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 muchas veces estos requieren de un procesamiento previo para identificar las características más relevantes. Ya sea por escoger un problema amplio y complejo, por falta de conocimiento o por falta de contexto a la hora de recolectar los datos, es posible que se llegue a aglomerar información de más, dando lugar a conjuntos de datos inmensos. Incluso aquellos algoritmos que internamente realizan esta identificación de características pueden beneficiarse de un procesamiento adicional. Este preprocesamiento, conocido como selección de características (features selection), se considera un problema complejo de optimización combinatoria.

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 like KNN or SVM often yield excellent results, but they often require preprocessing to identify the most relevant features. Whether due to tackling a broad and complex problem, lack of domain knowledge, or collecting data without proper context, it's possible to gather excessive information, resulting in vast datasets. Even algorithms internally performing feature identification can benefit from additional processing. This preprocessing, known as feature selection, is considered a complex combinatorial optimization 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

Índice de figuras

Índice de tablas

Capítulo 1

Introducción

1.1. Definición del problema

El problema de la selección de características se define como el proceso de seleccionar un subconjunto de características relevantes [1]. Una característica es una propiedad individual medible de un fenómeno concreto. Este problema es considerado un problema **NP duro** [2, 3]. 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 es inevitable [4].

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*.

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 sobreajuste. 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 [4].

La reducción de características ayuda a manejar ambos tipos de ruido [1, 4] 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 sobreajuste.

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 [1]. 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 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.

3. Evita la maldición de la dimensionalidad [5, 6]: Cuando la dimensionalidad se incrementa en un problema, el volumen del espacio también lo hace, y esto ocurre tan rápido que hace que los datos disponibles se vuelvan dispersos. De forma que para obtener un resultado seguro/-fiable, la cantidad de datos necesitados debe verse incrementada de manera exponencial con la dimensionalidad [7]. A menor dimensionalidad (características en el conjunto de datos) menos datos harán falta para obtener un buen modelo.

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 [1] 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 [1]. Por lo general, estos métodos realizan la selección de características antes de las tareas de clasificación y agrupamiento. Ejemplos de algoritmos de filtrado son *relieF* [8] o F-statistic [9].

Los métodos **wrapper**, en cambio, utilizan el algoritmo de aprendizaje usado postprocesamiento para evaluar las características y seleccionar así las más útiles [1].

Los algoritmos clasificatorios de aprendizaje utilizados en este trabajo son

SVM [10] y *kNN* [11, 12], 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.

1.1.1. Motivación

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.

Esta crecimiento viene acompañado, sin embargo, de comparaciones que distan de ser objetivas por varios motivos. 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 superanmal resto en alguna métrica concreta sin llegar a la rigurosidad adecuada.

Por ello, la motivación principal de este trabajo es la de proveer información no sesgada y todo lo objetiva posible por medio de un análisis comparativo entre los algoritmos optimizatorios metaheurísticos más populares y más citados junto con los algoritmos más robustos y clásicos en el campo de la optimización pseudo estocástica.

1.1.2. 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.

1.1.3. Planificación

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. Ha de tenerse en cuenta también que el alumno trabaja 25 horas semanales y debe superar algunas asignaturas además de su projecto 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 intensiva.

Se planifica una duración de 5 meses aproximadamente. Se utilizará un diagrama de Gantt [13] para describir la planificación del proyecto, de manera que se realizarán tareas en un orden cronológico. Sin embargo, se reconoce que algunas tareas probablemente requerirán iteraciones posteriores, ya que es probable que se mejore y perfeccione el proyecto a lo largo de su ciclo de vida.

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, siendo estas:

  - Binary Firefly Algorithm
  - Binary Whale Optimization Algorithm
  - Binary Bat Swarm Optimizer
  - Binary Grey Wolf Optimizer
  - Binary Dragonfly Algorithm
  - Binary Grasshoper Algorithm
  - Binary Cuckoo Search
  - Binary Differential Algorithm
  - Ant Colony Optimization
  - Binary Artificial Bee Colony Optimization
  - Binary Particle Swarn Optimization
  - Genetic Algorithm (binary & real)

  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.

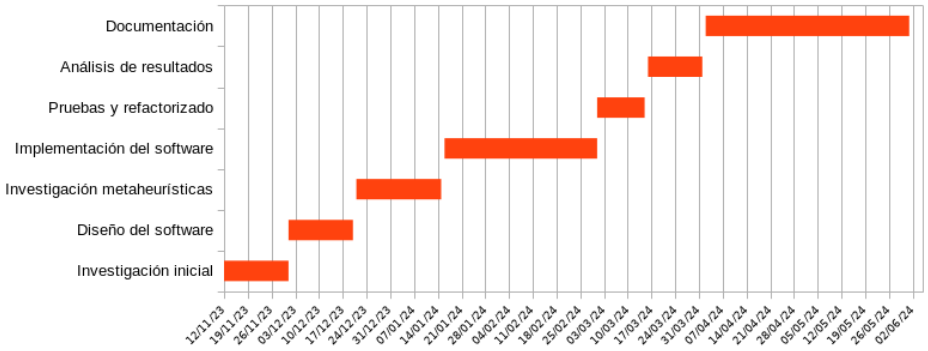

Figura 1.1: Diagrama de Gantt inicial

La planificación inicial tiene en cuenta un curso ideal del ciclo de vida del proyecto, siendo las estapas más extensas la de creación del software e implementación de la documentación. Son etapas excluyentes, no pueden ocurrir a la vez según este tipo de planificación. Al terminar una etapa se pasa inmediatamente a la siguiente.

La planificación final del proyecto se ha modificado significativamente debido a una serie de contratiempos y obstáculos surgidos durante su desarrollo, así como la influencia de numerosos eventos externos que han afectado a su cronograma. En particular, se han experimentado retrasos y bloqueos que han incidido en la duración prevista del proyecto. Por ejemplo, las etapas de implementación del software y la investigación de las metaheurísticas se han entrelazado debido a que la implementación efectiva del algoritmo se facilitaba una vez que se había estudiado a fondo la metaheurística correspondiente.

Esto ha llevado a una reevaluación de la estrategia de planificación original, reconociendo que no habría sido eficiente estudiar todas las metaheurísticas simultáneamente y luego proceder con su implementación.

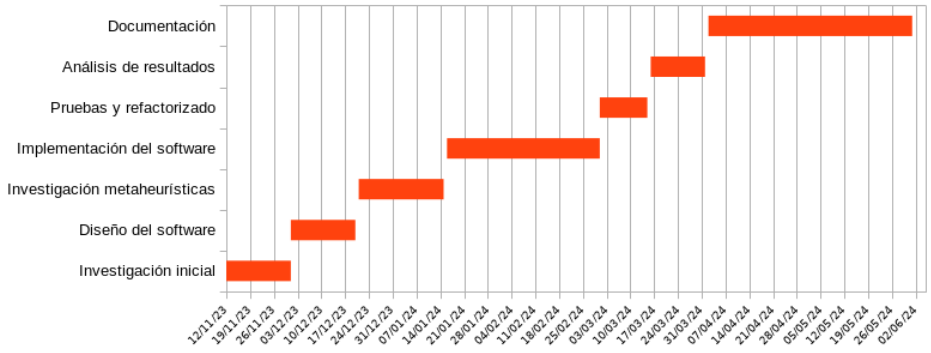

Figura 1.2: Diagrama de Gantt final

El coste estimado del proyecto se divide en varios subcostes:

- **Sueldo**: Considerando un precio de proyecto que deba cubrir gastos y compensar, se estima un salario por investigador de 52.000€ al año, es decir, 25€/hora. Al durar el proyecto aproximadamente 8 meses, el salario final total es de 34.000€.

- **Ordenador portátil**: El estudiante consta con un portátil HP 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 [14] entonces el costo del ordenador sería de 133.3€.

- **Servidor** Para la experimentación del proyecto se ha hecho uso de un servidor para ejecutar las distintas pruebas de los algoritmos. 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 1.1: 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 1.2: 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	34.000
Ordenador portátil	133.3
Servidor CPU - GC Compute Engine	75.84
Total	**34.209,14**

Tabla 1.3: Costo estimado del proyecto

Capítulo 2

Fundamentos Teóricos

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

2.1. 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 interdisciplinar 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.

2.1.1. 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 [15]:

Función objetivo a minimizar	$f(x)$	(2.1a)
Sujeto a		
s restricciones de desigualdad	$g_i(x) \leq 0, \quad j = 1, 2, ..., s$	(2.1b)
w restricciones de igualdad	$h_j(x) = 0, \quad j = 1, 2, ..., w$	(2.1c)
Donde el número de variables es dado por	$x_i, \quad i = 1, 2, ..., n$	

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.

2.1.2. 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* evalúa la idoneidad de una solución dentro de una población de soluciones. Es decir, determina la calidad relativa de la solución respecto a otras alternativas.

Debe destacarse que la métrica de la función de *fitness* suele ser estrictamente positiva, ya que representa la calidad de una solución, generalmente en un rango de valores entre 0 y 1. Por otro lado, la función objetivo puede ser positiva o negativa, dependiendo de si se busca maximizar o minimizar el objetivo. Además, la función de aptitud también puede llegar a ser una aproximación de la función objetivo, pero no necesariamente coinciden exactamente.

Resumiendo, podría decirse que la función *fitness* es un tipo particular de función objetivo que se utiliza como métrica de rendimiento [16].

2.1.3. Puntos 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 puntos locales en cambio son varios, no solo uno como es el global. Son soluciones máximas o mínimas dentro de una región de soluciones.

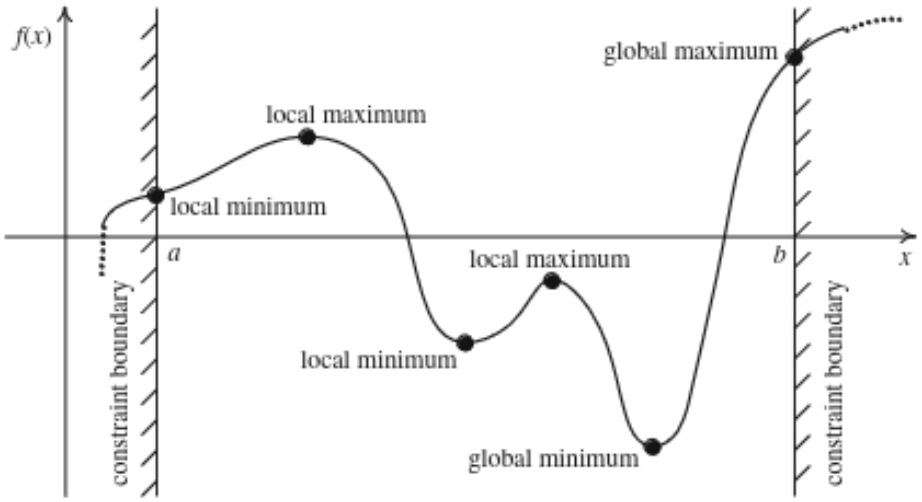

Figura 2.1: En esta figura extraida de [15] puede observarse de manera intuitiva la diferencia entre punto global y puntos locales. El máximo global es el valor más alto para $f(x)$ en el espacio, mientras que un máximo global solo representa el valor más alto dentro de un "vecindario"

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 [15]:

$$f(x^*) \geq f(x) \quad \forall x \tag{2.2}$$

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

$$\begin{aligned} f(x^*) \geq f(x) \quad & \forall x \\ & \text{dentro de un vecindario de } x^* \text{ [15]} \end{aligned} \tag{2.3}$$

2.2. Teoría de la computación

La teoría de la computación es un campo que estudia los fundamentos matemáticos y los límites de la computación. En esencia, se ocupa de preguntas fundamentales sobre qué se puede y qué no se puede computar, cómo se pueden resolver problemas de manera eficiente y qué tan difíciles son ciertos problemas en términos de recursos computacionales.

Suele usarse la máquina de Turing, como modelo abstracto de computadora, para las demostraciones realizadas, creando un marco sencillo en el que entender cómo se ejecutan los algoritmos y cómo se resuelven estos.

2.2.1. Máquina de Turing

Una máquina de Turing, cuyo nombre proviene de su creador Alan Turing, es un modelo de computación matemático que es capaz de ejecutar algoritmos mediante el control de una serie de símbolos en la tira de una cinta y de acuerdo a una serie de reglas establecidas [17].

Es capaz de general un modelo abstracto mediante el cual representar todo de algoritmos que resuelvan problemas computables.

2.2.2. Grupos de complejidad P y NP

Se puede definir un lenguaje como una serie de cadenas dentro de un alfabeto, siendo este último un conjunto finito de símbolos [2].

Dada esta definición se dice que un lenguaje que es *reconocible*, es decir, que una máquina de Turing puede determinar si una cadena pertenece al lenguaje o no, en un tiempo "polinómico determinista"son conocidos como problema de la clase P. Que un problema sea resoluble en tiempo polinómico determinista en una máquina de Turing, significa que para cada estado solo existe una posible acción y que su tiempo de resolución puede expresarse como una función polinómica [2].

En contraposición, un problema de la clase NP no es resoluble en tiempo polinómico determinista, sino en tiempo polinómico no determinista. Esto es, que puede resolverse expresando el tiempo como una función polinómica, pero con la diferencia de que el no determinismo implica que para pasar de un estado a otro es posible considerar múltiples acciones.

La diferencia entre los problemas de las clases P y NP es fundamental en la teoría de la complejidad computacional. Mientras que los problemas de la clase P son eficientemente solucionables, los problemas de la clase NP pueden ser verificados en tiempo polinómico, pero su resolución en tiempo polinómico no está demostrada y tiende a ser exponencial.

2.2.3. NP-Hard

Un problema de decisión H se considera NP-Hard o duro si, para cada problema L en NP, hay una reducción de muchos a uno de tiempo polinómico de L a H [2, 3]. Esto significa que cualquier problema en NP puede ser

transformado en H en tiempo polinómico, lo que sugiere que H es al menos tan difícil como cualquier problema en NP.

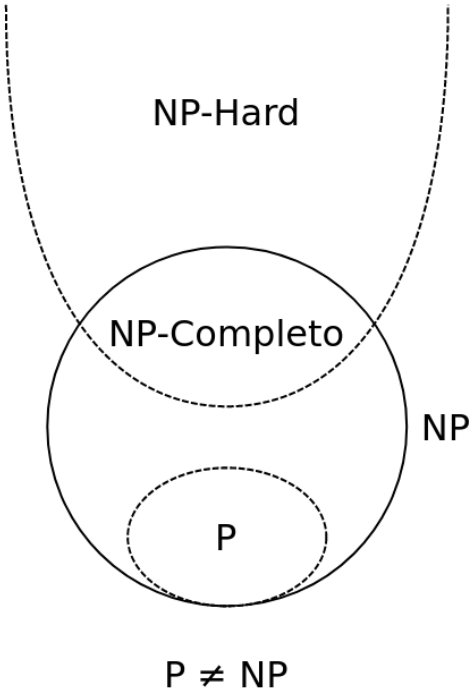

Figura 2.2: En esta figura se puede ver la relación de pertenencia entre clases de complejidad. Puede observarse que NP-Hard no es un subconjunto de NP. Una solución para un problema NP-Hard podría transformarse mediante reducción y solucionar un problema NP, pero no siempre es así al revés.

2.3. 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.

2.3.1. 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.

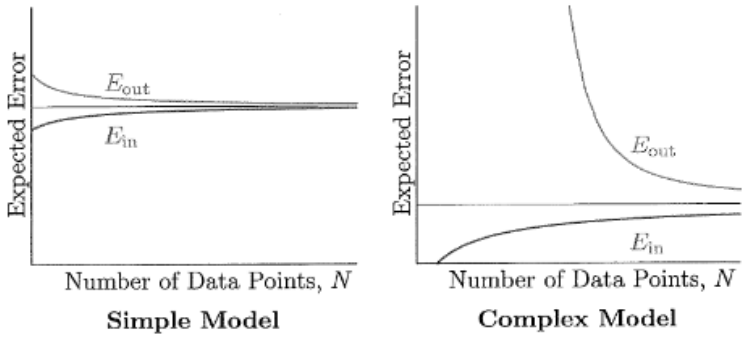

Figura 2.3: Esta figura extraída de [4] visualiza la relación directa entre calidad del modelo (menos error esperado) y número de datos que se le proporciona al algoritmo de aprendizaje.

Es inmediato pensar que a mayor cantidad de datos, mejor calidad del modelo, y así es. Sin embargo, tambíen 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 sobreajuste.
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 2.3, 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 puntos) suficientemente grande, la tendencia del error es a la baja [4, 18]. 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.

2.3.2. Concepto

El funcionamiento y motivo es explicado ya en la definición del problema en 1.1. Por evitar redundancias se procede a explicar el concepto de manera abreviada y puntualizando 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 2.2.2. 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** [5, 6].

2.3.3. Maldición de la dimensionalidad

El término fue acuñado por Bellman en 1961 [19]. 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 [20].

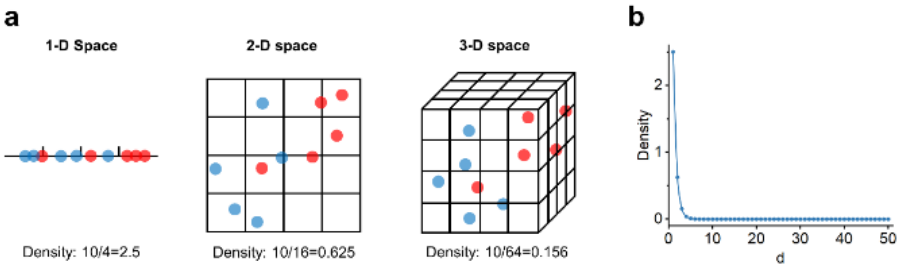

Figura 2.4: Esta figura extraída de [20] muestra la tendencia en la densidad a medida que se incrementa la dimensionalidad del espacio. En (a) se muestra la densidad con diez puntos de datos en un espacio $1D$, en (b) se muestra la tendencia al incrementar las dimensiones.

Además, en espacios de alta dimensión, la medición de distancias puede volverse inválida. 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 [5, 20]. En el trabajo de Beyer y compañeros [21] 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.

2.4. 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 [22]. 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 [22].

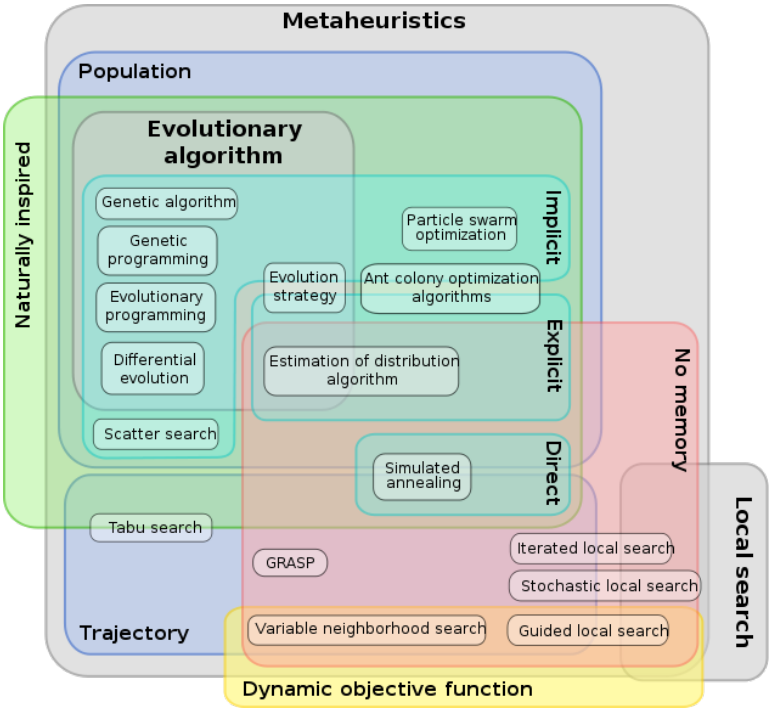

Figura 2.5: En esta figura de los autores Johann "nojhan" Dréo y Caner Candan, se clasifican las distintas metaheurísticas por inspiración.

2.4.1. 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 mejorarlas localmente, dentro de un "vecindario" [23].

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 [24]. Las interacciones entre estos individuos son las que conforman comportamientos exploratorios o explotativos y dependen del problema a solucionar y del propio algoritmo su equilibrio.

2.5. 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 [25].

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 [26].

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

2.6. Aprendizaje automático

El aprendizaje automático es una subrama 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 [27]. 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 recoleción 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.

2.6.1. 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 [28]. Es bastante costoso porque etiquetar cada dato es una tarea laboriosa y poco escalable.

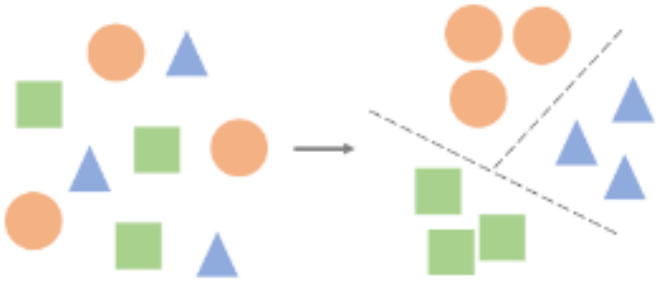

Figura 2.6: Figura extraída de [28] que muestra un ejemplo de aprendizaje supervisado.

2.6.2. 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* [28].

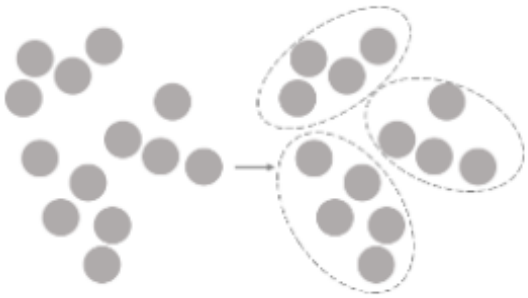

Figura 2.7: Figura extraída de [28] que muestra un ejemplo de aprendizaje no supervisado. Se trata de dividir en clústers sin información de etiquetas.

2.6.3. SVM

Las máquinas de vectores de soportes o *SVM* fueron introducidas por primera vez en [10]. 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

Capítulo 3

Bibliografía

[1] 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).

[2] 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.

[3] 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).

[4] Y. S. Abu-Mostafa, M. Magdon-Ismail y H.-T. Lin, *Learning From Data*. AMLBook, 2012.

[5] N. Venkat, *The Curse of Dimensionality: Inside Out*, sep. de 2018. DOI: 10.13140/RG.2.2.29631.36006.

[6] R. Bellman, *Dynamic Programming*. Princeton Univ Pr, 1957, ISBN: 978-0691079516. URL: http://libgen.li/file.php?md5=3a1794f608b48cbd4bce640735af75d2.

[7] Udacity, *Curse of Dimensionality - Georgia Tech - Machine Learning*, Retrieved 2022-06-29, feb. de 2015.

[8] K. Kira y L. Rendell, «A Practical Approach to Feature Selection,» English, 1992, págs. 249-256, ISBN: 978-1-55860-247-2. DOI: 10.1016/B978-1-55860-247-2.50037-1.

[9] 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.

[10] 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).

[11] 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).

[12] 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).

[13] W. Clark, W. N. Polakov y F. W. Trabold, *The Gantt chart, a working tool of management*, English. New York: The Ronald press company, 1922.

[14] 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).

[15] A. Kulkarni, G. Krishnasamy y A. Abraham, «Introduction to Optimization,» en sep. de 2017, vol. 114, págs. 1-7, ISBN: 978-3-319-44253-2. DOI: 10.1007/978-3-319-44254-9_1.

[16] 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.

[17] J. L. Stone, *Introduction to Computer Organization and Data Structures*. New York: McGraw-Hill, 1972.

[18] S. Shalev-Shwartz y S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. CUP, 2014, ISBN: 978-3-319-44253-2.

[19] R. E. Bellman, *Adaptive Control Processes*. Princeton University Press, 1961.

[20] 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).

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] D. Wolpert y W. 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.

[26] 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].

[27] K. P. Murphy, *Machine Learning: A Probabilistic Perspective* (Adaptive Computation and Machine Learning), 1.ª ed. The MIT Press, 2012, ISBN: 0262018020, 9780262018029. URL: https://libgen.li/file.php?md5=8ecfeeb2e1f9a19c770fba1ff85fa566.

[28] S. Sah, *Machine Learning: A Review of Learning Types*, jul. de 2020. DOI: 10.20944/preprints202007.0230.v1.