmos basados en poblaciones y evolutivos es el algoritmo genético. Los *algoritmos genéticos* o **GAs** ganaron popularidad en los años 70, particularmente en 1975 con el libro de John Holland [41]. Este tipo de metaheurísticas son diseñadas basándose en la selección natural. Un conjunto de fenotipos (soluciones) son evolucionadas durante generaciones para emular el cruce entre especies (cruce de soluciones mediante un intercambio común de cromosomas) dando lugar a nuevos individuos con características de ambos padres. Poco a poco este tipo de algoritmos fueron desarrollando nuevas características y aunque en un principio fueron diseñados para resolver problemas discretos, también tienen versiones que optimizan problemas continuos [10].

Uno de los primeros algoritmos metaheurísticos basados en poblaciones es el *sistema de hormigas* de Dorigo, que fue presentado en 1991 [46]. Esta metaheurística fue aplicada a la resolución del problema del vendedor ambulante con resultados muy prometedores, pues resolvía el problema de forma muy eficiente con una solución sub-óptima, pero suficientemente buena. A partir de este punto, surgieron numerosas mejoras y versiones de los algoritmos basados en colonias de hormigas.

Más tarde se presentó el conocido algoritmo de *optimización por enjambre de partículas*, nombrado como **PSO** por sus siglas en inglés y presentado por Kennedy y Eberhart en 1995 [40]. El algoritmo **PSO** surgió a partir de la observación del comportamiento social de las aves y peces. La idea principal del **PSO** es que cada solución candidata (o “partícula”) en el espacio de búsqueda se mueve a través del espacio en función de su propia mejor posición conocida y la mejor posición conocida de cualquier partícula en el enjambre, en lugar de depender de operadores de búsqueda convencionales. El algoritmo comenzaba con un conjunto de soluciones candidatas aleatorias, denominadas partículas, que procedían a moverse a través del espacio de búsqueda para encontrar la solución óptima. Cada partícula tenía una posición y una velocidad asociadas.

En 1999 se presentó el conocido **ACO** u *optimización de colonia de hormigas* [42]. Este algoritmo refinada la fórmula anteriormente usada en **AS** usando un grafo como representación del conjunto de soluciones, la actualización y rastro de feromona como heurística principal para escoger el camino en el grafo y una codificación discreta de las características.

La *evolución diferencial* fue un algoritmo de optimización propuesto por primera vez por Rainer Storn y Kenneth Price en el año 1997 [43]. Surgió como una alternativa a los algoritmos genéticos existentes en ese momento,

ya que se inspiraba en su diseño, pero siendo inicialmente diseñado para resolver problemas continuos, a diferencia de los **GAs**.

La idea detrás de la evolución diferencial se basaba en la exploración de un espacio de búsqueda mediante la manipulación de vectores de parámetros. A diferencia de los *GAs*, que utilizaban operadores genéticos como la cruce y la mutación, la **DE** se centraba principalmente en la mutación basada en la diferencia entre vectores de la población. Inicialmente, el algoritmo atrajo la atención debido a su simplicidad y eficiencia en la optimización de funciones complejas en espacios de búsqueda continuos.

El algoritmo de optimización basado en colonias de abejas es considerado otro gran clásico de las metaheurísticas poblacionales bio-inspiradas, pese a ello, es el más nuevo de los ya mencionados. El algoritmo de colonias de abejas artificiales o **ABCO** fue propuesto por primera vez por Dervis Karaboga en 2005 [44]. Surgió como una técnica inspirada en el comportamiento de las colonias de abejas para resolver problemas de optimización y se basaba en imitar el comportamiento de las abejas en la búsqueda de alimentos. En la naturaleza, las abejas utilizan la danza de reclutamiento para comunicar la ubicación de las fuentes de alimento a otras abejas en la colmena. Karaboga adaptó este proceso en un algoritmo de optimización que utiliza abejas artificiales para explorar y explotar un espacio de búsqueda.

4.2. Propuestas metaheurísticas modernas

Con el paso de los años se han ido refinando y probando nuevos tipos de algoritmos. Nuevas inspiraciones, basadas en el comportamiento animal, en teoría social, en teoría física o matemática. También se han creado nuevas versiones de algoritmos clásicos para resolver ciertos problemas que el algoritmo tenía de base o abordar otros nuevos. De cualquier manera, el número de nuevas metaheurísticas no ha dejado de crecer desde su origen, pues sus usos son muy extensos. Desde la selección de características hasta métodos de búsqueda de hiper-parámetros muy eficientes y con muy buenos resultados [47].

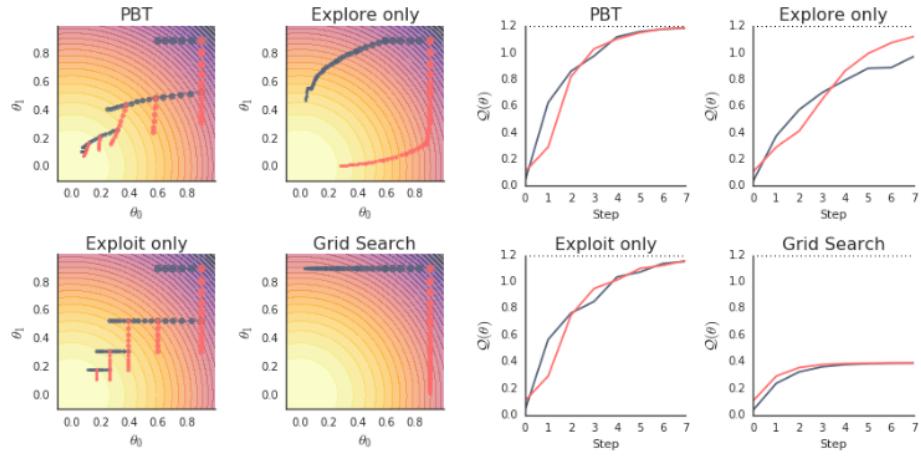


Figura 4.3: En esta figura se muestra varios métodos usados en la búsqueda de hiper-parámetros frente al método basado en poblaciones **PBT**. La zona más blanca representa el mínimo, ya que se trata de minimizar un error de pérdida a la vez que se maximiza una métrica concreta. Population-based training (**PBT**) es una técnica en el ámbito del aprendizaje automático y el entrenamiento de modelos que combina conceptos de algoritmos genéticos y métodos de ajuste de hiper-parámetros para mejorar el rendimiento y la eficiencia del entrenamiento de modelos. [47]

Dada una exhaustiva investigación, que incluye múltiples búsquedas en los últimos años en el motor de búsqueda de Scopus y una extensa lectura de varios estudios (“surveys”) [48-50] enfocados en metaheurísticas aplicadas a la **selección de características** se ha hecho una lista de los más usados, citados y con mejores resultados.

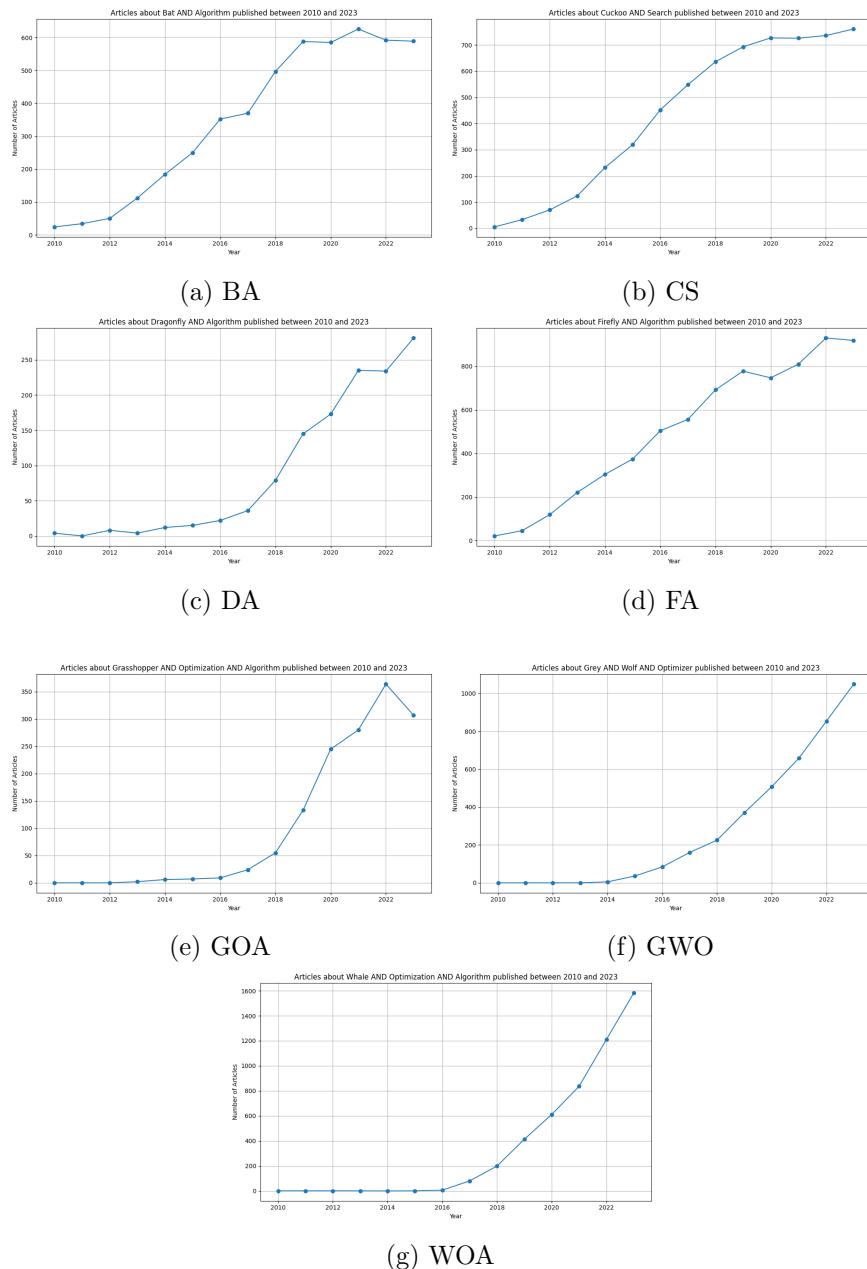


Figura 4.4: Número de búsquedas en Scopus de los diferentes algoritmos metaheurísticos seleccionados. Se puede ver una clara tendencia alcista en la popularidad de estos pese a su relativa novedad en el campo de la optimización.

El más “antiguo”, dentro de la novedad, de los siete algoritmos aquí citados,

es el de la *búsqueda cuco* o **CS** dadas sus siglas en inglés. Este algoritmo fue presentado a finales de 2009. En el artículo original [38] comparaba su rendimiento con un clásico ya mencionado en este capítulo, el **PSO**.

Esto no es novedad, todos los algoritmos más recientes hacen gala de sus buenos resultados en comparación a algoritmos clásicos como son el **ABCO**, **DE** o **PSO** entre otros muchos.

La búsqueda cuco se inspiraba en el comportamiento parasitario de los pájaros de cuyo nombre de apropió. Estos colocan sus huevos en los nidos de otras aves y además retiran los huevos ajenos para incrementar las posibilidades de incubación de los suyos. Esto pequeña premisa se simula matemáticamente por los agentes de la población y se combina con el uso de funciones que simulan vuelos Lévy (pues se ha observado que el patrón de vuelos de muchas aves e insectos tienden a encajar en ese tipo de distribución [51]).

El *algoritmo de los lobos grises* o **GWO** [33] y el *algoritmo de optimización de la ballena* o **WOA** [36] comparten similitudes en la forma que actúan sus operadores. Estos se basan en la búsqueda de una presa y en el rodeo de la misma, incorporando siempre una variable aleatoria que va dejando espacio, según avanzan las iteraciones, a la convergencia de la solución.

Otros, como el **DA** [37] y el **GOA** [35], comparten la idea de la atracción y la repulsión para equilibrar la balanza entre exploración y explotación. En el caso del algoritmo de la libélula es atracción por presas (soluciones muy prometedoras) y repulsión hacia enemigos (peores soluciones encontradas), mientras que el algoritmo de optimización del saltamontes la atracción y la repulsión están relacionadas directamente con el resto de soluciones (agentes). Se repelen entre sí para favorecer la exploración del espacio y a medida que se avanza en las iteraciones o se encuentran soluciones suficientemente buenas, se incrementa la atracción entre saltamontes para dar lugar a la explotación.

Estas nuevas metaheurísticas aportaron nuevas ideas muy novedosas e interesantes. Otros algoritmos incluidos en la recopilación de este proyecto y que han sido muy mencionadas y usadas son las de **BA** [39] y **FA** [34]. **BA** (algoritmo del murciélagos) es inspirada por la forma en la que se guían (y cazan) los murciélagos, la emisión de pulsos de ecolocalización. **FA** (algoritmo de las luciérnagas), en cambio, se basa en la emisión de luz para atraer presas, atraer parejas y repeler enemigos.

4.3. Versiones binarias

Esta selección de siete metaheurísticas es, según la investigación realizada, la más representativa de la actualidad en cuanto al problema de selección de

características se refiere. Su diseño permite fácilmente la modificación del algoritmo para adaptarlos al problema de selección de características, el cual requiere de una codificación discreta de los vectores soluciones. Algunas de las propuestas más novedosas para el problema de selección de características son las de **bGWO** [52], **bWOA** [53, 54], **bDA** [55], **bGOA** [56], **bBA** [57], **bCS** [58] y **bFA** [59]. Por su parte, los algoritmos clásicos también tienen múltiples versiones “binarias” o discretas, véase **bPSO** [60], **bABCO** [61] o **bDE** [62]. Nótese que los algoritmos genéticos no tienen como tal una versión binaria como el resto de algoritmos, pues estos fueron pensados originalmente para resolver problemas discretos. Gracias a que sus operadores son muy flexibles es posible también que acepten una codificación real y que resuelvan ese tipo de problemas con un rendimiento muy alto. En cuanto a los algoritmos de hormigas, tampoco necesitan una modificación especial, pues fueron creados también para resolver problemas discretos. Pese a que **ACO** tenga versiones del algoritmo con codificación continua [63], en este proyecto se ha preferido obviarlas, pues se considera que alteran demasiado el diseño del algoritmo original como para considerarlo una versión en vez de una metaheurística nueva.

Puede observarse que la técnica más empleada para pasar de un algoritmo con codificación continua a codificación binaria es la del uso de *funciones de transferencia* [48, 60, 64]. De este tipo de funciones hay varias familias, pero las más usadas sin lugar a duda son las funciones del grupo “s-shaped” y “v-shaped”, nombre que proviene de la forma en “s” y “v” que tienen este tipo de funciones representadas en una gráfica. La premisa es sencilla, la característica x_i tendrá más probabilidad de ser 0 o 1 dependiendo de su valor tras pasar por la función de transferencia. Se representa por tanto el recorrido desde los valores para la $y \in [0, 1]$ como una probabilidad dada una entrada x .

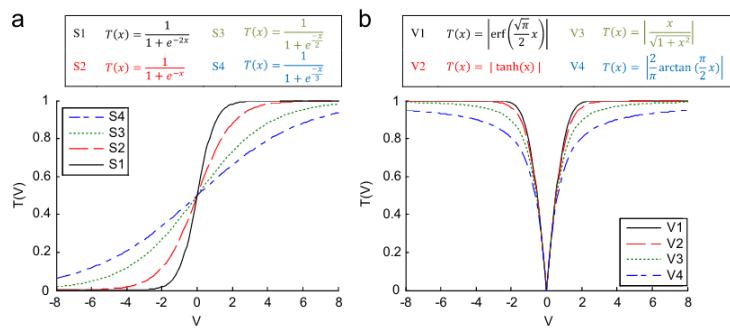


Figura 4.5: En esta figura extraída de [60] se muestran funciones del tipo **S-shaped** (a) y **V-Shaped** (b).

Capítulo 5

Descripción de los algoritmos

Para aclarar ciertas cuestiones referentes a las soluciones en este tipo de algoritmos estocásticos de tipo evolutivo, si no se especifica un tipo de representación se da por hecho que las soluciones se representan como vectores en un espacio n -dimensional. Un conjunto de individuos/vectores dan como resultado una población. La población es una matriz de $N \times M$, donde N es el número de filas o número de individuos en la población y M es el número de características del problema (M dimensiones). Algunos algoritmos admiten codificación binaria y otros continua, ese tipo de cuestiones si serán especificadas.

Por lo general, excepto para los algoritmos **GA** y **ACO**, se utilizarán funciones de transferencia a la hora de modificar los algoritmos para su versión binaria, tal y como se ha mencionado en el repaso bibliográfico.

5.1. Algoritmos genéticos

5.1.1. Introducción

Los algoritmos genéticos están inspirados en la selección natural y se utilizan tanto en problemas de optimización con restricciones como sin ellas. Es una metaheurística que modifica una población de soluciones individuales de manera repetida, seleccionando soluciones “padre” que dan lugar a la siguiente generación de soluciones en la siguiente iteración del algoritmo. En su forma más básica, un algoritmo genético opera sobre una población de soluciones potenciales a un problema dado. Cada solución potencial, a menudo llamada individuo o cromosoma, está representada como una cadena de símbolos, que puede ser binaria, numérica o simbólica [65].

5.1.2. Funcionamiento y Operadores

Las soluciones o **fenotipos** están compuestas por una serie de propiedades, características, cromosomas o **genotipos**, que pueden ser mutados y alterados. La representación más común es la binaria, pero otras codificaciones son posibles.

El proceso típico incluye [66]:

- Inicialización aleatoria de la población.
- Evaluación del *fitness* de cada individuo.
- Selección de padres basada en el *fitness*.
- Creación de una nueva generación mediante cruces y mutaciones.
- Elitismo: selección de los mejores individuos [67].

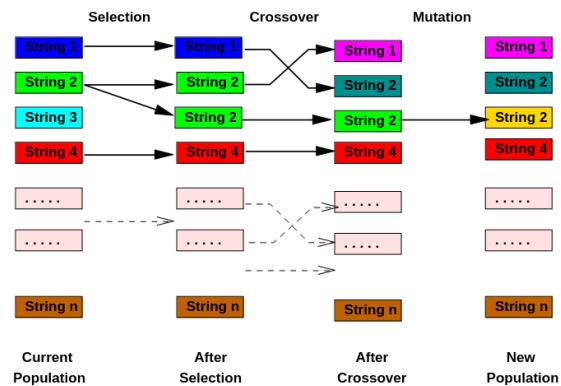


Figura 5.1: Figura obtenida de [66]

El algoritmo termina cuando se alcanza cierta tolerancia de error o número máximo de iteraciones.

Selección

La selección se puede llevar a cabo con varios métodos, uno de los más utilizados en selección por torneo.

En la selección por torneo, los individuos compiten entre sí en grupos pequeños. Cada grupo (torneo) tiene varios competidores (individuos). El competidor con la mejor aptitud (**fitness**) dentro del grupo tiene más probabilidad de ganar. Los ganadores son seleccionados para “reproducirse”.

Este proceso se repite hasta que se completa el *pool* de apareamiento con los ganadores de los torneos [68]. La probabilidad de cada individuo en base a su **fitness** se calcula de la siguiente manera:

$$P(\text{parent}_i) = \frac{\text{fitness}(\text{parent}_i)}{\sum_{j=1}^N \text{fitness}(\text{parent}_j)} \quad (5.1)$$

Para las versiones binaria y continua del algoritmo se utilizan diferentes operadores de mutación y cross-over.

Cruce binario

One-point crossover: Se elige al azar un punto en los cromosomas de ambos progenitores y se designa como "punto de cruce". Los bits a la derecha de ese punto se intercambian entre los dos cromosomas parentales. El resultado son dos descendientes, cada uno con información genética de ambos progenitores [69].

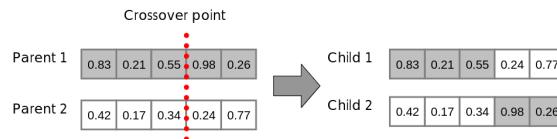


Figura 5.2: One-point crossover [70]

Cruce continuo

Blend crossover: Dado dos números reales para cada uno de los genes de los padres al hijo se le asignará un número aleatorio entre ese rango de gen para cada gen que conforme al vector cromosómico [70].

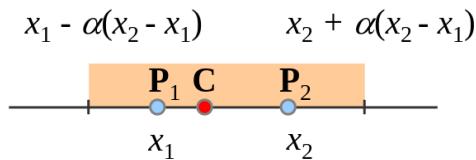


Figura 5.3: Blend crossover [70]

Mutación binaria

Su propósito es mantener la diversidad genética de los cromosomas de una población de un algoritmo evolutivo. El operador más básico y clásico consiste en cambiar un bit arbitrario de un genotipo o solución de un algoritmo genético binario a su estado inverso dada una probabilidad de mutación [67].

Mutación continua

Esta mutación se utiliza especialmente en problemas en los que se busca una exploración más amplia del espacio de búsqueda, ya que la distribución de Cauchy tiene colas pesadas y puede generar valores alejados del centro con mayor probabilidad en comparación con una distribución normal:

$$m(x_i) = \begin{cases} (2r)^{\frac{1}{\eta+1}} - 1 & \text{if } p \leq 0.5 \\ 1 - (2r)^{\frac{1}{\eta+1}} & \text{if } p > 0.5 \end{cases} \quad (5.2)$$

Esta función de mutación está diseñada para generar una exploración más amplia del espacio de búsqueda, con la posibilidad de generar valores más alejados del centro de la distribución. La variable r es un número a aleatorio y η es una variable de control.

Elitismo

El elitismo garantiza que las soluciones de alta calidad sobrevivan en las generaciones futuras, ayudando en la explotación del espacio de búsqueda. La cantidad de individuos élite seleccionados debe ser elegida con precisión [67].

5.2. Algoritmos de colonias de hormigas

5.2.1. Introducción

El algoritmo de colonia de hormigas o **ACO** es una técnica probabilística que trata de resolver problemas computacionales que pueden ser reducidos a la búsqueda de caminos óptimos a través de grafos. Se basa en el comportamiento de las hormigas reales y su comunicación vía feromonas. Las hormigas vagan por el mundo en búsqueda de comida de forma aleatoria. Cuando estas encuentran comida, dejan trazas de feromonas en su camino, de esta forma, si otras hormigas encuentran ese rastro, se hace más probable que recorran ese camino y dejen de vagar de forma aleatoria. Sin embargo, las trazas de feromonas se evaporan con el tiempo, perdiendo su fuerza de atracción sobre otras hormigas. De esta forma, si el camino es muy largo, la hormiga tardará

mucho en recorrerlo y la traza de feromonas tendrá más tiempo a evaporarse. De manera análoga, a más corto es el camino, más rápido se recorre y más densidad de feromonas acumula [71].

5.2.2. Funcionamiento y Operadores

Para poder adaptar el algoritmo ACO a un problema es necesario reducir ese problema a la búsqueda del camino más corto dentro de un grafo ponderado. De esta forma, la **representación** en este tipo de algoritmo debe ser la de un grafo. Cada característica del problema original se representa como un nodo n . Los caminos que conectan a los nodos (e) representan la elección del subconjunto de características. El camino deberá ser lo más corto posible maximizando a su vez el **accuracy** [71].

Regla de transición probabilística

$$P_{ij}^k(t) = \begin{cases} \frac{\tau_{ij}^\alpha \eta_{ij}^\beta}{\sum_l \tau_{il}^\alpha \eta_{il}^\beta} & \text{Si } l \text{ y } k \text{ son nodos admisibles} \\ 0 & \text{De lo contrario} \end{cases} \quad (5.3)$$

Donde:

- $P_{ij}^k(t)$ denota la probabilidad de transición de un nodo de i a j en la k -hormiga (agente) en el instante de tiempo t .
- τ_{ij} es la cantidad de traza de feromona en la arista (i, j) en el momento t . η_{ij} es la heurística de deseabilidad o visibilidad de la arista.
- β y α son dos parámetros que controlan la importancia relativa del valor de la feromona vs la información de la heurística.

Evaporación de feromona

Después de que todas las hormigas hayan terminado su camino, la evaporación de feromonas comienza. El contenido de feromonas del camino (i, j) en el instante $t + 1$ es:

$$\tau_{ij}(\text{new}) = (1 - p)\tau_{ij}(t) + \sum_{k=1}^m \Delta\tau_{ij}^k(t) + \Delta\tau_{ij}^g(t) \quad (5.4)$$

Donde:

- $\Delta\tau_{ij}^k = \begin{cases} \frac{Q}{F^k} & \text{Si la hormiga } k \text{ pasa por la arista } (i,j) \text{ en } T^k \\ 0 & \text{De lo contrario} \end{cases}$

- $p \in (0, 1]$ es el ritmo de evaporación.
- m es el número de hormigas.
- $\Delta\tau_{ij}^k$ y $\Delta\tau_{ij}^g$ son respectivamente, la cantidad de feromonas colocadas en la arista (i, j) por la hormiga k y la cantidad de feromonas depositadas por la mejor hormiga g en el instante t sobre la arista (i, j) .
- Q es una constante y F^k es el valor de coste de la solución encontrada por la hormiga k en el tour T^k , es decir, F^k es el *fitness* de la hormiga k .

Se utiliza el sistema *max-min*, solo la mejor hormiga puede depositar feromona.

En el algoritmo original, no había heurística de deseabilidad o visibilidad (η), solo información en términos de feromona [42].

5.3. Optimización por enjambre de partículas

5.3.1. Introducción

El algoritmo conocido como optimización por enjambre de partículas o en inglés *particle swarm optimization* fue concebido en 1995 por James Kennedy y Russel Eberhart [40]. En este algoritmo, un conjunto de soluciones candidatas, llamadas partículas, se mueven en un espacio de búsqueda multidimensional. Cada partícula ajusta su posición de acuerdo con su experiencia personal y la experiencia del grupo.

Para la versión binaria se utilizan funciones de transferencia.

5.3.2. Funcionamiento y Operadores

Las soluciones son descritas como individuos dentro de un enjambre (conjunto total de posibles soluciones) que tienen una posición y una velocidad en todo momento. La posición de estos individuos representa la solución en sí.

Velocidad

$$v_i(t+1) = v_i(t) + c_1 \cdot \text{rand}() \cdot (pbest_i - x_i(t)) + c_2 \cdot \text{rand}() \cdot (gbest - x_i(t)) \quad (5.5)$$

Donde:

- $v_i(t + 1)$ es la velocidad de la partícula i en la siguiente iteración.
- $v_i(t)$ es la velocidad de la partícula i en la iteración actual.
- c_1 y c_2 son factores de aceleración que controlan la influencia de las mejores posiciones.
- $\text{rand}()$ es un número aleatorio en el rango $[0, 1]$.
- $pbest_i$ es la mejor posición histórica de la partícula i .
- $x_i(t)$ es la posición actual de la partícula i en la iteración t .
- $gbest$ es la mejor posición global del grupo.

Posición

$$x_i(t + 1) = x_i(t) + v_i(t + 1) \quad (5.6)$$

Estas ecuaciones se utilizan para guiar el movimiento de las partículas a través del espacio de búsqueda hacia las mejores soluciones, adaptándose tanto a la mejor solución encontrada por la propia partícula ($pbest$) como a la mejor solución encontrada por cualquier partícula en el enjambre ($gbest$) [40]. Con el tiempo, las partículas tienden a converger hacia las mejores soluciones conocidas.

5.4. Evolución diferencial

La evolución diferencial surgió como un algoritmo de optimización para problemas no diferenciables y no lineales, con pocas variables de control, robusto, paralelizable y muy eficaz. El algoritmo de evolución diferencial o *differential evolution* en inglés fue diseñado para resolver problemas continuos, a diferencia del propósito original de su “hermano gemelo” el algoritmo genético, cuya finalidad inicial era resolver problemas binarios. Este algoritmo fue presentado en 1996 por Storn y Price en [43].

En la evolución diferencial, la diferencia de vectores es un concepto clave que se utiliza en el operador de mutación. Esta se calcula restando dos vectores de la población. Gracias a esta operación puede calcularse la distancia entre vectores y esto conlleva una serie de ventajas [43]:

- Las posiciones de los individuos proporcionan información ya que si los individuos están bien distribuidos (y si la población es lo suficientemente grande), la población inicial será una buena representación de todo el espacio de búsqueda.

- Las distancias entre estos individuos serán inversamente proporcionales al tamaño de la población.
- A medida que avanza la búsqueda y los individuos comienzan a converger hacia un óptimo local, las distancias entre los individuos comenzarán a disminuir.

Para la versión binaria se utilizan funciones de transferencia.

5.4.1. Funcionamiento y Operadores

La representación suele ser la misma que en otros algoritmos metaheurísticos. DE fue diseñado para optimizar un problema continuo, por lo que la población debería ser un conjunto de números reales.

Mutación

La operación de mutación produce un vector, conocido como vector de prueba o en inglés *trial vector*. Para ello se hace uso de un vector objetivo y una diferencia de vectores ponderada. La mutación elige a un parente x_i , generando un vector de prueba u_i siguiendo los siguientes pasos:

1. Se selecciona un vector objetivo x_{i_1} tal que $i \neq i_1$.
2. Se selecciona aleatoriamente dos individuos x_{i_2} y x_{i_3} de la población tal que $i \neq i_1 \neq i_2 \neq i_3$ y i_2, i_3 sean seleccionados con una probabilidad uniforme. Es necesario que todos los individuos tengan la misma probabilidad de selección.

Finalmente, se calcula el vector de prueba:

$$u_i = x_{i_1} + \beta(x_{i_2} - x_{i_3}) \quad (5.7)$$

Donde β es un factor de escalado que controla la amplificación de la variación diferencial (mutación más diversa/grande) [43].

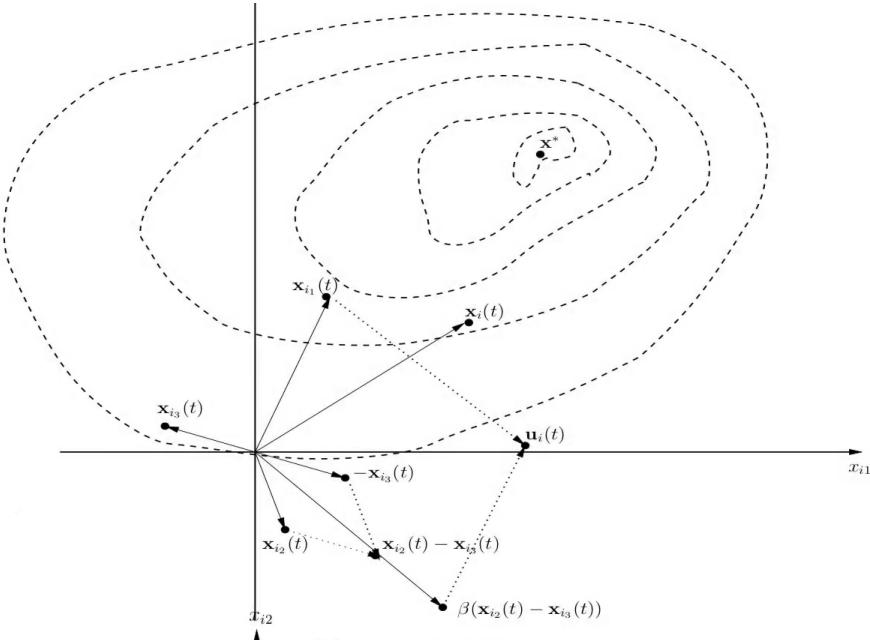


Figura 5.4: Esta figura ha sido seleccionada de [72]. Muestra de manera gráfica en un espacio bidimensional como funciona el operador de mutación y cómo se crea el “trial vector”.

Una ventaja que tiene el uso de diferencias de vectores en la mutación, es que si todos los individuos tienen la misma probabilidad de selección, la media de la distribución será cero. Con ello se consigue mantener la deriva genética lo más pequeña posible. La deriva genética es un proceso estocástico en el que la frecuencia de alelos (forma alternativa de un gen) en una población cambia aleatoriamente con el tiempo debido a eventos aleatorios. Esto puede llevar a cambios en la composición genética de la población a lo largo de las generaciones, especialmente en poblaciones pequeñas. Por tanto, una media distinta de cero daría pie a sesgos en la deriva genética [72, 73].

Cruce

El cruce implementa una recombinación del vector de prueba u_i y el vector padre x_i para producir un vector hijo x'_i .

$$x'_{ij} = \begin{cases} u_{ij} & \text{Si } j \in \mathcal{J} \\ x_{ij} & \text{De lo contrario} \end{cases} \quad (5.8)$$

Donde \mathcal{J} es un conjunto de puntos de perturbación. Estos índices pueden ser calculados de muchas formas, por ejemplo, de manera aleatoria.

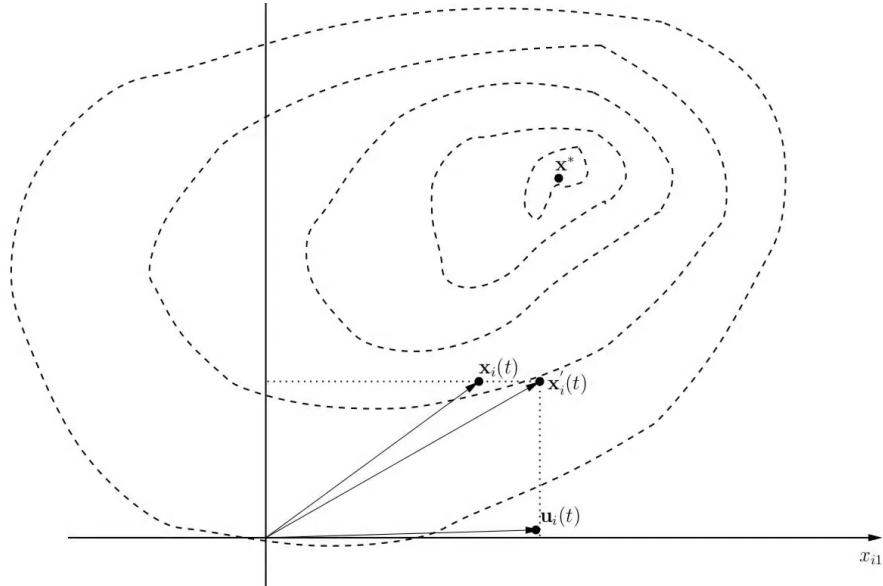


Figura 5.5: Esta figura ha sido seleccionada de [72]. Muestra de manera gráfica en un espacio bidimensional como funciona el operador de cruce al recombinar los genes/características de los vectores padre y “trial”.

Selección

Si el vector hijo, al evaluarlo, tiene un valor **fitness** mejor que el vector padre, entonces se sustituye el vector x_i por el vector x'_i . De otra forma, el vector padre continuará en la población para la siguiente generación [43].

5.5. Optimización por colonia de abejas artificial

5.5.1. Introducción

El algoritmo del enjambre de abejas artificial es un algoritmo estocástico evolutivo, al igual que los ya mencionados anteriormente, que se basa en el comportamientos de los enjambres de abejas para la optimización de problemas. El modelo usado en este algoritmo depende de tres factores, las fuentes de alimento, las abejas buscadores empleadas y las abejas buscadoras desempleadas (hay dos tipos) [44].

Para la versión binaria se utilizan funciones de transferencia.

5.5.2. Funcionamiento y Operadores

Los operadores del algoritmo son los siguientes:

1. **Abejas Empleadas (Employed Bees)**: Estas abejas están asociadas con fuentes de alimento específicas y son responsables de explotar estas fuentes. Llevan información sobre la fuente de alimento, incluyendo la distancia, la dirección y la rentabilidad, y comparten esta información dentro de la colmena, influyendo en las decisiones de otras abejas.
2. **Abejas Observadoras (Outlooker Bees)**: Estas abejas esperan dentro de la colmena y toman decisiones sobre a qué fuentes de alimento dirigirse basándose en la información proporcionada por las abejas empleadas. Seleccionan las fuentes de alimento según la probabilidad que está directamente relacionada con la rentabilidad de las fuentes.
3. **Abejas Exploradoras (Scout Bees)**: Son abejas que buscan nuevas fuentes de alimento sin información previa. Su rol es crucial para la exploración y el descubrimiento de nuevas fuentes de alimento. Cuando una fuente se agota o no mejora después de cierto número de intentos (determinado por un parámetro de control llamado "límite"), una abeja empleada puede convertirse en exploradora.

Estos operadores trabajan en un ciclo repetitivo donde las abejas empleadas primero van a las fuentes de alimento y actualizan la información sobre su rentabilidad. Luego, las abejas observadoras seleccionan fuentes de alimento basadas en esta información actualizada y las abejas exploradoras buscan nuevas fuentes [44]. Este proceso se repite hasta cumplir con los criterios de terminación del algoritmo.

5.6. Algoritmo de optimización del saltamontes

5.6.1. Introducción

Los **grasshoppers** (saltamontes o langosta cuando están en enjambre) son insectos cuyo ciclo de vida consiste en tres etapas: *huevo, ninfa, adulto*. Un saltamontes puede ser encontrado en el enjambre en su estado ninfa y adulto. La diferencia entre ninfa y adulto es la velocidad de movimiento con pasos pequeños en la ninfa y lo contrario en su vida adulta (Las soluciones adultas se encargan de la exploración y las ninfas de la explotación) [35].

Para la versión binaria se utilizan funciones de transferencia.

5.6.2. Funcionamiento y Operadores

Operador de posición (primera propuesta)

La posición de un individuo se define en la siguiente ecuación:

$$x_i = r_1 s_1 + r_2 g_i + r_3 a_i \quad (5.9)$$

Donde x_i es la posición del agente número i , s la interacción social, g la fuerza gravitacional, a la advección del viento y r_1, r_2, r_3 son números aleatorios entre $[0, 1]$. Esta función de actualización de posición no puede ser usada para la optimización debido a que se alcanza la zona de comfort de manera prematura y no se converge en un punto específico [35].

Posición (propuesta final)

$$X_i^d = c_1 \left(\sum_{j=1, j \neq i}^N c_2 \frac{ub_d - lb_d}{2} s(|x_j^d - x_i^d|) \frac{x_j - x_i}{d_{ij}} \right) + \hat{T}_d \quad (5.10)$$

Donde ud_b es la cota superior en la dimensión D , lb_d es la cota inferior en la dimensión D , \hat{T}_d es el valor de la dimensión D en el objetivo (la mejor solución encontrada hasta la fecha) y c es el coeficiente decreciente para hacer cada vez más pequeña la zona de comfort, zona de atracción y zona de repulsión.

No se considera la gravedad (G) y la dirección del viento siempre es a favor del objetivo \hat{T}_d .

Una diferencia con otros algoritmos de enjambres como PSO es que **GAO** actualiza la posición de cada saltamontes a partir de la posición actual, el mejor punto encontrado hasta el momento y la posición de todos los otros agentes, mientras que **PSO** lo hace a partir de la posición actual del agente y la mejor posición encontrada [35].

Esto significa que PSO no tienen en cuenta al resto de agentes, mientras que GAO sí.

Los parámetros c_1, c_2 tienen dos propósitos:

- $c_1 \rightarrow$ Reduce el movimiento del enjambre entero alrededor del objetivo, es decir, maneja la relación entre exploración y explotación para el conjunto global de soluciones.
- $c_2 \rightarrow$ Reduce la zona de comfort y las regiones de repulsión y atracción. s es la función que decide si un agente debe sentirse atraído

o repelido al objetivo, mientras que $\frac{ub_d - lb_d}{2}$ decrece de manera lineal el espacio de búsqueda a medida que c_2 decrece.

Ha de tenerse en cuenta también que en el proceso de búsqueda, la exploración debe ir antes que la explotación. En los saltamontes sin embargo, la fase explotativa que es la **ninfa** es la que se da primero. Por ello el parámetro c debe decrecer a medida que las iteraciones sobre el algoritmo van pasando. Este mecanismo hace que se promueva la explotación a medida que las iteraciones se incrementan, ya que se va reduciendo cada vez más el movimiento general del enjambre y se reduce la zona de comfort. c se actualiza por tanto así:

$$c = c_{\max} - l \cdot \frac{c_{\max} - c_{\min}}{L} \quad (5.11)$$

Donde l es la iteración actual y L el número máximo de iteraciones.

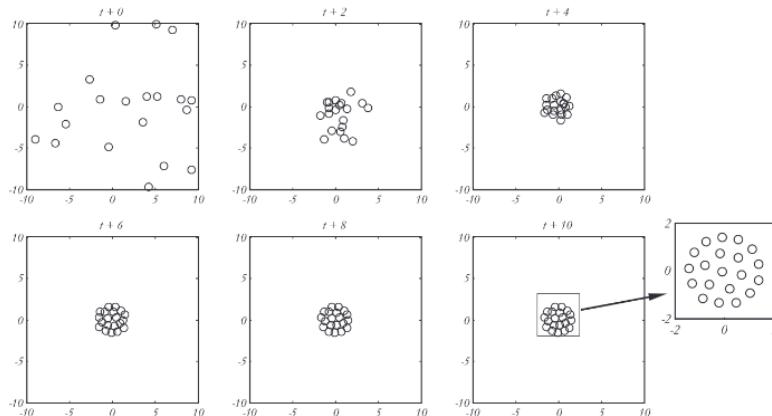


Figura 5.6: Esta figura ha sido seleccionada de [35]. Muestra la convergencia de los “saltamontes” en tan solo unas pocas iteraciones usando la fórmula 5.10.

5.7. Algoritmo de la libélula

5.7.1. Introducción

Los **dragonflies** (libélulas en español) son pequeños insectos que se pueden ver en enjambre solo cuando cazan o migran, dando lugar a dos tipos de enjambre:

- Las diferencias dadas entre ambos enjambres son notorias. En caza, se organizan en enjambres pequeños que vuelan en muchas direcciones en busca de comida. Este tipo de enjambre es llamado **estático**. Cuando

se organizan en enjambres para migrar, estos son numerosos y vuelan en una sola dirección. Estos enjambres en contraposición a los anteriores son llamados **dinámicos** [37].

- Es obvio el paralelismo que se da entre las principales fases de la búsqueda optimizatoria (**exploración y explotación**). La fase dinámica es un paralelismo con la exploración, ya que se buscan pequeños objetivos en grupos por el espacio. La fase estática (explotación) se da cuando todos los agentes empiezan a migrar a una dirección concreta.

Para la versión binaria se utilizan funciones de transferencia.

5.7.2. Funcionamiento y Operadores

Separación

$$S_i = - \sum_{j=1}^N X - X_j \quad (5.12)$$

Se refiere a la elusión estática de colisión entre individuos con su vecindario. Donde X es la posición del agente actual y X_j la posición de su vecino número j . Mientras que N es el número total de agentes.

Alineamiento

$$A_i = \frac{\sum_{j=1}^N V_i}{N} \quad (5.13)$$

Se refiere a la coordinación de velocidad entre individuos de un vecindario. Donde V_j es la velocidad del individuo número j en el vecindario.

Cohesión

$$C_i = \frac{\sum_{j=1}^N X_j}{N} - X \quad (5.14)$$

Se refiere a la fuerza de atracción de los individuos hacia el centro de masa del vecindario. Se da por hecho un radio alrededor de cada libélula agrupando su vecindario, pues este es muy importante para el comportamiento de esta. En un enjambre dinámico, las libélulas tienden a alinear su vuelo al tiempo que mantienen una separación y cohesión adecuadas.

En un enjambre estático, sin embargo, las alineaciones son muy bajas mientras

que la cohesión es alta para atacar a las presas. Por lo tanto, se asigna a las libélulas pesos de alta alineación y baja cohesión cuando exploran el espacio de búsqueda y de baja alineación y alta cohesión cuando explotan el espacio de búsqueda. Para la transición entre exploración y explotación, los radios de vecindad se incrementan proporcionalmente al número de iteraciones [37].

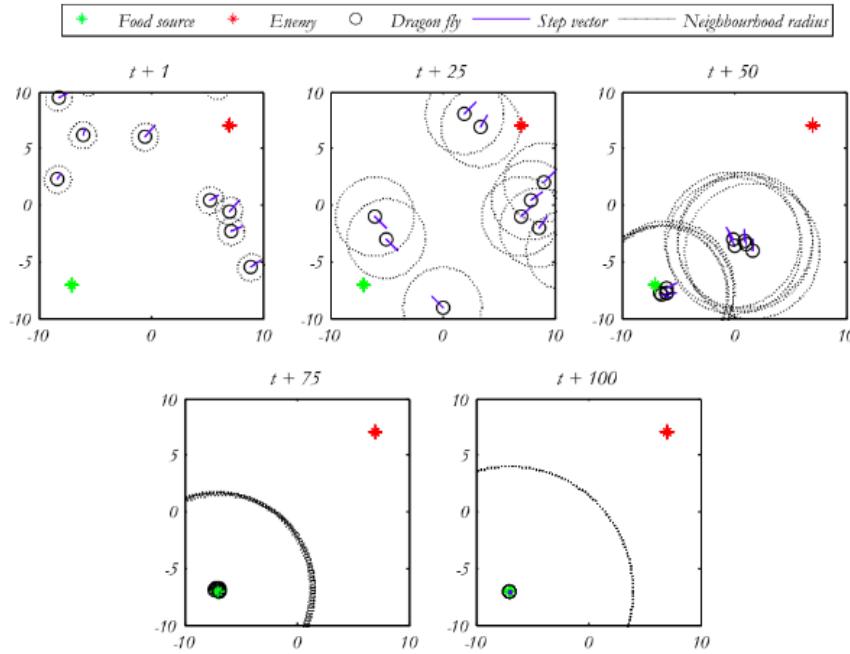


Figura 5.7: Vecindario de libélulas [37].

Atracción comida

$$F_i = X^+ - X \quad (5.15)$$

La fuerza de atracción hacia la comida (la mejor posición encontrada hasta el momento). Donde X^+ es la mejor posición encontrada hasta el momento.

Repulsión depredador

$$E_i = X^- + X \quad (5.16)$$

La fuerza de repulsión hacia un depredador (la peor posición encontrada hasta el momento). Donde X^- es la peor posición encontrada hasta el momento.

Dirección

$$\Delta X_{t+1} = (sS_i + aA_i + cC_i + fF_i + eE_i) + w\Delta X_t \quad (5.17)$$

El delta de X indica el vector dirección del movimiento de las libélulas. Cada elemento en minúscula es el factor del operador al que multiplica, de forma que es posible acentuar el efecto de cada uno de estos con múltiples combinaciones.

Posición

$$X_{t+1} = X_t + \Delta X_{t+1} \quad (5.18)$$

El operador de actualización de posición.

5.8. Algoritmo de optimización de ballenas

5.8.1. Introducción

Simula el comportamiento de las ballenas jorobadas en la caza, utilizando tanto el azar como el mejor agente de búsqueda para perseguir a la presa. Además, emplea una espiral para simular el mecanismo de ataque de la red de burbujas de las ballenas jorobadas [36].

Para la versión binaria se utilizan funciones de transferencia.

5.8.2. Funcionamiento y Operadores

Rodeado de presa

$$\vec{D} = |\vec{C} \cdot \vec{X}^*(t) - \vec{X}(t)| \quad (5.19)$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A} \cdot \vec{D} \quad (5.20)$$

Donde t indica la iteración actual, \vec{A} y \vec{C} son vectores de coeficientes, \vec{X}^* es el vector de posición de la mejor solución encontrada hasta el momento y \vec{X} es el vector posición.

Los vectores \vec{A} y \vec{C} se calculan como sigue:

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \quad (5.21)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (5.22)$$

Donde \vec{a} es decrementado linealmente desde 2 hasta 0 y \vec{r} es un vector aleatorio con valores en $[0, 1]$.

Ataque de red de burbujas (explotación)

- **Mecanismo de encogimiento del cerco** Este mecanismo es conseguido mediante el decremento del valor de \vec{a} . El vector \vec{A} es un vector de valores aleatorios en el intervalo $[-a, a]$ donde a es decrementado desde 2 hasta 0.

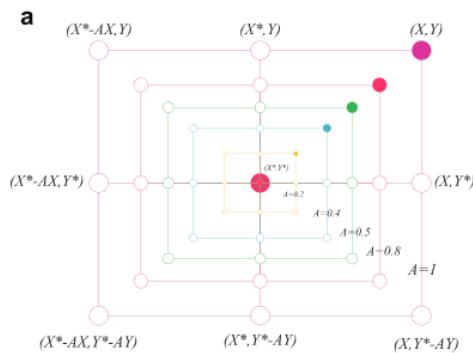


Figura 5.8: Ataque de red de burbujas usando el mecanismo de encogimiento del cerco [36].

- **Mecanismo de actualización de posición en espiral** Este mecanismo calcula la distancia entre la ballena (x, y) y la presa (x^*, y^*). Una ecuación con forma de espiral es creada entonces entre la ballena y la presa para simular el comportamiento de estas al cazar.

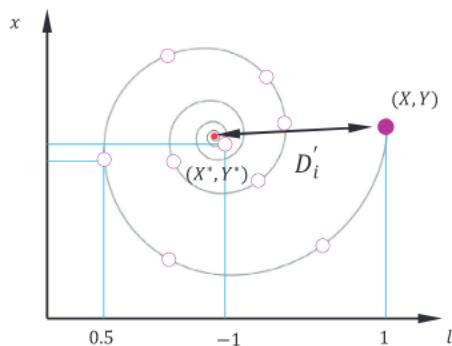


Figura 5.9: Ataque de red de burbujas usando el mecanismo de actualización de posición en espiral [36].

$$\vec{X}(t+1) = \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) \quad (5.23)$$

Donde $\vec{D}' = |\vec{X}^*(t) - \vec{X}(t)|$ e indica la distancia entre presa y ballena. b es una constante para definir la forma de la espiral, l es un número aleatorio entre $[-1, 1]$.

Se utilizan ambos mecanismos, dependiendo de si cierto número p aleatorio es mayor o menor a 0.5.

Búsqueda de la presa (exploración)

El mismo método utilizando el vector \vec{A} puede ser utilizado para la exploración. Se puede usar \vec{A} con valores más grandes a 1 o menores a -1 para forzar al agente a moverse lejos de otro agente.

$$\vec{D} = |\vec{C} \cdot \vec{X}_{rand} - \vec{X}| \quad (5.24)$$

$$\vec{X}(t+1) = \vec{X}_{rand} - \vec{A} \cdot \vec{D} \quad (5.25)$$

Cuando $|A| \geq 1 \rightarrow$ Exploración, cuando $|A| < 1 \rightarrow$ Explotación.

5.9. Algoritmo del murciélagos

5.9.1. Introducción

El algoritmo *metaheurístico* de los murciélagos se basa en la capacidad de eco-localización de estos para detectar a su presa y cazarla finalmente. La eco-localización de los murciélagos no solo les permite detectar a su presa, sino discriminar entre distintos tipos de insectos. Más concretamente, este algoritmo se basa en los micro murciélagos, pues estos usan más frecuentemente la eco-localización en comparación a otras especies de murciélagos. Estos murciélagos emiten un pulso de sonido, el cual, rebota en los objetos y entidades cercanas, permitiendo al murciélagos componer una “visión” general de sus alrededores. La frecuencia del pulso emitido podría estar relacionada con distintas estrategias de caza [39].

Acústica de la eco-localización

Aunque cada pulso dura solo unos pocos milisegundos (entre 8 y 10 ms), mantiene una frecuencia constante, generalmente entre 25 kHz y 150 kHz. La longitud de onda (λ) de estos impulsos ultrasónicos, con una frecuencia (f) y velocidad del sonido (v) en el aire a 340 m/s, está dada por

$$\lambda = \frac{v}{f} \quad (5.26)$$

con longitudes entre 2 mm y 14 mm para frecuencias de 25 kHz a 150 kHz, respectivamente. Esta longitud de onda coincide con el tamaño de las presas. Los murciélagos pueden detectar obstáculos tan pequeños como cabellos humanos, utilizando la eco-localización para construir un modelo 3D del entorno, y discriminando distancias, orientaciones y tipos de presas. Aunque los murciélagos tienen buena vista y olfato, utilizan principalmente la eco-localización para la detección de presas y la navegación eficiente [39].

Para la versión binaria se utilizan funciones de transferencia.

5.9.2. Funcionamiento y Operadores

Para simplificar:

- Los murciélagos usan la ecolocación para percibir la distancia y distinguir alimento de barreras.
- Vuelan aleatoriamente con velocidad v_i y frecuencia fija f_{min} , ajustando longitud de onda λ y sonido A_0 .
- Se asume que la intensidad del sonido varía desde A_0 hasta un mínimo A_{min} .

Movimiento

Las soluciones (posiciones) y las velocidades de los murciélagos en un instante i son:

$$\begin{aligned} x_i^t &= x_i^{t-1} + v_i^t \\ v_i^t &= v_i^{t-1} + (x_i^t - x_*) f_i \\ f_i &= f_{min} + (f_{max} - f_{min}) \beta \end{aligned} \quad (5.27)$$

Donde β es un vector aleatorio escogido de una distribución uniforme, x_* es la mejor posición global que es seleccionada tras comprobar los n murciélagos. Como el productor de $\lambda_i f_i$ es el incremento de la velocidad, se puede usar tanto uno como otro para ajustar la velocidad mientras el otro es fijado, todo depende del problema de interés [39]. En el documento original, se utiliza un rango para f de $[0, 100]$. A cada murciélagos se le asigna una frecuencia uniformemente de $[f_{min}, f_{max}]$. Una vez se ha seleccionado una mejor solución, se actualizan los vectores posición de cada murciélagos:

$$x_{\text{new}} = x_{\text{old}} + \epsilon A^t \quad (5.28)$$

Donde $\epsilon \in [-1, 1]$ es un número aleatorio y A^t es la intensidad sonora media de los murciélagos.

Intensidad sonora y Emisión de pulso

La intensidad sonora y el ratio de emisión de pulsos deben actualizarse según avance el número de iteraciones. A_i debe decrementarse a la vez que el ratio r_i se incrementa según el murciélagos se acerca a su presa [39]. Se actualizan del siguiente modo:

$$A_i^{t+1} = \alpha A_i^t, \quad r_i^{t+1} = r_i^0[1 - \exp(-\gamma t)] \quad (5.29)$$

Donde γ, α son constantes. Estos valores solo se actualizan si las nuevas soluciones son mejores, de manera que se guíe hacia un óptimo.

5.10. Optimización de los lobos grises

5.10.1. Introducción

- El algoritmo se inspira en el comportamiento de la manada de los lobos grises y sus escala jerárquica. Los lobos grises son super-depredadores, es decir, están en la cima de la cadena alimentaria [33]. Su escala jerárquica es la siguiente:
 1. $\alpha \rightarrow$ Los lobos alfa son aquellos que lideran la manada, no por su fuerza, si no por su capacidad organizativa dentro de la manada. Estos solo puede procrear dentro de la manada y son los encargados de tomar decisiones importantes sobre todo el conjunto.
 2. $\beta \rightarrow$ Los lobos beta son subordinados directos de los alfa. Son la opción más directa cuando se trata de sustituir a un alfa. Deben mostrar respeto a los alfa y comandan al resto de la manada según dicte el alfa.
 3. $\omega \rightarrow$ El nivel más bajo de la cadena. Son los últimos en comer, pero no por ello carecen de importancia, pues la manada puede llegar a enfrentarse a graves problemas cuando este nivel de lobos falta. Son totalmente necesarios para mantener la estructura.
 4. $\delta \rightarrow$ Lobos que someten a los omegas, pero que deben someterse a los alfas y omegas. A este grupo pertenecen los cazadores, ancianos, centinelas y cuidadores.
- Modelado de jerarquía:
 - Se considera la solución más buena como la solución α , siendo las siguientes en orden de mejor a peor **fitness** la β , la δ y por último la ω .
 - La caza (optimización) es guiada por las soluciones α, β y δ mientras que las soluciones ω siguen a estos tres lobos.

Para la versión binaria se utilizan funciones de transferencia.

5.10.2. Funcionamiento y Operadores

Rodeo de la presa

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)| \quad (5.30)$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \quad (5.31)$$

Donde t indica la iteración actual, \vec{A} y \vec{C} son vectores de coeficientes, \vec{X}_p es el vector posición de la presa y \vec{X} el vector posición del lobo gris. Los vectores \vec{A} y \vec{C} se calculan:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \quad (5.32)$$

$$\vec{C} = 2 \cdot \vec{r}_2 \quad (5.33)$$

Donde los componentes del vector \vec{a} son decrementados de 2 a 0 linealmente según avanzan las iteraciones. Las variables r_1 y r_2 son vectores aleatorios entre $[0, 1]$.

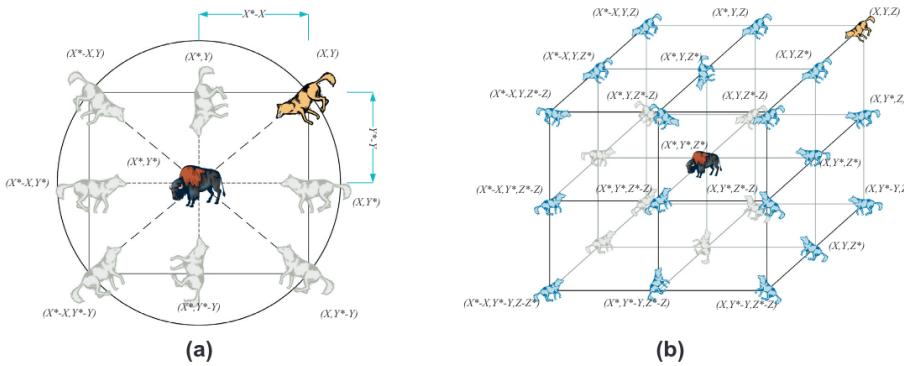


Figura 5.10: Esta figura ha sido seleccionada de [33]. Muestra en espacios 2D (a) y 3D (b) diferentes posiciones que un lobo puede tomar alrededor de la presa dada la ecuación 5.31. Define una vecindad alrededor de la solución en forma de hiper-esfera.

Caza

Se supone que el alfa, beta y delta tienen mayor conocimiento de donde está la presa (óptimo) y por tanto la posición del resto de lobos debe actualizarse

siguiendo a estos tres primeros [33].

$$\vec{D}_\alpha = |\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}|, \quad \vec{D}_\beta = |\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}|, \quad \vec{D}_\delta = |\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}| \quad (5.34)$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \vec{D}_\alpha, \quad \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta, \quad \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta \quad (5.35)$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (5.36)$$

Ataque a la presa (explotación)

Cuando los lobos están muy cerca de la presa (solución óptima), el valor de A disminuye, llevando a los lobos a converger hacia la presa. Esto se refleja en una reducción en la amplitud de los saltos en el espacio de búsqueda, facilitando una búsqueda local intensiva cerca de la mejor solución encontrada [33].

Búsqueda de la presa (exploración)

Para fomentar la exploración y evitar los óptimos locales, los vectores A pueden tomar valores que obliguen a los lobos a alejarse de la presa, explorando nuevas áreas del espacio de búsqueda. Este comportamiento se activa cuando los elementos de A son mayores que 1 o menores que -1 , lo que amplía la capacidad del algoritmo para explorar el espacio de búsqueda de manera más amplia [33].

5.11. Algoritmo de la luciérnaga

5.11.1. Introducción

Las luciérnagas son insectos capaces de producir un brillo único, muy visible y característico sobre todo de noche. El brillo que estas producen puede llegar a tener significados que hoy día siguen debatiéndose en la comunidad científica. Pese a ello, se conoce que este brillo tiene dos funciones fundamentales y bien definidas:

1. **Comunicación:** Sirve como instrumento para la comunicación entre dos luciérnagas para el apareamiento.
2. **Atracción:** Se usa para atraer a potenciales presas de la luciérnaga.

Por simplicidad del algoritmo, se asumen tres reglas sobre el comportamiento y naturaleza de las luciérnagas [34]:

1. Las luciérnagas solo tienen un sexo, por lo que todas se ven atraídas por todas y no hay discriminación más allá de que una luciérnaga sea más atractiva que otra.
2. La luminosidad es proporcional a lo atractiva que es una luciérnaga para el resto.
3. La luminosidad de una luciérnaga es determinada por los límites de la función objetivo.

Para la versión binaria se utilizan funciones de transferencia.

5.11.2. Funcionamiento y Operadores

Intensidad de la luz

$$\beta = \beta_0 e^{-\gamma r^2} \quad (5.37)$$

Donde β_0 es el atractivo en $r = 0$, es decir, en distancia igual a 0 y γ es un coeficiente fijo de absorción de luz [34]. La distancia entre dos luciérnagas es la distancia cartesiana:

$$r_{ij} = \|x_i - x_j\|_2 = \sqrt{\sum_{k=1}^d (x_{i,k} - x_{j,k})^2} \quad (5.38)$$

Movimiento

El movimiento de un agente (luciérnaga) es descrito por la siguiente ecuación, que según aumente el brillo y con ello el atractivo, hace que el vector posición se dirija hacia la luciérnaga atractiva:

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{ij}^2} (x_j^t - x_i^t) + \alpha \epsilon_i^t \quad (5.39)$$

Donde α es un parámetro de aleatoriedad y ϵ un vector de números alatorios sacados de una distribución Gaussiana o uniforme.

El parámetro α debe ser decrementado gradualmente de forma que las luciérnagas pasen de moverse tan aleatoriamente en un principio a converger en un punto óptimo al final [34]:

$$\alpha = \alpha_\infty + (\alpha_0 - \alpha_\infty) e^{-t} \quad (5.40)$$

Donde $t \in (0, t_{max}]$, siendo t_{max} el número máximo de generaciones, α_0 el parámetro inicial y α_∞ el valor final. De esta forma la exploración da paso a la explotación a medida que el número de iteraciones avanza.

5.12. Búsqueda Cuckoo

5.12.1. Introducción

El algoritmo de búsqueda de los cucos se basa en el tipo de ave de cuyo nombre se inspira el algoritmo. Se basa en el tipo de crianza de estos pájaros y su interesante comportamiento parasitario, así como en los vuelos Lévy [74].

El parasitismo de cría que adoptan estos pájaros implica poner huevos en nidos de otras aves. Los tipos incluyen parasitismo intra-específico, cría cooperativa y toma de nidos. Las aves anfitrionas pueden rechazar los huevos ajenos, lo que las lleva a abandonar el nido o construir uno nuevo. Algunos cucos parasitarios imitan los huevos de las aves anfitrionas para evitar el rechazo y aumentar el éxito reproductivo [38].

Para la versión binaria se utilizan funciones de transferencia.

5.12.2. Funcionamiento y Operadores

En el algoritmo de búsqueda cuco se implementa un tipo de vuelo Lévy, debido a que se ha observado que el vuelo de ciertos insectos y aves sigue este tipo de movimiento [38].

El algoritmo de búsqueda cuco (Cuckoo) se basa en tres reglas ideales simplificadas:

1. **Puesta de huevos:** Cada cuco pone un huevo a la vez y lo deposita en un nido elegido al azar. Nuevo huevo en nido $x_i^{t+1} = x_i^t + \alpha \oplus Levy(\lambda)$, siendo α el tamaño de paso y siendo el vuelo Lévy:

$$Levy \sim u = t^{-\lambda} \quad (5.41)$$

Siendo t la iteración actual y λ un parámetro de estabilidad. El símbolo \oplus se refiere a una operación de multiplicación entrada por entrada, lo que significa que cada entrada de dos matrices (o vectores) se multiplica una con la otra, resultando en una nueva matriz (o vector) con las mismas dimensiones [38].

2. **Selección de nidos:** Los nidos con huevos de alta calidad se pasan a las generaciones siguientes. Si $f(x_j) > f(x_i)$, entonces $x_i = x_j$.
3. **Descubrimiento de huevos:** Los huevos son descubiertos por aves anfitrionas con una probabilidad p_a . Probabilidad de descubrimiento: $p_a \in [0, 1]$

Cuando un huevo es descubierto, el ave anfitriona puede desecharlo o abandonar el nido y construir uno nuevo $\rightarrow rand() < p_a$. El objetivo es usar las

nuevas soluciones (huevos de cuco) para reemplazar soluciones no tan buenas en los nidos [38].

Capítulo 6

Diseño experimental

En este capítulo se describirá el proceso de experimentación llevado a cabo, los conjuntos de datos usados para los distintos experimentos y los parámetros usados en cada uno de los algoritmos.

Los conjuntos de datos usados son lo siguientes:

Dataset	Instancias	Características	Clases	Área
breast-cancer	286	9	2	Medicina
ecoli	336	7	8	Biología
parkinsons	200	22	2	Medicina
spectf-heart	348	44	2	Medicina
wine	182	13	3	Alimentación
dermatology	366	34	6	Medicina
ionosphere	350	34	2	Meteorología
sonar	207	60	2	Biología
waveform5000	5000	40	3	Física
yeast	1483	8	10	Biología
diabetes	768	8	2	Medicina
iris	149	4	3	Biología
spambase-460	459	54	2	Informática
wdbc	568	29	2	Medicina
zoo	101	18	7	Biología

Tabla 6.1: Información de los conjuntos de datos

Se han seleccionado los conjuntos de datos descritos en la tabla 6.1 por los siguientes motivos:

- **Variedad de áreas:** Se ha buscado representar una amplia gama de áreas, como Medicina, Biología, Alimentación, Meteorología, Física e

Informática. Esto asegura que los problemas abordados sean variados y abarquen diferentes dominios de aplicación.

- **Diversidad de problemas:** Los conjuntos de datos elegidos cubren una variedad de problemas, desde diagnósticos médicos y clasificación de especies hasta predicciones en meteorología y detección de spam. Esta diversidad permite evaluar el rendimiento de los algoritmos en múltiples contextos.
- **Número de características:** Algunos conjuntos de datos, como *spectf-heart* y *spambase-460*, tienen un gran número de características, lo que es particularmente útil para evaluar la eficacia de las técnicas de selección de características. Esta variedad en el número de características permite probar la robustez de los algoritmos en situaciones de alta dimensionalidad. La inclusión de otros como *iris* con tan pocas características hacen de conjunto de datos de referencia, todos los algoritmos deberían optimizar este *dataset*.
- **Popularidad y uso en investigaciones:** Los conjuntos de datos seleccionados son populares y ampliamente utilizados en estudios y artículos de referencia [52, 54, 56, 75]. Esto facilita la comparación de resultados con otros trabajos y asegura que los datos son reconocidos y aceptados en la comunidad científica.
- **Variedad en el número de instancias y clases:** La selección incluye conjuntos de datos con diferentes números de instancias y clases. Esta variedad es crucial para probar la escalabilidad y versatilidad de los algoritmos.
- **Relevancia práctica:** Muchos de estos conjuntos de datos, como *diabetes* y *breast-cancer*, tienen una alta relevancia práctica en sus respectivas áreas, lo que subraya la importancia de encontrar soluciones eficientes y precisas para problemas del mundo real.

Los datos han sido normalizados entre 0 y 1 siguiendo la función de normalización *MinMaxScaler* de *scikit-learn* [76]. Los conjuntos de datos se han dividido en 20 % test, 20 % validación y 60 % entrenamiento. Se han llevado a cabo 10 ejecuciones sobre cada algoritmo en cada conjunto de datos, tanto en su versión original como en su versión binaria. Estas ejecuciones se han procesado de forma paralela en el servidor de *Hércules* [8].

Se han recabado datos de todas las ejecuciones para poder comparar los resultados. En forma de gráficos se ha representado la curva de convergencia de cada algoritmo además de una curva de decrecimiento de las características de cada conjunto de datos. Ambos gráficos son muy interesantes a la hora de comparar algoritmos para un mismo conjunto de datos.

Además se han recabado los siguientes datos en ficheros de texto para su posterior análisis:

- **classifier**: Clasificador usado (kNN o SVC).
- **dataset**: El conjunto de datos utilizado para el proceso de selección de características.
- **optimizer**: Algoritmo metaheurístico usado como optimizador.
- **all_fitness**: Lista que contiene los valores *fitness* de todas las soluciones generadas durante la optimización.
- **best**: Mejor valor encontrado durante la optimización.
- **avg**: Promedio de los valores de *fitness* de todas las soluciones generadas durante la optimización.
- **std_dev**: Desviación estándar de los valores de *fitness* de todas las soluciones generadas durante la optimización. Indica qué tan dispersos están los valores alrededor del promedio.
- **acc**: Precisión (accuracy) del modelo de clasificación, es decir, la proporción de instancias correctamente clasificadas sobre el total de instancias.
- **n_features**: Número de características seleccionadas.
- **selected_rate**: Tasa de selección de características, que representa la proporción de características seleccionadas respecto al total de características disponibles.
- **execution_time**: Tiempo de ejecución del algoritmo optimizatorio.

Para la función de *fitness* se ha usado:

$$\text{fitness} = \text{acc} \cdot \alpha + \text{red} \cdot (1 - \alpha) \quad (6.1)$$

Donde *acc* quiere decir precisión y *red* el porcentaje de características reducidas. La variable α actúa como medio de ponderación y ha sido fijado en $\alpha = 0.9$, dando un 90 % de prioridad a la precisión. Se procede a mostrar los parámetros fijados para cada algoritmo (como referencia se han seguido los parámetros de los artículos originales y binarios ya citados para cada algoritmo).

Los parámetros fijados en cada uno de los algoritmos de optimización metaheurísticos han sido elegidos por las recomendaciones dadas en los artículos de referencia de los que se sacan, más concretamente los artículos que presentan la versión binaria. Estos son citados en la descripción de cada tabla.

Parámetro	Valor
c_{min}	0.00001
c_{max}	1
F	0.5
L	1.5

Tabla 6.2: Parámetros de GOA [56]

Parámetro	Valor
Parámetro espiral	1

Tabla 6.3: Parámetros de WOA. En el paper original [36] y en el de la versión binaria [54] no se da un valor. El valor por defecto ha sido sacado de la implementación de [77]

Parámetro	Valor
Abeja empleada	3
Abeja vigilante	3
Límite	3

Tabla 6.4: Parámetros de ABCO [44]

Parámetro	Valor
α	0.9
γ	0.9
f_{min}	0
f_{max}	2

Tabla 6.5: Parámetros de BA [57]

Parámetro	Valor
w	0.9
c_1	2
c_2	2

Tabla 6.6: Parámetros de PSO [60]

Parámetro	Valor
α_0	0.5
β_0	0.2
γ_0	1

Tabla 6.7: Parámetros de FA [59]

Parámetro	Valor
Ratio de cruce	1
Ratio de mutación	0.05
Elite	2
η	1
α	$\sqrt{0.3}$

Tabla 6.8: Parámetros de GA. Dada la cantidad de variantes y cantidad de valores de parámetros propuestos, estos han sido escogidos de manera empírica y basándose en recomendaciones del tutor del proyecto.

Parámetro	Valor
α	1
Q	1
Feromonía inicial	0.1
Ratio de evaporación	0.049

Tabla 6.9: Parámetros de ACO [71]

Parámetro	Valor
Ratio de descubrimiento	0.25
α	1
λ	1.5

Tabla 6.10: Parámetros de CS [58]

Parámetro	Valor
F	0.5
Cr	0.1

Tabla 6.11: Parámetros de DE [43]

Los algoritmos de **DA** y **GWO** no tienen parámetros configurables.

Capítulo 7

Análisis y resultados

En este capítulo se van a discutir los resultados obtenidos tras realizar los experimentos descritos en el capítulo anterior. Se mostrarán gráficas que ilustren la curva de convergencia de los algoritmos en algunos conjunto de datos, para cada clasificador (kNN y SVC) y en cada versión del algoritmo, es decir, original y versión binaria. Junto a eso también se han elaborado una gráficas de “cajas y bigotes” o *boxplots* que ilustran no solo los mejores resultados, sino su robustez y estabilidad tras varias iteraciones.

Además se mostrarán tablas y gráficas de los ranking obtenidos en versión binaria y en versión continua para cada clasificador tras hacer la media de los resultados en todos los conjuntos de datos.

Por último, los resultados vendrán expuestos en forma de tabla junto a las métricas correspondientes, de forma que se facilite el análisis de los algoritmos, de las versiones reales y continuas y de los clasificadores usados.

También se realizarán análisis estadísticos de los resultados obtenidos utilizando como herramienta la página web **TACOlаб** [78], facilitada por el tutor del proyecto, de forma que se puedan hacer los mejores y más robustos tests de forma automática y sencilla.

Como detalle a tener en cuenta, se ha añadido en las comparaciones un optimizador al que se le ha llamado “Dummy”, debido a su naturaleza totalmente aleatorio y sin criterio alguno. Se añade como algoritmo de referencia, ninguna otra metaheurística debería ser peor que esta en promedio. En caso contrario podría decirse que el algoritmo en cuestión está mal diseñado.

7.1. Binario

En esta sección se discutirán y analizarán los resultados para las metaheurísticas de codificación binaria o discreta. Estas versiones tienen un enfoque del problema distinto al de las reales, pues codifican cada característica con un 0 o un 1, o lo que es lo mismo, descartan características. Dada su naturaleza, cabría pensar que son los algoritmos más específicos que podrían usarse para el problema de selección de características.

7.1.1. General

En esta sub-sección se presentarán los resultados generales obtenidos con las metaheurísticas binarias. Esto permitirá una visión más amplia de su desempeño y proporcionará intuiciones iniciales sobre la efectividad de cada algoritmo en distintos contextos.

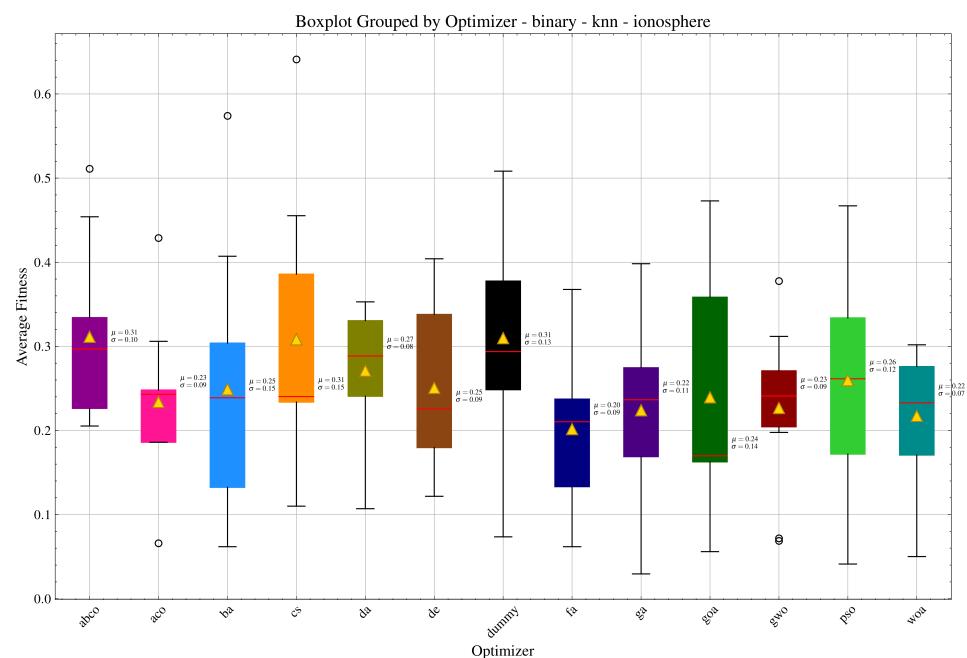


Figura 7.1: Boxplot ionosphere en kNN

Como ya se aclaró en capítulos anteriores, para obtener estos resultados se han ejecutado 10 veces cada algoritmo en cada conjunto de datos.

Se muestran algunos de los resultados obtenidos con los algoritmos binarios en 7.1, 7.2, 7.3 y 7.4 con el clasificador **kNN**. Obviamente, no es posible incluir gráficas para todas las combinaciones posibles sin saturar el capítulo; para ello, se remite al lector al capítulo de apéndices.

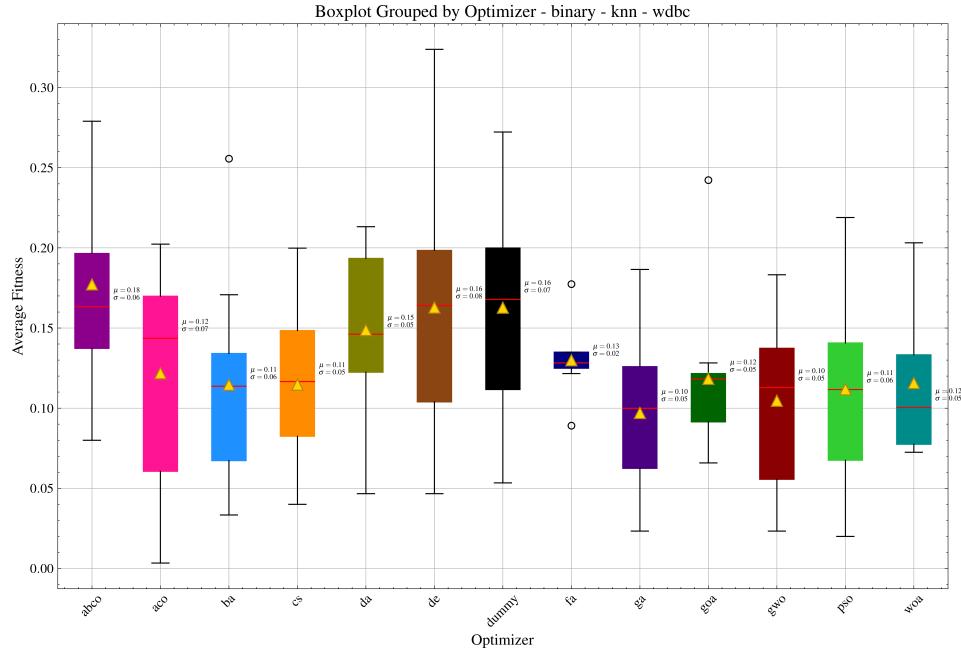


Figura 7.2: Boxplot wdbc en kNN

Se puede observar que la mayoría de los algoritmos obtienen resultados bastante buenos en los distintos conjuntos de datos, considerándose los valores de fitness suficientemente buenos para considerarlos soluciones de calidad, obviamente considerando cada problema por separado, con las características de cada uno. Es importante destacar que no todos los conjuntos de datos presentan la misma dificultad, ya que distintos tipos de problemas presentan diferentes desafíos. En particular, se puede notar que los algoritmos aplicados al conjunto de datos *ionosphere* (figura 7.1) muestran menos variabilidad en sus valores de fitness, reflejándose en una desviación estándar menor en comparación con los resultados de *diabetes* (figura 7.3) o *parkinsons* (figura 7.4). Con ello, queda reflejado como a los algoritmos les cuesta menos optimizar ciertos problemas y suelen ser más estables en estos. Esto puede deberse muchos factores, por ejemplo, mayor cantidad de mínimos locales en algunos problemas en comparación a otros.

También se pueden identificar ciertos algoritmos que parecen tener un mejor desempeño en general. Por ejemplo, algoritmos como **GWO** y **PSO** tienden a obtener resultados muy buenos, mientras que otros como **ACO** y **ABCO** parecen ser menos efectivos en los conjuntos de datos analizados.

En los resultados obtenidos con el clasificador **SVC**, se observa un rendimiento ligeramente mejor en comparación con **kNN**, al menos en los conjuntos de datos presentados. Esta mejora inicial sugiere que **SVC** puede ser más efectivo en ciertos contextos. Los algoritmos dominantes continúan mostrando un

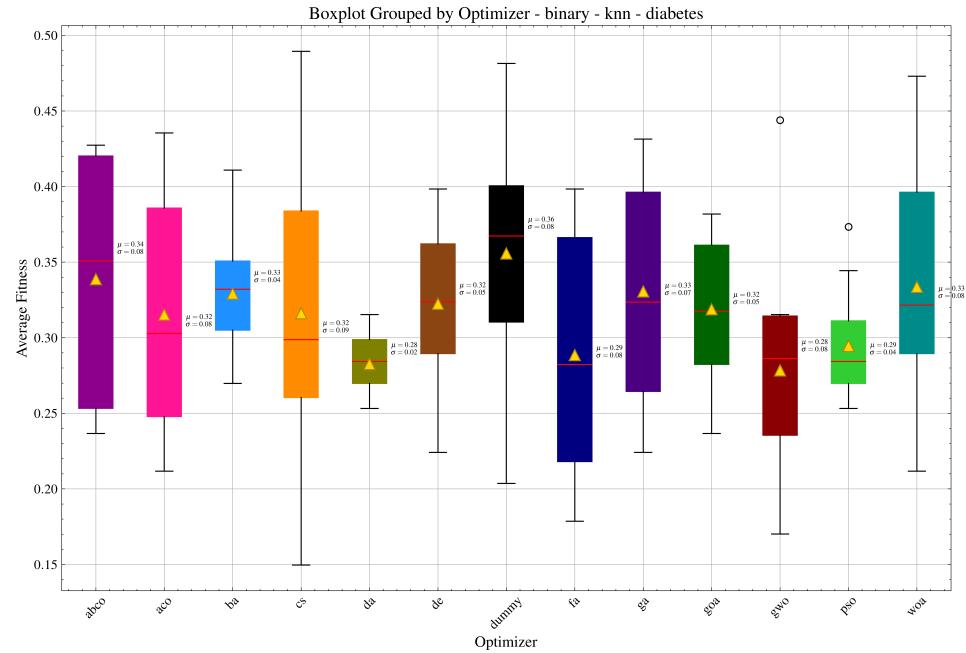


Figura 7.3: Boxplot diabetes en kNN

rendimiento superior, como era de esperarse, aunque se pueden apreciar algunas mejoras en la variabilidad de algoritmos específicos. Por ejemplo, **WOA** muestra una menor variabilidad en el conjunto de datos *parkinsons*, como se puede observar en la figura 7.8. Por el contrario, otros algoritmos como **GWO** parecen ser menos estables en el mismo conjunto de datos, indicando una mayor variabilidad en su desempeño.

Aunque estas observaciones iniciales proporcionan algunas ideas sobre el comportamiento de los algoritmos, es importante tener en cuenta que no se pueden sacar conclusiones definitivas solo a partir de las gráficas presentadas. Para obtener una comprensión más profunda y precisa del rendimiento de cada algoritmo, se deben realizar análisis estadísticos adicionales y considerar variaciones en los conjuntos de datos.

Por lo tanto, en las siguientes secciones, se analizarán métricas específicas como el fitness, la precisión, la tasa de reducción y el tiempo para obtener una comprensión más completa del desempeño de los algoritmos en estudio.

7.1.2. Fitness

Se procede a comparar y analizar los distintos algoritmos basándose en la métrica de *fitness*. Como se explica en la ecuación 6.1, este valor está compuesto por otras dos métricas a tener en cuenta, la precisión o *accuracy*

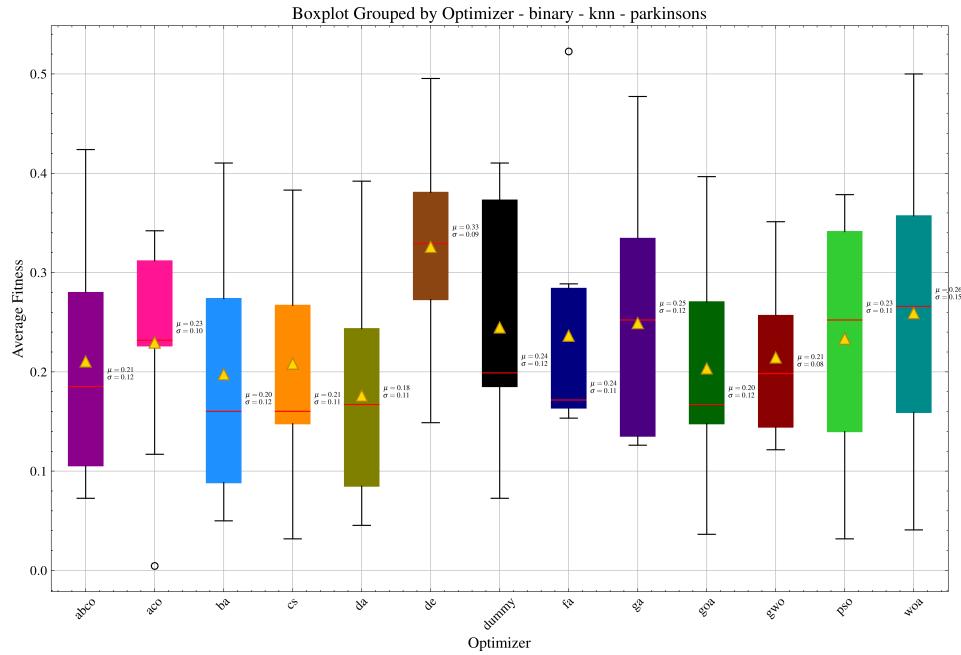


Figura 7.4: Boxplot parkinsons en kNN

(no confundir con *precision*) y el ratio de reducción de características. La métrica de *fitness* es el marcador más importante para comprobar la calidad de un algoritmo. Pese a ello, más tarde se analizarán ambas métricas que lo componen por separado, pues es interesante obtener una visión más precisa de los resultados.

Debe tenerse en cuenta que, al estar compuesto el *fitness* en un 90 % por el *accuracy*, ambas métricas estarán altamente relacionadas entre sí en la gran mayoría de los casos. Aún así, ha de tenerse en cuenta el 10 % restante.

Como puede observarse en la tabla 7.2, los algoritmos que consiguen mejores resultados son **ACO**, **GWO**, **PSO**, **FA**, **DE**, **GA**, **CS** y **WOA** con el clasificador **SVC**. Los más destacables en cuanto al número de veces que estos se repiten por ser los mejores por *dataset* son **CS**, **ACO**, **DE** y **GWO**.

Las tablas de convergencia para **SVC** presentadas en las figuras 7.10 y 7.11 ofrecen una visión detallada del desempeño de cada algoritmo en los diferentes conjuntos de datos. Es evidente que el algoritmo de referencia, **Dummy**, muestra consistentemente el peor rendimiento en la mayoría de los casos, con mejoras mínimas o nulas.

Es interesante destacar que el algoritmo **ACO** exhibe variabilidad en su rendimiento, siendo el mejor en algunos conjuntos de datos y el peor en

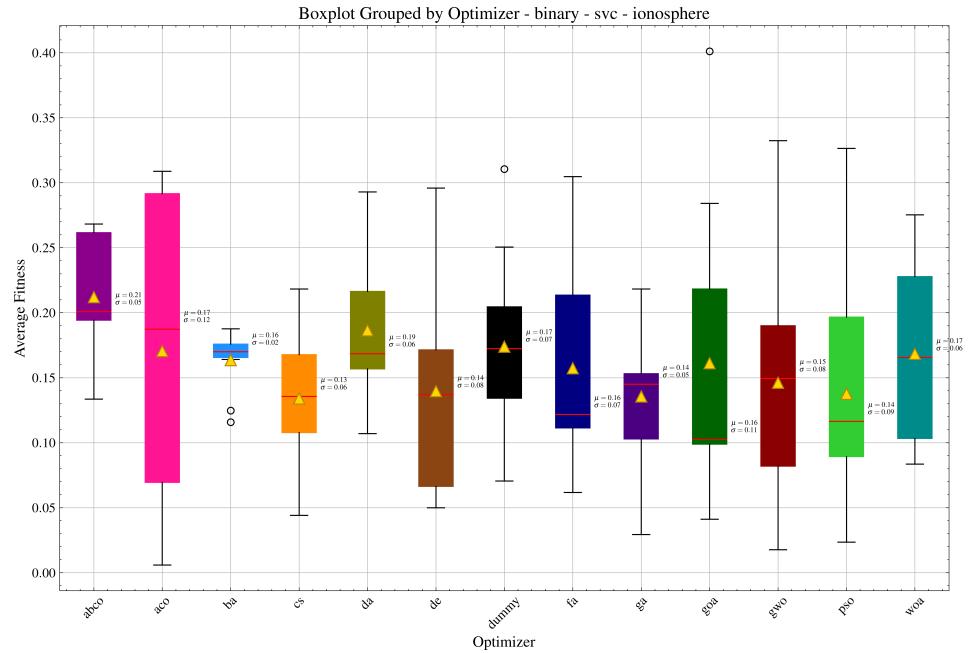


Figura 7.5: Boxplot ionosphere en SVC

otros. Resulta notable su inferioridad sobre el algoritmo aleatorio en algunos problemas tales como *waveform5000* o *yeast*. Estos observaciones pueden relacionarse con el principio del *no free lunch*, discutido en la sección 3.4, el cual establece que no existe un algoritmo universalmente óptimo para todos los problemas. En cambio, cada problema requiere una solución específica, y es improbable encontrar un algoritmo que supere a todos los demás en todos los contextos.

Clásicos vs Modernos

Los algoritmos clásicos de optimización como el **PSO**, **GA** o **DE** han demostrado ser extremadamente efectivos en una amplia gama de problemas. Su capacidad para explorar y explotar el espacio de búsqueda ha llevado a soluciones de alta calidad en diversas aplicaciones, desde la ingeniería hasta la economía. Estos métodos son conocidos por su robustez, simplicidad y eficacia, lo que los convierte en herramientas valiosas para investigadores y profesionales.

Sin embargo, en los últimos años, han surgido nuevos algoritmos de optimización que prometen mejorar aún más estos resultados. Algoritmos como el **WOA**, **GWO** y **CS** han sido diseñados para abordar algunas de las limitaciones de los métodos clásicos, como la convergencia prematura y la exploración insuficiente del espacio de búsqueda. Estos algoritmos modernos

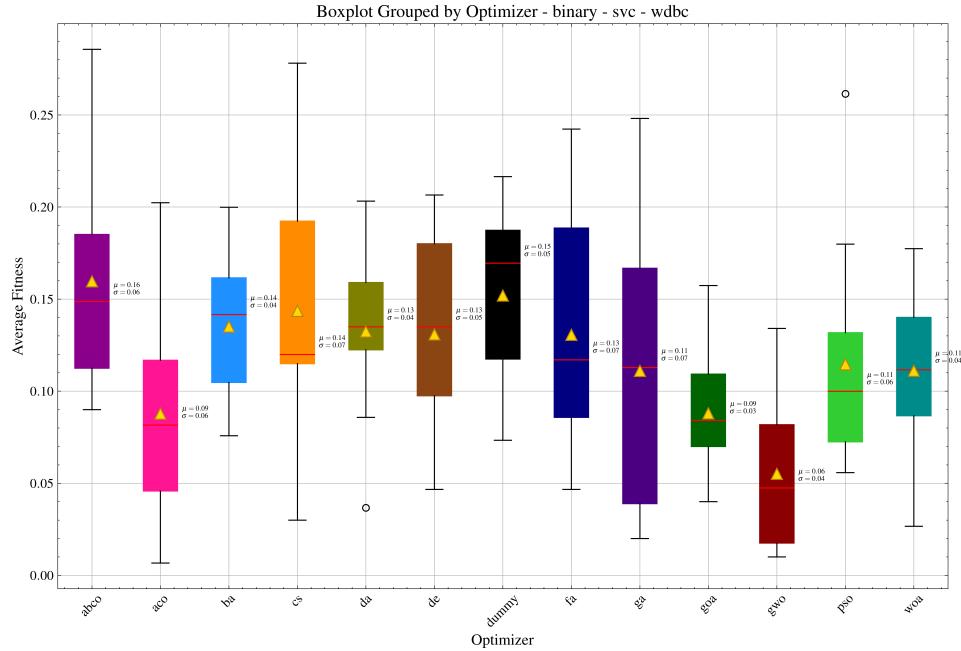


Figura 7.6: Boxplot wdbc en SVC

incorporan nuevas estrategias y mecanismos de búsqueda inspirados en la naturaleza, que buscan mejorar la eficiencia y la precisión de las soluciones obtenidas.

Los rankings han sido creados basándose en los valores *fitness* de los algoritmos para todos los conjuntos de datos. De forma visual es capaz de verse de un vistazo aquellos más versátiles, los que mejor se han comportado en distintos escenarios.

En los rankings que se pueden observar en las gráficas 7.9a para **kNN** y en 7.9b para **SVC** puede verse como el mejor algoritmo es el **GWO**, el algoritmo de los lobos grises, considerado como moderno. Puede observarse además el valor *fitness* obtenido en las tablas 7.3 y 7.2, con gran diferencia en comparación al resto. Le siguen algoritmos como **PSO**, **CS** y **GA**.

El **PSO**, a pesar de ser uno de los algoritmos más antiguos en la optimización metaheurística, sigue manteniendo una posición destacada en los rankings. Estos hechos junto a su simplicidad y efectividad en la convergencia hacia soluciones óptimas, hacen ver que **PSO** sigue siendo una opción muy interesante, incluso en problemas de tipo binario como la selección de carac-

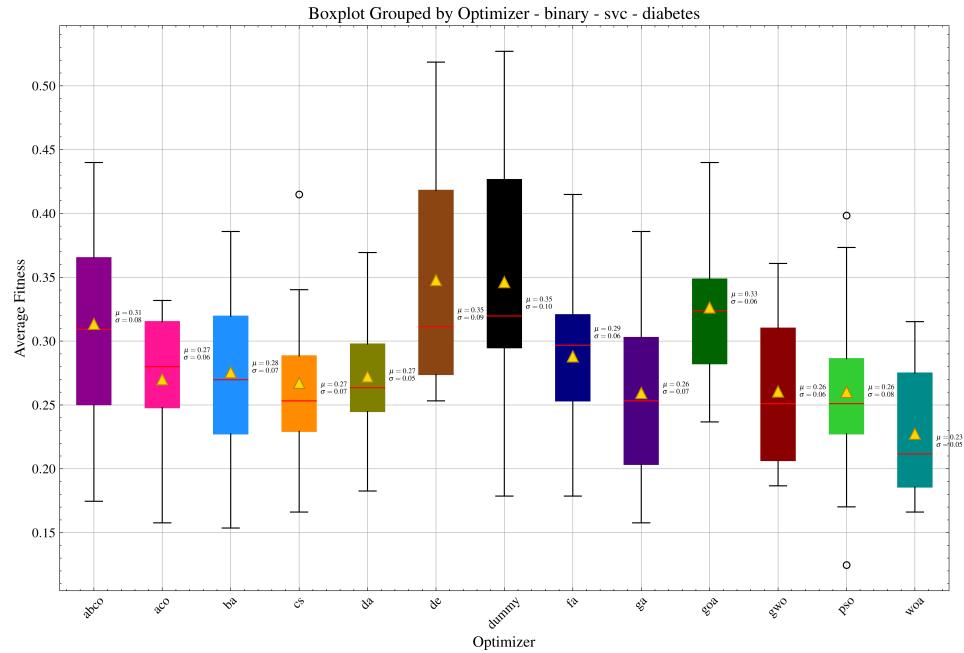


Figura 7.7: Boxplot diabetes en SVC

terísticas, escenario para los que **PSO** no fue diseñado. Esto hace de esta metaheurística una opción que parece sugerir mucha robustez.

El **CS**, basado en el comportamiento de anidación de ciertos pájaros, también se muestra como una opción fuerte, compitiendo por el podio con **GA**.

Finalmente, el **GA**, uno de los algoritmos más estudiados y aplicados en el ámbito de la optimización, sigue siendo relevante. Su capacidad para encontrar soluciones óptimas a través de mecanismos inspirados en la evolución natural, como la selección, el cruce y la mutación, le permite abordar una amplia gama de problemas con éxito. Además, este algoritmo al ser inicialmente pensado para aplicarse en dominios de codificación discreta, hace de esta metaheurística una opción muy interesante y robusta en el problema de selección de características, además de llevar años siendo aplicado con éxito en otros dominios.

Es visible que algunas de estas nuevas propuestas son muy potentes, llegando a mejorar en la propuesta binaria en comparación a algoritmos clásicos que llevan tiempo en la escena como las opciones *de facto*. Pese a ello, el grueso de los algoritmos comparados son opciones modernas, y muchos de

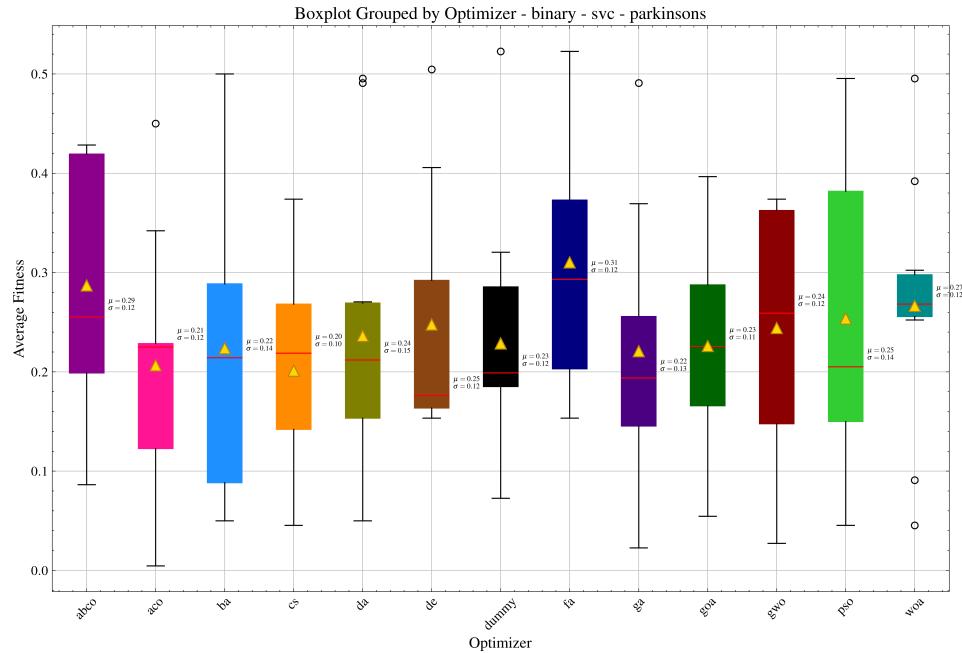


Figura 7.8: Boxplot parkinsons en SVC

estos algoritmos clásicos obtienen mejores resultados que la mayoría de este reciente grupo. De las cinco mejores, tres de ellos son clásicos, lo cual es más impresionante si se considera que hay un desbalance en número de *clásicos vs modernos*. *DE* es otro algoritmo asentado, que es capaz de clasificar en posiciones medias de *fitness*.

El peor de todos los algoritmos comparados es, sin embargo, **ABCO**, un algoritmo clásico. No parece poder alcanzar en calidad ni a modernos ni a algoritmos más antiguos como **PSO**.

Mejor vs Peor

Como ya se ha podido observar en, el algoritmo con mejor puntuación es **GWO** y el peor **ABCO** (7.2, 7.3, 7.4).

En las figuras de 7.11 y 7.10 se aprecia como **ABCO** muestra rendimientos muy buenos en según que conjunto de datos. Por ejemplo, en *yeast* o en *sonar*, el algoritmo obtiene resultados muy decentes y competitivos siendo estos dos conjuntos de datos de los más complicados de la selección, por convergencia y número de características. Pese a ello, en conjunto, no consigue superar al resto.

En cambio **GWO** parece un adaptarse bien a todos los problemas seleccionados sin excepción. Es capaz de obtener el mínimo, en el valor *fitness* de evaluación durante el entrenamiento, en casi todos los *datasets*.

	Original	Holm	Hommel	Hochberg
dummy_knn	6.104E-05	0.002	0.002	0.002
dummy_svc	3.052E-04	0.007	0.007	0.007
goa_knn	0.001	0.027	0.024	0.027
abco_knn	0.002	0.034	0.032	0.034
abco_svc	0.002	0.045	0.043	0.045
ba_svc	0.004	0.085	0.078	0.085
da_svc	0.005	0.102	0.097	0.102
goa_svc	0.008	0.149	0.149	0.149
de_knn	0.026	0.435	0.307	0.409
aco_knn	0.026	0.435	0.307	0.409
da_knn	0.064	0.956	0.573	0.887
woa_svc	0.073	1.000	0.588	0.887
fa_knn	0.083	1.000	0.653	0.887
ga_knn	0.107	1.000	0.686	0.887
cs_knn	0.107	1.000	0.686	0.887
de_svc	0.109	1.000	0.686	0.887
fa_svc	0.121	1.000	0.719	0.887
ba_knn	0.135	1.000	0.719	0.887
ga_svc	0.252	1.000	0.799	0.887
woa_knn	0.256	1.000	0.799	0.887
cs_svc	0.364	1.000	0.887	0.887
pso_knn	0.489	1.000	0.887	0.887
aco_svc	0.490	1.000	0.887	0.887
pso_svc	0.599	1.000	0.887	0.887
gwo_knn	0.887	1.000	0.887	0.887

Tabla 7.1: P-valores del mejor vs el resto

Se procede a hacer una comparativa estadística mediante la prueba de *Wilcoxon* y el test post hoc de *Holm* para determinar las diferencias significativas entre el mejor algoritmo (**GWO**) y el resto. Los resultados, presentados en la tabla 7.1, muestran los p-valores originales y ajustados, lo que nos permite identificar cuáles de las diferencias observadas son estadísticamente significativas con un error corregido. Este análisis es crucial para validar los resultados y asegurar que las conclusiones obtenidas no sean producto del azar, proporcionando una base sólida para afirmar la superioridad de **GWO** frente a los demás algoritmos en los conjuntos de datos evaluados, en frente al peor de ellos.

Para evaluar el rendimiento de los algoritmos, se plantean las siguientes hipótesis:

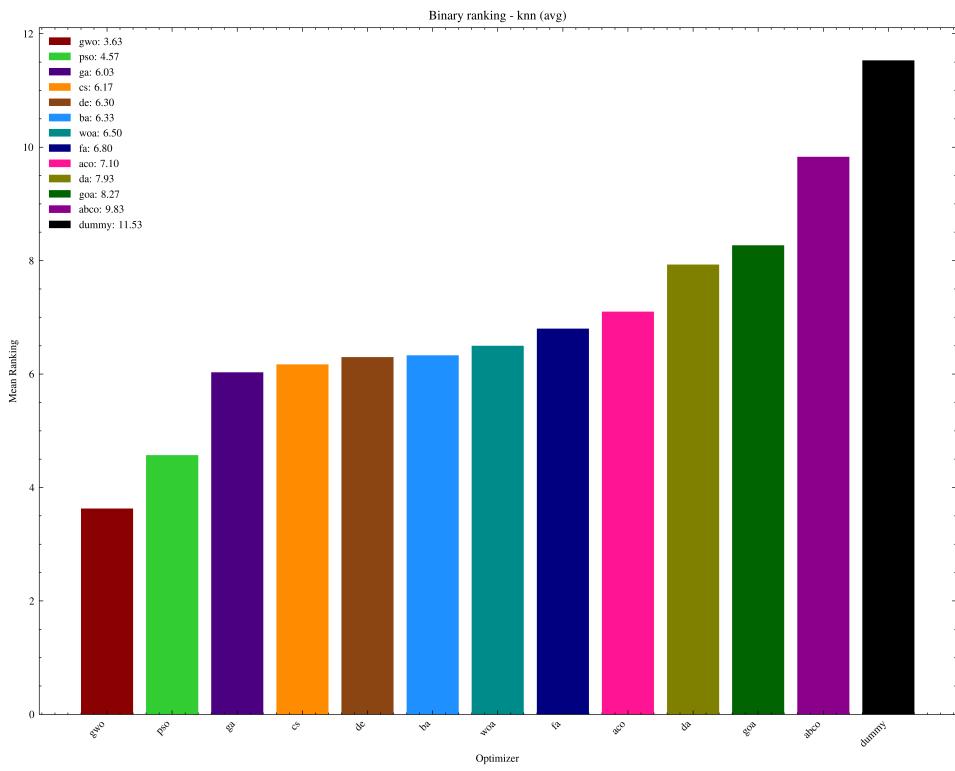
Hipótesis nula (H_0):

$$H_0 : \text{Rendimiento}_{GWO} \leq \text{Rendimiento}_{ABCO} \quad (7.1)$$

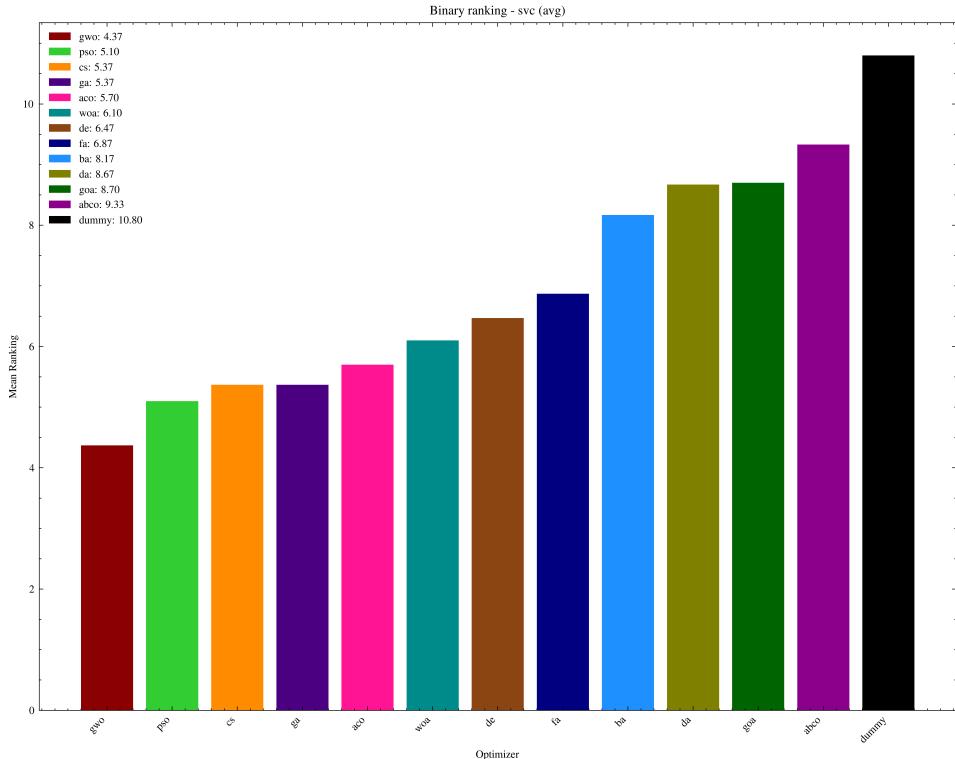
Hipótesis alternativa (H_1):

$$H_1 : \text{Rendimiento}_{GWO} > \text{Rendimiento}_{ABCO} \quad (7.2)$$

Si bien no se observan diferencias significativas sobre la mayoría de algoritmos, sí sobre el peor (**ABCO**) y sobre la versión de **GOA** usando **kNN**. Tanto en la versión que utiliza el clasificador **SVC** como en **kNN** se obtienen resultados suficientemente significativos como para poder considerar que los resultados de **GWO** frente a los de **ABCO** no son producto del azar. Por tanto, utilizando un valor crítico de $\alpha = 0.05$ se puede rechazar la hipótesis nula.



(a) Ranking por fitness para knn



(b) Ranking por fitness para svc

Figura 7.9: Rankings para fitness

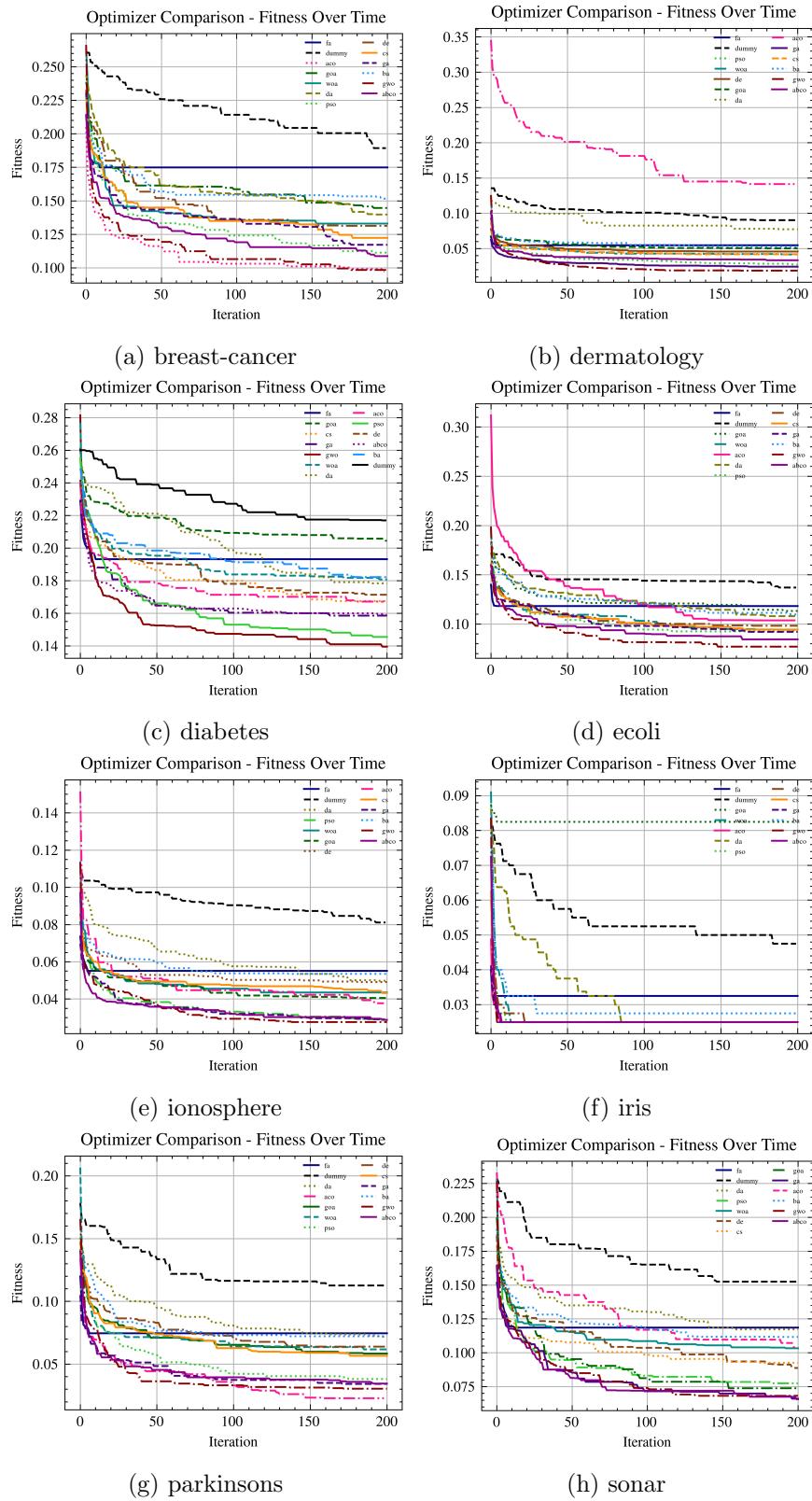


Figura 7.10: Convergencia con SVC parte 1

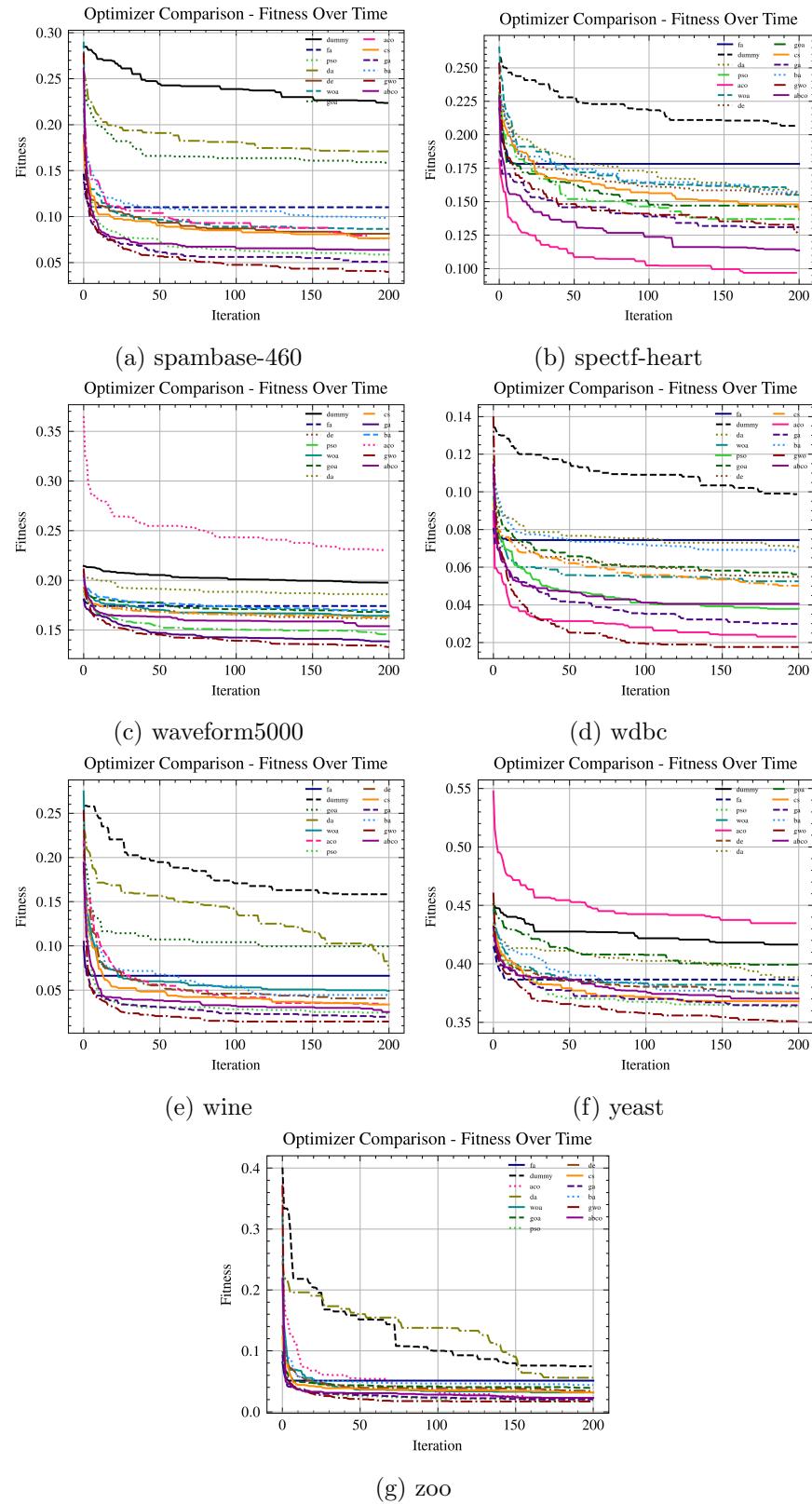


Figura 7.11: Convergencia con SVC parte 2

optimizer	breast-cancer	dermatology	diabetes	ecoli	ionosphere	iris	parkinsons	sonar	spambase-460	spectfheart	waveforms5000	wdbc	wine	yeast	zoo
abco	3.08E-01	1.76E-01	3.13E-01	2.99E-01	2.12E-01	1.00E-01	2.87E-01	4.53E-01	3.19E-01	2.80E-01	2.37E-01	1.60E-01	3.23E-01	5.26E-01	3.21E-01
aco	1.94E-01	2.04E-01	2.70E-01	3.53E-01	1.71E-01	1.15E-01	2.07E-01	3.33E-01	1.74E-01	2.27E-01	2.48E-01	8.78E-02	2.20E-01	4.91E-01	5.69E-01
ba	3.46E-01	1.62E-01	2.76E-01	2.68E-01	1.64E-01	1.45E-01	4.47E-01	2.24E-01	2.58E-01	3.13E-01	2.04E-01	1.35E-01	2.70E-01	5.09E-01	3.84E-01
cs	3.63E-01	9.01E-02	2.67E-01	2.31E-01	1.34E-01	1.15E-01	2.01E-01	4.26E-01	2.08E-01	2.91E-01	1.92E-01	1.44E-01	2.96E-01	5.00E-01	3.01E-01
da	2.17E-01	1.84E-01	2.72E-01	3.01E-01	1.87E-01	1.45E-01	2.36E-01	4.44E-01	2.74E-01	2.90E-01	2.02E-01	1.32E-01	3.77E-01	5.12E-01	3.36E-01
de	3.47E-01	1.91E-01	3.48E-01	2.82E-01	1.40E-01	7.00E-02	2.48E-01	3.63E-01	2.67E-01	2.74E-01	1.92E-01	1.31E-01	2.05E-01	5.22E-01	3.24E-01
dummy	2.92E-01	2.70E-01	3.46E-01	3.49E-01	1.74E-01	2.40E-01	2.29E-01	3.88E-01	3.82E-01	3.93E-01	2.35E-01	1.52E-01	5.75E-01	4.86E-01	
fa	2.58E-01	1.21E-01	2.88E-01	3.23E-01	1.57E-01	1.20E-01	3.10E-01	3.86E-01	2.14E-01	2.84E-01	2.04E-01	1.31E-01	4.10E-01	4.98E-01	2.46E-01
ga	3.02E-01	1.45E-01	2.50E-01	3.45E-01	1.36E-01	1.30E-01	2.21E-01	4.26E-01	1.81E-01	2.85E-01	1.80E-01	1.11E-01	3.04E-01	4.99E-01	3.07E-01
goa	2.98E-01	2.19E-01	3.26E-01	3.01E-01	1.61E-01	3.03E-01	2.26E-01	3.99E-01	3.69E-01	2.41E-01	2.00E-01	8.80E-02	4.06E-01	5.41E-01	4.34E-01
gwo	3.11E-01	1.73E-01	2.61E-01	2.79E-01	1.40E-01	1.15E-01	2.44E-01	3.34E-01	1.72E-01	2.77E-01	1.92E-01	5.51E-02	2.64E-01	4.81E-01	3.04E-01
pso	2.82E-01	1.30E-01	2.60E-01	2.92E-01	1.38E-01	1.30E-01	2.54E-01	3.89E-01	2.40E-01	2.78E-01	1.87E-01	1.15E-01	2.14E-01	4.59E-01	3.24E-01
woa	2.64E-01	1.61E-01	2.27E-01	2.61E-01	1.68E-01	1.45E-01	2.66E-01	3.68E-01	1.91E-01	3.54E-01	1.84E-01	1.11E-01	3.33E-01	5.22E-01	3.55E-01

Tabla 7.2: Fitness medio con clasificador SVC

optimizer	breast-cancer	dermatology	diabetes	ecoli	ionosphere	iris	parkinsons	sonar	spambase-460	spectfheart	waveform500	wdbc	wine	yeast	zoo
abco	3.64E-01	1.76E-01	3.39E-01	3.15E-01	3.11E-01	1.00E-01	2.10E-01	4.22E-01	2.88E-01	3.44E-01	2.72E-01	1.77E-01	2.66E-01	5.07E-01	3.94E-01
aco	2.35E-01	2.88E-01	3.15E-01	2.80E-01	2.34E-01	8.50E-02	2.30E-01	3.75E-01	2.60E-01	3.55E-01	2.85E-01	1.22E-01	2.07E-01	5.12E-01	3.90E-01
ba	2.96E-01	1.59E-01	3.20E-01	3.31E-01	2.49E-01	2.50E-02	1.98E-01	4.05E-01	2.38E-01	3.20E-01	2.44E-01	1.15E-01	1.85E-01	5.05E-01	4.21E-01
cs	3.33E-01	1.90E-01	3.16E-01	1.95E-01	3.08E-01	1.00E-01	2.08E-01	4.28E-01	1.86E-01	2.82E-01	2.42E-01	1.15E-01	2.41E-01	5.25E-01	3.57E-01
da	3.56E-01	2.00E-01	2.83E-01	2.51E-01	2.71E-01	5.50E-02	1.76E-01	4.62E-01	2.97E-01	3.20E-01	2.64E-01	1.49E-01	1.88E-01	5.46E-01	4.20E-01
de	3.54E-01	1.46E-01	3.23E-01	2.81E-01	2.51E-01	1.00E-01	3.26E-01	3.70E-01	1.94E-01	2.79E-01	2.34E-01	1.63E-01	1.79E-01	5.04E-01	3.96E-01
dummy	4.33E-01	2.34E-01	3.56E-01	3.03E-01	3.10E-01	1.45E-01	2.45E-01	4.76E-01	3.25E-01	3.04E-01	2.81E-01	1.63E-01	3.33E-01	5.13E-01	4.83E-01
fa	4.21E-01	1.46E-01	2.89E-01	2.68E-01	2.02E-01	2.75E-02	2.36E-01	4.53E-01	2.42E-01	3.05E-01	2.40E-01	1.30E-01	1.61E-01	5.13E-01	4.81E-01
ga	2.64E-01	1.25E-01	3.31E-01	2.94E-01	2.24E-01	8.50E-02	2.49E-01	3.82E-01	2.03E-01	3.03E-01	2.09E-01	9.70E-02	2.43E-01	5.13E-01	4.52E-01
goa	3.57E-01	2.16E-01	3.19E-01	3.14E-01	2.39E-01	1.32E-01	2.03E-01	3.90E-01	2.96E-01	3.41E-01	2.29E-01	1.18E-01	2.40E-01	5.61E-01	3.79E-01
gwo	2.63E-01	1.01E-01	2.75E-01	2.29E-01	2.27E-01	8.50E-02	2.14E-01	4.41E-01	1.89E-01	2.89E-01	2.06E-01	1.05E-01	1.05E-01	4.92E-01	4.31E-01
pso	3.41E-01	1.21E-01	2.95E-01	2.16E-01	2.60E-01	8.50E-02	2.33E-01	3.78E-01	2.21E-01	2.99E-01	2.11E-01	1.12E-01	2.07E-01	4.56E-01	3.60E-01
woa	3.23E-01	1.58E-01	3.34E-01	2.63E-01	2.17E-01	1.00E-01	2.50E-01	4.05E-01	2.02E-01	3.09E-01	2.48E-01	1.04E-01	5.26E-01	2.13E-01	

Tabla 7.3: Fitness medio con clasificador kNN

	abcs_knn	abcs_svc	acc_knn	acc_svc	ba_knn	ba_svc	es_knn	es_svc	du_knn	du_svc	de_knn	de_svc	dummy_knn	dummy_svc	fa_knn	fa_svc	ga_knn	ga_svc	geno_knn	geno_svc	pso_knn	pso_svc	wao_knn	wao_svc		
breast-cancer	24	13	3	1	10	18	16	23	21	2	20	19	26	9	25	4	6.5	12	22	11	5	14	17	8	15	6.5
dermatology	15.5	15.5	26	21	11	13	18	1	20	17	8.5	19	24	8.5	3.5	5	7	22	23	2	14	3.5	6	10	12	
diabetes	23	14	15	6	20	8	16	5	10	7	18	25	26	24	12	11	21	19	9	4	13	3	22	1	1	
ecoli	21	16	11	26	23	8.5	1	4	5	17.5	12	13	19	25	8.5	22	15	24	20	17.5	3	10	2	14	7	6
ionosphere	26	14	18	10	20	8	24	1	23	12	21	4	25	11	11	6	16	13	19	7	17	5	22	3	15	9
iris	11	11	6.5	15.5	1	23	11	15.5	3	23	11	4	23	25	2	18	6.5	19.5	21	26	6.5	15.5	6.5	19.5	11	18.5
parkinsons	7	24	13	5	2	10	6	3	1	15.5	26	19	18	12	16.5	25	20	9	4	11	8	17	14	21	22	23
sonar	16	23.5	6	1	14.5	22	19	17.5	25	21	5	3	26	10	23.5	9	8	17.5	12	13	20	2	7	11	14.5	4
spambase-460	20	23	17	3	13	16	1	10	22	19	7	18	24	26	15	11	9	4	21	25	5	2	12	14	8	6
spectfheart	23	7	25	1	20	19	8	13	21	12	6	3	16	26	17	9	15	10	22	2	11	4	14	5	18	24
waveform5000	24	17	26	21.5	20	9.5	19	5	23	8	15	5	25	16	18	9.5	12	1	14	7	11	5	13	3	21.5	2
wdbc	26	23	14	2	10	19	10	20	21	18	24.5	16.5	22	16	16.5	4	6.5	13	3	5	1	8	10	12	6.5	
wine	16	20	8.5	11	5	17	13	18	6	24	4	7	21.5	23	3	26	14	19	12	25	2	15	8.5	10	1	21.5
yeast	11	16	21	13.5	19	14	14	11	3	18	9	17	7.5	24	25	16	6	16	7	23	5	3	1	2	22	19
zoo	16	6	15	26	19	14	12	20	8	24	13.5	9.5	18	26	16	6	16	7	23	2	12	4	12	7.5	1	10
Mean	18.633	16.533	14.500	10.267	13.200	14.467	12.867	9.800	16.200	14.507	13.633	12.067	22.533	20.333	14.333	11.900	12.667	9.700	17.133	15.567	8.633	7.700	10.233	9.133	13.333	11.067

Tabla 7.4: Ranking de los algoritmos en *fitness*

7.1.3. Clasificación y reducción

Como se ha mencionado al principio de la anterior sección y como se explica en la fórmula 6.1, el *fitness* está compuesto por las métricas de *accuracy* en la clasificación y por el ratio de reducción de las características. Con ello se intenta reducir tantas características como sea posible, manteniendo un valor lo más alto posible de precisión.

Con estos aspectos aclarados, en esta sección se analizarán los algoritmos desde la perspectiva individual de cada una de las métricas que componen la función *fitness*. De esta manera, se proveerá información de cuál o cuáles algoritmos son los mejores para cada una de las métricas.

Esta visión es interesante ya que, los pesos fijados por el autor para cada métrica son los estándar en la mayoría de artículos, pero es posible que en otros problemas sea necesario ajustarlos a las necesidades del mismo. Por tanto si se conocen las características de los diferentes algoritmos para cada métrica por separado, se obtendrá un conocimiento que proporcionará un uso de los mismos más flexible.

Clásicos vs Modernos

Se procede a comparar los algoritmos clásicos frente a los algoritmos modernos en cuanto a su capacidad clasificatoria, es decir, el valor de cada uno de ellos obtenido en *accuracy*.

En los rankings definidos en la tabla 7.6, queda claro que **CS** usando el algoritmo **SVC** es capaz de obtener los mejores resultados en cuanto a la precisión en la clasificación. Le siguen algoritmos como **DE**, **GWO** o **FA**.

Si bien es cierto que la mayoría en el top de mejores algoritmos en *accuracy* son algoritmos modernos, los clásicos no se quedan atrás. En las comparaciones obtienen muy buenos resultados, no siendo los mejores y muy lejos de ser los peores. En los algoritmos modernos hay mayor diversidad en este punto. Hay unos pocos que destacan, pero el resto parece verse superado en muchas ocasiones por los algoritmos clásicos.

Pese a ello, se puede apreciar en la tabla 7.7 como algoritmos como **PSO** no obtienen resultados significativos con respecto al resto de sus competidores. Si se comparan los p-valores de los algoritmos modernos con los de los clásicos, se aprecia que, a excepción de algunos algoritmos como **Dummy** (obviamente) y **ACO**, no hay diferencia estadísticamente suficientemente

significativa como para determinar que un grupo de algoritmos es mejor que otro. De hecho, al corregir estos valores usando un post hoc como **Holms**, las diferencias son incluso menores.

Por tanto tanto un grupo como otro parece alcanzar resultados bastante satisfactorios en este sentido.

En cuanto a la reducción de características, según se observa en los rankings 7.12 como **ACO** es sin duda el mejor, por mucho además. Es capaz de reducir en algunos casos hasta el 90 % de las características.

Si se observan los p-valores obtenidos en la tabla 7.8 y se agrupan los modernos y los clásicos, se obtienen los siguientes resultados:

- **PSO:** A excepción del **ACO** (el mejor), **GWO** (segundo mejor) y **GA**, es superior a todo el resto de algoritmos con una diferencia significativa más que suficiente, incluso tras corrección post hoc con el método **Holms**.
- **ACO:** Es el mejor con diferencia significativa más que suficiente como para poder afirmar que este algoritmo es muy destacable en la reducción de características.
- **GA:** Ocurre lo mismo que con **PSO**. Pese a excepciones de algoritmos por encima suyo en el ranking, **GA** es muy superior al resto de algoritmos, con diferencias muy notables y de nuevo significativas.
- **DE:** Obtiene resultados muy buenos y estadísticamente significativos con respecto a **BA**, **WOA** y **FA**. Pese a ser mejor que otros algoritmos modernos, la diferencia no es suficientemente como para rechazar la hipótesis nula.
- **ABCO:** El peor parado del grupo de los algoritmos clásicos. No parece ser una buena propuesta a la hora de reducir.

Dada estos resultados, los algoritmos clásicos salen muy bien parados en la reducción de características frente a los modernos. Si bien es cierto que hay dos algoritmos que destacan sobre el resto de los modernos (**CS** y **GWO**) en la reducción de características, los clásicos en conjunto, con excepción de **ABCO**, parecen una opción robusta y que, en combinación a sus resultados de precisión (*accuracy*), son muy precisos a la hora de seleccionar características importantes.

Mejor vs resto

Como se ha mencionado anteriormente, la metaheurística **ACO** es sin lugar a dudas la mejor. Un ejemplo de esto es en el conjunto de datos de *ionosphere*. Se puede observar como el segundo algoritmo que más reduce, que es **GWO**, obtiene una reducción media del 74 %, mientras que el **ACO** obtiene una reducción media del 91.5 %, con tan solo una diferencia de 4.6 % en precisión a favor de **GWO** 7.9.

Si bien es cierto que su precisión no es de las mejores, se mantiene en mitad de los rankings con el resto de algoritmos. Teniendo ese aspecto en cuenta, añadido a que es capaz de reducir características a un nivel muy significativo con respecto al resto de algoritmos, **ACO** se presenta como un algoritmo muy interesante para aquellos problemas en los que se pueda permitir cierto nivel de error en la precisión, pero en cambio, sea crucial una compresión del tamaño del modelo o un tiempo den entrenamiento corto.

Si comparamos los p-valores 7.5, tras aplicar post hoc para mayor seguridad, de **ACO** con respecto al resto de algoritmos, vemos que claramente la mejora en reducción no se debe al azar de manera muy probable.

	Original	Holm	Hommel	Hochberg
dummy_svc	6.104E-05	0.002	0.001	0.001
goa_svc	6.104E-05	0.002	0.001	0.001
goa_knn	6.104E-05	0.002	0.001	0.001
fa_knn	6.104E-05	0.002	0.001	0.001
dummy_knn	6.104E-05	0.002	0.001	0.001
fa_svc	0.001	0.014	0.002	0.003
cs_svc	0.001	0.021	0.002	0.003
pso_svc	0.001	0.021	0.002	0.003
pso_knn	0.001	0.021	0.002	0.003
gwo_svc	0.001	0.021	0.002	0.003
ga_svc	0.001	0.021	0.002	0.003
ga_knn	0.001	0.021	0.002	0.003
abco_knn	0.001	0.021	0.002	0.003
de_svc	0.001	0.021	0.002	0.003
de_knn	0.001	0.021	0.002	0.003
da_svc	0.001	0.021	0.002	0.003
da_knn	0.001	0.021	0.002	0.003
cs_knn	0.001	0.021	0.002	0.003
ba_svc	0.001	0.021	0.002	0.003
ba_knn	0.001	0.021	0.002	0.003
abco_svc	0.001	0.021	0.002	0.003
woa_knn	0.001	0.021	0.002	0.003
woa_svc	0.001	0.021	0.002	0.003
gwo_knn	0.002	0.021	0.003	0.003
aco_svc	0.530	0.530	0.530	0.530

Tabla 7.5: P-valores de ACO vs el resto de algoritmos en reducción de características

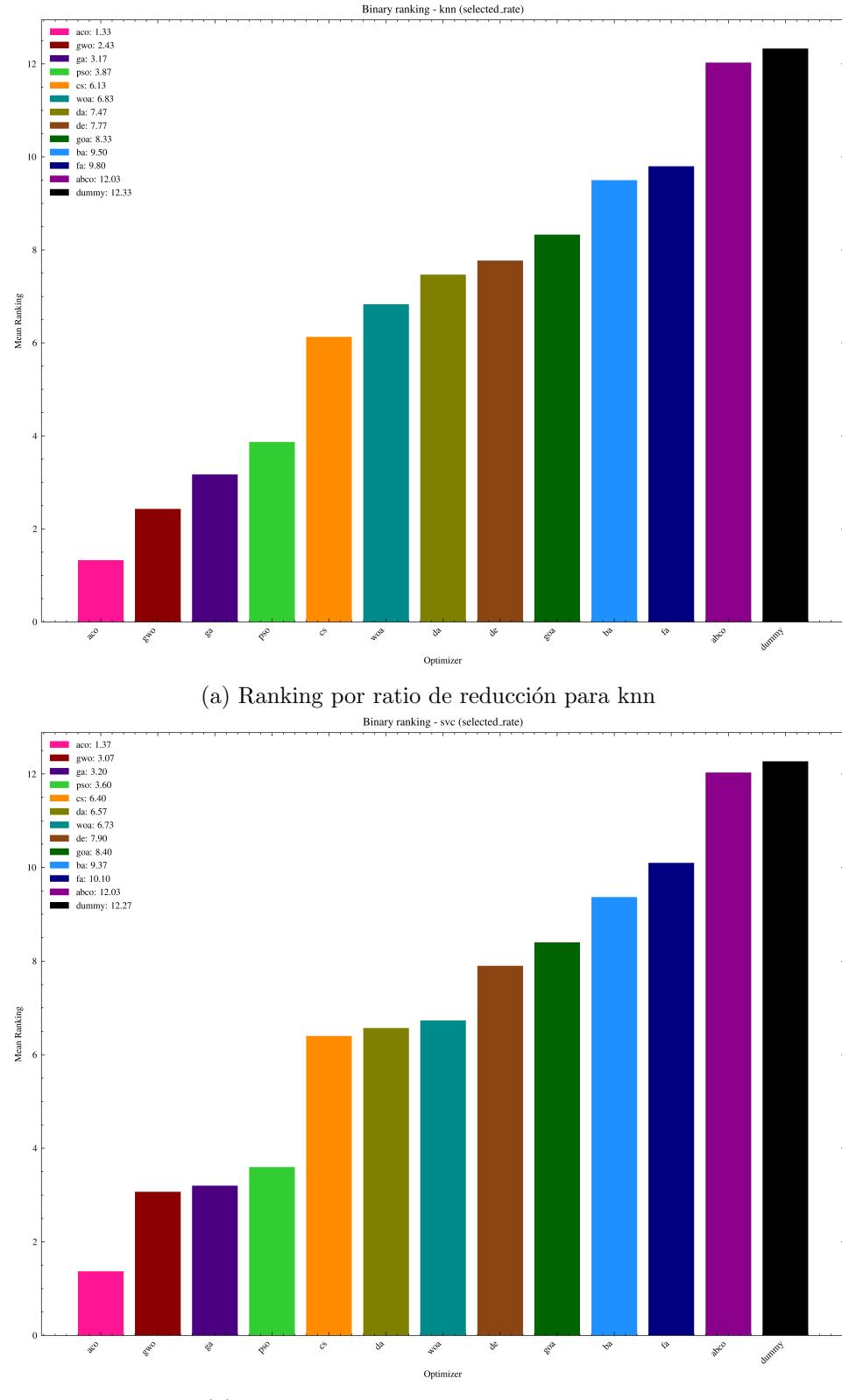


Figura 7.12: Rankings para reducción de características

	abco- skun	abco- swc	aco- skun	aco- swc	ba- skun	ba- swc	cs- skun	cs- swc	da- skun	da- swc	de- skun	de- swc	dum- my-skun	dum- my-swc	fa- skun	fa- swc	ga- skun	ga- swc	go- skun	go- swc	gw- skun	gw- swc	wo- skun	wo- swc		
breast-cancer	15	7	6	26	1.5	11.5	19	17	23	24	1.5	21	19	25	3.5	26	3.5	9	13	19	9	15	22	11.5	5	
dermatology	9	11.5	10	24.5	23.5	21	4.5	11.5	18	1	18	6.5	24	18.5	2.5	4.5	8.5	2.5	6.5	22.5	2.5	21	4.5	9	13.5	
diabetes	18.5	10	20.5	26	22.5	8	1	3	5.5	20.5	13	11	13	24	8	22.5	18	25	18	13	4	15.5	2	18	5.5	8
ecoli	15.5	10	20.5	11	24.5	14	16	26	1.5	22	12	19	1.5	24.5	3	13	4	17.5	6	17.5	8.5	20	8.5	22	6	15
ionosphere	22	12	7	17.5	1.5	23.5	12	17.5	3	23.5	19	4	17.5	25	1.5	17.5	7	21.5	7	26	7	17.5	7	21.5	12	17.5
lymphoma	12	17	23.5	17	3.5	8.5	5.5	1	13.5	26	13.5	8.5	11	23.5	13.5	5.5	8.5	13.5	23.5	13.5	11	14	20.5	17	20.5	20.5
parkinsons	2	16	18	11	20	18	1.5	22.5	4.5	26	22.5	6	21.5	22.5	6	14	22.5	6	20.5	1.5	20.5	1	8	11	14	4.5
spambase-460	16.5	20	24	11	12.5	15	1	9	21.5	19	2.5	18	21.5	25	14	7.5	10	5	23	26	7.5	2.5	12.5	6	16.5	4
spambase-600	17.5	1	26	8	17.5	19	1.5	10	10	22	12	14	5	25	14	5	21	14	23	2	17.5	5	20	10	17.5	24
spec-fleat	17.5	20	8	26	25	19	4.5	22	2	24	11	15.5	6.5	13.5	14	1.5	18	10	15.5	12	13.5	6.5	21	1	1	
waveform5000	22	6.5	25.5	13.5	3	16	6.5	24	23	16	25.5	13.5	7	19	21	10.5	10.5	6.5	19.5	10.5	2	16	19.5	6.5	4	
wdbc	13.5	19	9.5	12	5	15.5	13.5	19	5	24	5	14.5	4.5	24.5	1.5	22	3	25.5	2.5	15.5	22	17	9.5	1	1	23
year	6	13	23	19.5	10	7	22	8.5	25	14.5	12	14.5	4.5	26	16	8.5	24	18	11	3	1	2	10.5	1	19.5	17
zoo	13.5	3.5	18	13.5	10.5	3.5	18	6.5	16	6.5	22	22	24.5	2	24.5	6.5	13.5	20	13.5	9	11	3	10.5	1	10.5	17
Mean	14.067	10.033	15.883	12.067	12.433	13.633	9.833	16.600	14.467	13.567	10.767	17.883	16.300	12.700	10.033	15.567	13.400	15.700	14.733	12.867	10.300	12.233	11.867	12.733	10.900	

Tabla 7.6: Ranking de los algoritmos en *accuracy*

Tabla 7.7: P-valores para todos los algoritmos en *accuracy*

Tabla 7.8: P-valores para todos los algoritmos en *reducción de características*

optimizer	Data	dataset	breast-cancer	dermatology	diabetes	ecoli	ionosphere	iris	parkinsons	sonar	spambase-460	spectf-heart	waveform5000	wdlc	wine	yeast	zoo
abco	Average - acc	74.17%	88.00%	73.23%	77.14%	85.33%	91.67%	77.50%	60.00%	74.74%	79.29%	83.75%	91.74%	71.25%	51.33%	70.00%	
	Average - selected.rate	75.56%	67.65%	72.50%	92.86%	80.00%	25.00%	84.55%	92.67%	91.58%	93.64%	91.00%	85.33%	63.85%	87.50%	51.18%	
aco	Average - acc	79.17%	78.67%	71.94%	65.71%	82.00%	90.00%	77.50%	63.33%	81.05%	75.00%	73.45%	90.87%	77.50%	50.00%	38.00%	
	Average - selected.rate	6.67%	12.35%	17.50%	44.20%	8.53%	25.00%	4.09%	2.50%	3.33%	1.82%	9.25%	5.67%	17.69%	41.25%	10.50%	
ba	Average - acc	66.67%	88.00%	74.52%	77.86%	88.00%	86.67%	81.25%	57.78%	78.42%	72.86%	84.70%	90.43%	75.00%	52.67%	62.00%	
	Average - selected.rate	45.56%	53.82%	46.25%	68.57%	55.59%	25.00%	55.00%	66.83%	63.86%	68.86%	66.50%	49.00%	44.62%	82.50%	41.76%	
cs	Average - acc	64.17%	94.07%	75.16%	81.43%	90.00%	82.50%	58.89%	82.63%	74.29%	85.35%	88.26%	71.25%	52.33%	70.00%		
	Average - selected.rate	40.00%	42.06%	43.75%	64.20%	44.12%	25.00%	43.64%	55.83%	52.11%	59.77%	38.00%	37.69%	71.25%	31.18%		
da	Average - acc	79.17%	85.33%	74.19%	72.58%	84.67%	86.67%	78.75%	56.67%	76.32%	73.57%	83.70%	90.43%	62.50%	50.67%	68.00%	
	Average - selected.rate	30.00%	52.06%	40.00%	57.14%	48.53%	25.00%	45.00%	54.17%	60.53%	51.82%	55.00%	46.33%	39.23%	67.50%	47.65%	
de	Average - acc	66.67%	84.00%	66.77%	76.43%	90.00%	95.00%	78.75%	66.67%	76.84%	75.71%	85.25%	90.87%	81.25%	50.67%	68.00%	
	Average - selected.rate	46.67%	47.06%	48.75%	70.00%	49.71%	25.00%	56.36%	62.50%	58.25%	55.68%	58.75%	48.67%	36.15%	77.50%	35.88%	
dummy	Average - acc	76.67%	79.33%	70.00%	69.29%	89.33%	80.00%	83.75%	66.67%	67.37%	65.71%	83.15%	91.30%	70.00%	46.17%	54.00%	
	Average - selected.rate	82.22%	83.82%	76.25%	72.86%	77.94%	60.00%	82.27%	88.33%	88.77%	84.55%	83.50%	73.67%	68.46%	90.00%	71.76%	
fa	Average - acc	76.67%	92.67%	73.55%	71.43%	88.67%	90.00%	72.50%	64.44%	83.16%	75.71%	91.30%	60.00%	53.00%	78.00%	48.24%	
	Average - selected.rate	47.78%	55.29%	50.00%	65.71%	55.29%	30.00%	62.73%	66.00%	62.28%	65.45%	60.25%	52.33%	50.00%	75.00%	60.25%	
ga	Average - acc	69.17%	86.67%	74.52%	67.86%	88.00%	88.33%	78.75%	56.67%	84.21%	72.86%	84.75%	90.00%	68.75%	52.33%	68.00%	
	Average - selected.rate	24.44%	25.29%	30.00%	55.71%	27.65%	25.00%	29.55%	35.83%	38.77%	40.23%	42.75%	21.00%	22.31%	70.00%	19.41%	
goa	Average - acc	73.33%	81.33%	70.97%	75.71%	86.67%	75.00%	81.25%	61.11%	64.74%	78.57%	83.60%	94.78%	60.00%	50.17%	56.00%	
	Average - selected.rate	57.78%	50.88%	65.00%	82.58%	41.18%	77.50%	57.27%	49.33%	51.93%	48.64%	52.50%	41.00%	46.15%	92.50%	38.24%	
gwo	Average - acc	68.33%	82.67%	73.55%	75.00%	86.67%	90.00%	76.25%	68.89%	85.26%	75.71%	82.65%	95.65%	72.50%	53.83%	68.00%	
	Average - selected.rate	25.50%	22.50%	17.06%	54.29%	26.18%	25.00%	30.45%	54.50%	59.65%	58.18%	36.00%	16.00%	16.92%	65.00%	16.47%	
pso	Average - acc	71.67%	88.67%	74.84%	73.57%	88.33%	76.25%	62.22%	77.89%	74.29%	84.00%	90.00%	78.75%	55.50%	66.00%	18.24%	
	Average - selected.rate	26.67%	28.24%	33.75%	54.29%	29.71%	25.00%	40.00%	48.67%	41.40%	46.36%	42.75%	24.67%	22.31%	58.75%	18.24%	
woa	Average - acc	75.83%	87.35%	79.03%	77.86%	86.00%	90.00%	76.25%	65.36%	82.74%	67.14%	85.90%	92.17%	67.50%	50.33%	64.00%	
	Average - selected.rate	46.67%	47.06%	38.75%	61.43%	42.35%	25.00%	52.73%	58.68%	57.00%	40.67%	40.00%	31.18%	31.18%	50.00%	31.18%	

Tabla 7.9: Porcentajes de características seleccionadas y precisión en clasificación para cada algoritmo binario

Capítulo 8

Bibliografía

- [1] J. E. Hopcroft y J. D. Ullman, *Introduction to Automata Theory, Languages and Computation* (Addison-Wesley Series in Computer Science), 1st. Addison-Wesley Publishing Company, 1979.
- [2] J. v. Leeuwen, *Algorithms and complexity* (Handbook of theoretical computer science / ed. by Jan. van Leeuwen), eng, 1. MIT Press paperback ed., 2. printing. Amsterdam: Elsevier, 1998, OCLC: 247934368, ISBN: 978-0-262-22038-5. URL: <http://www.gbv.de/dms/bowker/toc/9780444880710.pdf> (visitado 20-04-2024).
- [3] Y. Abu-Mostafa, M. Magdon-Ismail y H. Lin, *Learning From Data*. AMLBook, 2012.
- [4] Udacity, *Curse of Dimensionality - Georgia Tech - Machine Learning*, [Online]. Available: Retrieved 2022-06-29, feb. de 2015.
- [5] D. Molina, J. Poyatos, J. Ser, S. García, A. Hussain y F. Herrera, «Comprehensive Taxonomies of Nature- and Bio-inspired Optimization: Inspiration Versus Algorithmic Behavior, Critical Analysis Recommendations,» English, *Cognitive Computation*, vol. 12, n.º 5, págs. 897-939, 2020, ISSN: 1866-9956. DOI: 10.1007/s12559-020-09730-8.
- [6] W. Clark, W. Polakov y F. Trabold, *The Gantt Chart, a Working Tool of Management*, English. New York: The Ronald Press Company, 1922.
- [7] J. Woidasky y E. Cetinkaya, «Use Pattern Relevance for Laptop Repair and Product Lifetime,» *Journal of Cleaner Production*, vol. 288, pág. 125 425, mar. de 2021, ISSN: 0959-6526. DOI: 10.1016/j.jclepro.2020.125425. URL: <https://www.sciencedirect.com/science/article/pii/S0959652620354718> (visitado 16-04-2024).
- [8] CITIC-UGR Sala de Servidores y Clústeres de Computadores, http://citic.ugr.es/pages/informacion_general/equipamiento/cluster, [Online; accessed 12-May-2024].

- [9] A. A. Kulkarni, G. G. Krishnasamy y A. A. Abraham, «Introduction to Optimization,» en *Optimization*. sep. de 2017, vol. 114, págs. 1-7, ISBN: 978-3-319-44253-2. DOI: 10.1007/978-3-319-44254-9_1.
- [10] A. Eiben y J. Smith, *Introduction to Evolutionary Computing* (Natural Computing Series), 2nd. Berlin, Heidelberg: Springer, 2015, pág. 30, S2CID 20912932, ISBN: 978-3-662-44873-1. DOI: 10.1007/978-3-662-44874-8.
- [11] S. Shalev-Shwartz y S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. CUP, 2014, ISBN: 978-3-319-44253-2.
- [12] J. Miao y L. Niu, «A Survey on Feature Selection,» *Procedia Computer Science*, Promoting Business Analytics and Quantitative Management of Technology: 4th International Conference on Information Technology and Quantitative Management (ITQM 2016), vol. 91, págs. 919-926, ene. de 2016, ISSN: 1877-0509. DOI: 10.1016/j.procs.2016.07.111. URL: <https://www.sciencedirect.com/science/article/pii/S1877050916313047> (visitado 02-04-2024).
- [13] N. Venkat, *The Curse of Dimensionality: Inside Out*, sep. de 2018. DOI: 10.13140/RG.2.2.29631.36006.
- [14] R. Bellman, *Dynamic Programming*. Princeton Univ Pr, 1957, ISBN: 978-0691079516.
- [15] R. E. Bellman, *Adaptive Control Processes*. Princeton University Press, 1961.
- [16] D. Peng, Z. Gui y H. Wu, *Interpreting the Curse of Dimensionality from Distance Concentration and Manifold Effect*, en-US, arXiv:2401.00422 [cs], ene. de 2024. DOI: 10.48550/arXiv.2401.00422. URL: <http://arxiv.org/abs/2401.00422> (visitado 20-04-2024).
- [17] K. Beyer, J. Goldstein, R. Ramakrishnan y U. Shaft, «When Is “Nearest Neighbor” Meaningful?» Inglés, en *Proceedings of the 7th International Conference on Database Theory*, ép. Lecture Notes in Computer Science, vol. 1540, Springer, 1999, págs. 217-235. URL: <http://www.springerlink.com/link.asp?id=04p94cqnbge862kh>.
- [18] L. Bianchi, M. Dorigo, L. M. Gambardella y W. J. Gutjahr, «A survey on metaheuristics for stochastic combinatorial optimization,» *Natural Computing*, vol. 8, n.º 2, págs. 239-287, 2009. DOI: 10.1007/s11047-008-9098-4.
- [19] J. Xu y J. Zhang, «Exploration-exploitation tradeoffs in metaheuristics: Survey and analysis,» en *Proceedings of the 33rd Chinese control conference*, IEEE, 2014, págs. 8633-8638.
- [20] J. Xu y J. Zhang, «Exploration-exploitation tradeoffs in metaheuristics: Survey and analysis,» en *Proceedings of the 33rd Chinese Control Conference*, 2014, págs. 8633-8638. DOI: 10.1109/ChiCC.2014.6896450.

- [21] D. H. Wolpert y W. G. Macready, «No free lunch theorems for optimization,» *IEEE Transactions on Evolutionary Computation*, vol. 1, n.º 1, págs. 67-82, 1997. DOI: 10.1109/4235.585893.
- [22] M. Goldblum, M. Finzi, K. Rowan y A. G. Wilson, *The No Free Lunch Theorem, Kolmogorov Complexity, and the Role of Inductive Biases in Machine Learning*, 2023. arXiv: 2304.05366 [cs.LG].
- [23] K. P. Murphy, *Machine Learning: A Probabilistic Perspective* (Adaptive Computation and Machine Learning), 1.^a ed. The MIT Press, 2012.
- [24] S. Sah, *Machine Learning: A Review of Learning Types*, jul. de 2020. DOI: 10.20944/preprints202007.0230.v1.
- [25] C. Cortes y V. Vapnik, «Support-Vector Networks,» en, *Machine Learning*, vol. 20, n.º 3, págs. 273-297, sep. de 1995. DOI: 10.1007/BF00994018. URL: <http://link.springer.com/10.1007/BF00994018> (visitado 02-04-2024).
- [26] scikit-learn Developers, *Support Vector Machines — scikit-learn documentation*, <https://scikit-learn.org/stable/modules/svm.html>, Archived from the original on 2017-11-08, Retrieved 2024-04-27, 2024.
- [27] T. Hastie, R. Tibshirani y J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd. New York: Springer, 2009.
- [28] E. Fix y J. L. Hodges, «Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties,» *International Statistical Review / Revue Internationale de Statistique*, vol. 57, n.º 3, págs. 238-247, 1989, Publisher: [Wiley, International Statistical Institute (ISI)], ISSN: 0306-7734. DOI: 10.2307/1403797. URL: <https://www.jstor.org/stable/1403797> (visitado 02-04-2024).
- [29] T. Cover y P. Hart, «Nearest Neighbor Pattern Classification,» en, *IEEE Transactions on Information Theory*, vol. 13, n.º 1, págs. 21-27, ene. de 1967. DOI: 10.1109/TIT.1967.1053964. URL: <http://ieeexplore.ieee.org/document/1053964/> (visitado 02-04-2024).
- [30] G. Guo, H. Wang, D. Bell, Y. Bi y K. Greer, «KNN Model-Based Approach in Classification,» en *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, R. Meersman, Z. Tari y D. C. Schmidt, eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, págs. 986-996, ISBN: 978-3-540-39964-3.
- [31] K. Kira y L. A. Rendell, «A Practical Approach to Feature Selection,» English, en *Proc. 9th International Workshop on Machine Learning (ICML 1992)*, 1992, págs. 249-256, ISBN: 978-1-55860-247-2. DOI: 10.1016/B978-1-55860-247-2.50037-1.

- [32] C. Ding y H. Peng, «Minimum Redundancy Feature Selection from Microarray Gene Expression Data,» English, *Journal of Bioinformatics and Computational Biology*, vol. 3, n.º 2, págs. 185-205, 2005, ISSN: 0219-7200. DOI: 10.1142/S0219720005001004.
- [33] S. Mirjalili, S. M. Mirjalili y A. Lewis, «Grey Wolf Optimizer,» en-US, *Advances in Engineering Software*, vol. 69, págs. 46-61, mar. de 2014, ISSN: 0965-9978. DOI: 10.1016/j.advengsoft.2013.12.007. URL: <https://www.sciencedirect.com/science/article/pii/S0965997813001853> (visitado 18-11-2023).
- [34] X.-S. Yang, «Chapter 8 - Firefly Algorithms,» en-US, en *Nature-Inspired Optimization Algorithms*, X.-S. Yang, ed., Oxford: Elsevier, ene. de 2014, págs. 111-127, ISBN: 978-0-12-416743-8. DOI: 10.1016/B978-0-12-416743-8.00008-7. URL: <https://www.sciencedirect.com/science/article/pii/B9780124167438000087> (visitado 23-01-2024).
- [35] S. Saremi, S. Mirjalili y A. Lewis, «Grasshopper Optimisation Algorithm: Theory and application,» en-US, *Advances in Engineering Software*, vol. 105, págs. 30-47, mar. de 2017, ISSN: 0965-9978. DOI: 10.1016/j.advengsoft.2017.01.004. URL: <https://www.sciencedirect.com/science/article/pii/S0965997816305646> (visitado 04-11-2023).
- [36] S. Mirjalili y A. Lewis, «The Whale Optimization Algorithm,» en-US, *Advances in Engineering Software*, vol. 95, págs. 51-67, mayo de 2016, ISSN: 0965-9978. DOI: 10.1016/j.advengsoft.2016.01.008. URL: <https://www.sciencedirect.com/science/article/pii/S0965997816300163> (visitado 14-10-2023).
- [37] S. Mirjalili, «Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems,» en, *Neural Computing and Applications*, vol. 27, n.º 4, págs. 1053-1073, mayo de 2016, ISSN: 1433-3058. DOI: 10.1007/s00521-015-1920-1. URL: <https://doi.org/10.1007/s00521-015-1920-1> (visitado 06-11-2023).
- [38] X.-S. Yang y S. Deb, *Cuckoo Search via Levy Flights*, en-US, arXiv:1003.1594 [math] version: 1, mar. de 2010. DOI: 10.48550/arXiv.1003.1594. URL: <http://arxiv.org/abs/1003.1594> (visitado 13-03-2024).
- [39] X.-S. Yang, *A New Metaheuristic Bat-Inspired Algorithm*, en-US, Is-sue: arXiv:1004.4170 arXiv:1004.4170 [physics], abr. de 2010. DOI: 10.48550/arXiv.1004.4170. URL: <http://arxiv.org/abs/1004.4170> (visitado 22-01-2024).

- [40] J. Kennedy y R. Eberhart, «Particle swarm optimization,» en-US, en *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, nov. de 1995, 1942-1948 vol.4. DOI: 10.1109/ICNN.1995.488968. URL: <https://ieeexplore.ieee.org/document/488968>.
- [41] J. H. Holland, *Adaptation in Natural and Artificial Systems*, 2nd. Ann Arbor, MI: University of Michigan Press, 1975, second edition, 1992.
- [42] M. Dorigo y G. D. Caro, «Ant colony optimization: a new metaheuristic,» en *Proceedings of the 1999 Congress on Evolutionary Computation- CEC99 (Cat. No. 99TH8406)*, vol. 2, jul. de 1999, 1470-1477 Vol. 2. DOI: 10.1109/CEC.1999.782657. URL: <https://ieeexplore.ieee.org/document/782657>.
- [43] R. Storn y K. Price, «Differential Evolution - A Simple and Efficient Heuristic for global Optimization over Continuous Spaces,» *Journal of Global Optimization*, vol. 11, n.^o 4, págs. 341-359, dic. de 1997, ISSN: 1573-2916. DOI: 10.1023/A:1008202821328. URL: <https://doi.org/10.1023/A:1008202821328>.
- [44] D. Karaboga, «AN IDEA BASED ON HONEY BEE SWARM FOR NUMERICAL OPTIMIZATION,» en,
- [45] D. Simon, *Evolutionary Optimization Algorithms*. Wiley, 2013, ISBN: 9781118659502. URL: <https://books.google.es/books?id=gwUwIEPqk30C>.
- [46] A. Colorni, M. Dorigo y V. Maniezzo, «Distributed Optimization by Ant Colonies,» en *Proceedings of the First European Conference on Artificial Life*, ene. de 1991.
- [47] M. Jaderberg et al., *Population Based Training of Neural Networks*, 2017. arXiv: 1711.09846 [cs.LG].
- [48] T. Dokeroglu, A. Deniz y H. E. Kiziloz, «A comprehensive survey on recent metaheuristics for feature selection,» en-US, *Neurocomputing*, vol. 494, págs. 269-296, jul. de 2022, ISSN: 0925-2312. DOI: 10.1016/j.neucom.2022.04.083. URL: <https://www.sciencedirect.com/science/article/pii/S092523122200474X> (visitado 22-09-2023).
- [49] I. Boussaïd, J. Lepagnot y P. Siarry, «A survey on optimization metaheuristics,» *Information Sciences, Prediction, Control and Diagnosis using Advanced Neural Computations*, vol. 237, págs. 82-117, jul. de 2013, ISSN: 0020-0255. DOI: 10.1016/j.ins.2013.02.041. URL: <https://www.sciencedirect.com/science/article/pii/S0020025513001588> (visitado 20-04-2024).
- [50] P. Agrawal, H. Abutarboush, T. Ganesh y A. Mohamed, «Metaheuristic algorithms on feature selection: A survey of one decade of research (2009-2019),» en-US, *IEEE Access*, vol. 9, págs. 26 766-26 791, 2021, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3056407.

- [51] A. G. Guy, D. A. Bohan, S. J. Powers y A. M. Reynolds, «Avoidance of conspecific odour by carabid beetles: a mechanism for the emergence of scale-free searching patterns,» *Animal Behaviour*, vol. 76, n.º 3, págs. 585-591, 2008, ISSN: 0003-3472. DOI: 10.1016/j.anbehav.2008.04.004. URL: <https://www.sciencedirect.com/science/article/pii/S0003347208001668>.
- [52] E. Emary, H. M. Zawbaa y A. E. Hassanien, «Binary grey wolf optimization approaches for feature selection,» en-US, *Neurocomputing*, vol. 172, págs. 371-381, ene. de 2016, ISSN: 0925-2312. DOI: 10.1016/j.neucom.2015.06.083. URL: <https://www.sciencedirect.com/science/article/pii/S0925231215010504> (visitado 18-11-2023).
- [53] A. G. Hussien, A. E. Hassanien, E. H. Houssein, S. Bhattacharyya y M. Amin, «S-shaped Binary Whale Optimization Algorithm for Feature Selection,» en, en *Recent Trends in Signal and Image Processing*, S. Bhattacharyya, A. Mukherjee, H. Bhaumik, S. Das y K. Yoshida, eds., ép. *Advances in Intelligent Systems and Computing*, Singapore: Springer, 2019, págs. 79-87, ISBN: 978-981-10-8863-6. DOI: 10.1007/978-981-10-8863-6_9.
- [54] M. Mafarja y S. Mirjalili, «Whale optimization approaches for wrapper feature selection,» *Applied Soft Computing*, vol. 62, págs. 441-453, ene. de 2018, ISSN: 1568-4946. DOI: 10.1016/j.asoc.2017.11.006. URL: <https://www.sciencedirect.com/science/article/pii/S1568494617306695> (visitado 20-11-2023).
- [55] M. Mafarja et al., «Binary dragonfly optimization for feature selection using time-varying transfer functions,» English, *Knowledge-Based Systems*, vol. 161, págs. 185-204, 2018, ISSN: 0950-7051. DOI: 10.1016/j.knosys.2018.08.003.
- [56] M. Mafarja, I. Aljarah, H. Faris, A. I. Hammouri, A. M. Al-Zoubi y S. Mirjalili, «Binary grasshopper optimisation algorithm approaches for feature selection problems,» *Expert Systems with Applications*, vol. 117, págs. 267-286, mar. de 2019, ISSN: 0957-4174. DOI: 10.1016/j.eswa.2018.09.015. URL: <https://www.sciencedirect.com/science/article/pii/S0957417418305864> (visitado 23-10-2023).
- [57] S. Mirjalili, S. Mirjalili y X.-S. Yang, «Binary bat algorithm,» en-US, *Neural Computing and Applications*, vol. 25, n.º 3-4, págs. 663-681, 2014, Number: 3-4, ISSN: 0941-0643. DOI: 10.1007/s00521-013-1525-5.
- [58] D. Rodrigues et al., «BCS: A Binary Cuckoo Search algorithm for feature selection,» en *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, ISSN: 2158-1525, mayo de 2013, págs. 465-468. DOI: 10.1109/ISCAS.2013.6571881. URL: <https://ieeexplore.ieee.org/document/6571881> (visitado 13-03-2024).

- [59] L. Zhang, L. Shan y J. Wang, «Optimal feature selection using distance-based discrete firefly algorithm with mutual information criterion,» *Neural Computing and Applications*, 2016. DOI: 10.1007/s00521-016-2204-0.
- [60] S. Mirjalili y A. Lewis, «S-shaped versus V-shaped transfer functions for binary Particle Swarm Optimization,» *Swarm and Evolutionary Computation*, vol. 9, págs. 1-14, abr. de 2013, ISSN: 2210-6502. DOI: 10.1016/j.swevo.2012.09.002. URL: <https://www.sciencedirect.com/science/article/pii/S2210650212000648> (visitado 29-10-2023).
- [61] M. S. Kiran, «A binary artificial bee colony algorithm and its performance assessment,» en, *Expert Systems with Applications*, vol. 175, pág. 114817, ago. de 2021, ISSN: 09574174. DOI: 10.1016/j.eswa.2021.114817. URL: <https://linkinghub.elsevier.com/retrieve/pii/S095741742100258X> (visitado 28-11-2023).
- [62] G. Pampara, A. Engelbrecht y N. Franken, «Binary Differential Evolution,» en *2006 IEEE International Conference on Evolutionary Computation*, ISSN: 1941-0026, jul. de 2006, págs. 1873-1879. DOI: 10.1109/CEC.2006.1688535. URL: <https://ieeexplore.ieee.org/document/1688535> (visitado 25-03-2024).
- [63] K. Socha, «ACO for Continuous and Mixed-Variable Optimization,» en, en *Ant Colony Optimization and Swarm Intelligence*, M. Dorigo, M. Birattari, C. Blum, L. M. Gambardella, F. Mondada y T. Stützle, eds., ép. Lecture Notes in Computer Science, Berlin, Heidelberg: Springer, 2004, págs. 25-36, ISBN: 978-3-540-28646-2. DOI: 10.1007/978-3-540-28646-2_3.
- [64] Y. He, F. Zhang, S. Mirjalili y T. Zhang, «Novel binary differential evolution algorithm based on Taper-shaped transfer functions for binary optimization problems,» *Swarm and Evolutionary Computation*, vol. 69, pág. 101022, mar. de 2022, ISSN: 2210-6502. DOI: 10.1016/j.swevo.2021.101022. URL: <https://www.sciencedirect.com/science/article/pii/S221065022100184X> (visitado 27-03-2024).
- [65] M. Mitchell, *An Introduction to Genetic Algorithms*. Cambridge, MA, USA: MIT Press, 1998, ISBN: 0262631857.
- [66] T. V. Mathew, «Genetic algorithm,» *Report submitted at IIT Bombay*, vol. 53, 2012.
- [67] S. Mirjalili y S. Mirjalili, «Genetic algorithm,» *Evolutionary algorithms and neural networks: Theory and applications*, págs. 43-55, 2019.
- [68] B. L. Miller, «Genetic Algorithms, Tournament Selection, and the Effects of Noise,» en,

- [69] Z. C. Dagdia y M. Mirchev, «Chapter 15 - When Evolutionary Computing Meets Astro- and Geoinformatics,» en *Knowledge Discovery in Big Data from Astronomy and Earth Observation*, P. Škoda y F. Adam, eds., Elsevier, 2020, págs. 283-306, ISBN: 978-0-12-819154-5. DOI: 10.1016/B978-0-12-819154-5.00026-6. URL: <https://www.sciencedirect.com/science/article/pii/B9780128191545000266>.
- [70] Purdue University College of Engineering, *Lecture 4: Real-Coded Genetic Algorithms*, Lecture notes, Accessed on April 27, 2024. URL: <https://engineering.purdue.edu/~sudhoff/ee630/Lecture04.pdf>.
- [71] S. Kashef y H. Nezamabadi-pour, «An advanced ACO algorithm for feature subset selection,» *Neurocomputing*, vol. 147, págs. 271-279, ene. de 2015, ISSN: 0925-2312. DOI: 10.1016/j.neucom.2014.06.067. URL: <https://www.sciencedirect.com/science/article/pii/S0925231214008601>.
- [72] A. P. Engelbrecht, *Computational Intelligence: An Introduction*, 2nd. Wiley Publishing, 2007, ISBN: 0470035617.
- [73] K. Price, «Eliminating Drift Bias from the Differential Evolution Algorithm,» vol. 143, ene. de 2008. DOI: 10.1007/978-3-540-68830-3_2.
- [74] *Lévy Flight - an overview — ScienceDirect Topics*. URL: <https://www.sciencedirect.com/topics/physics-and-astronomy/levy-flight> (visitado 06-05-2024).
- [75] Y. Zhang, X.-f. Song y D.-w. Gong, «A return-cost-based binary firefly algorithm for feature selection,» en, *Information Sciences*, vol. 418-419, págs. 561-574, dic. de 2017, ISSN: 00200255. DOI: 10.1016/j.ins.2017.08.047. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0020025516314098> (visitado 21-03-2024).
- [76] *MinMaxScaler*, Scikit-learn Documentation, URL: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>.
- [77] Y. Li, T. Han, H. Zhao y H. Gao, «An Adaptive Whale Optimization Algorithm Using Gaussian Distribution Strategies and Its Application in Heterogeneous UCAVs Task Allocation,» *IEEE Access*, vol. 7, págs. 110138-110158, 2019. URL: <https://api.semanticscholar.org/CorpusID:201621393>.
- [78] D. Molina, *TACO Website*, <https://tacolab.org/>, Accessed: 2024-06-01, 2024.