

Mathematical Foundations for Machine Learning

Helga Kelly Quiroz-Chavil¹ , Carlos Enrique Capuñay-Uceda² , Luis Jaime Collantes Santisteban³ , Samuel Collantes Santisteban⁴ , Carlos Alberto Collantes Santisteban⁵ , Kelly Scarlett Collantes Alvarado⁶ 

^{2,3,5,6} Universidad Nacional Pedro Ruiz Gallo, ccapunay@unprg.edu.pe, lcollantes@unprg.edu.pe, ccollantess@unprg.edu.pe, kcollantes@unprg.edu.pe

^{1,2} Universidad Tecnológica del Perú, c22682@utp.edu.pe

⁴ Universidad San Martín de Porras, scollantess@usmp.pe

Abstract– This paper has conducted a comprehensive review of more than 30 machine learning (ML) and deep learning (DL) algorithms across four main categories: supervised, unsupervised, reinforcement and deep learning. The discussion focused on the description of each algorithm, the application of mathematical foundations in their implementation, and their relevance to various practical applications. Through this analysis, the critical importance of underlying mathematical and statistical concepts, such as optimization, probability theory, and geometry, in the development of ML and DL models was highlighted. A key conclusion of the paper is the diversity and adaptability of ML and DL algorithms in a wide range of fields, including computer vision, natural language processing, robotics, and medicine. This analysis underscores how rapidly the field is advancing, marked by the evolution of complex models that have transformed machine learning and adaptive capabilities. However, the work also recognizes the challenges and limitations that exist in the design and implementation of these algorithms, including overfitting, interpretability, and computational consumption. These challenges underscore the need for continued research and development to optimize and create new techniques that overcome these barriers. Looking to the future, the work suggests a focus on developing even more generative models.

Keywords– Mathematics, machine learning, algorithms.

Digital Object Identifier: (only for full papers, inserted by LACCEI).

ISSN, ISBN: (to be inserted by LACCEI).

DO NOT REMOVE

Fundamentos matemáticos para Machine Learning

Helga Kelly Quiroz-Chavil¹ , Carlos Enrique Capuñay-Uceda² , Luis Jaime Collantes Santisteban³ , Samuel Collantes Santisteban⁴ , Carlos Alberto Collantes Santisteban⁵ , Kelly Scarlett Collantes Alvarado⁶ 

^{2,3,5,6} Universidad Nacional Pedro Ruiz Gallo, ccapunay@unprg.edu.pe, lcollantes@unprg.edu.pe, ccollantess@unprg.edu.pe, kcollantes@unprg.edu.pe

^{1,2} Universidad Tecnológica del Perú, c22682@utp.edu.pe

⁴ Universidad San Martín de Porras, scollantess@usmp.pe

Resumen— Este trabajo ha realizado una revisión exhaustiva de más de 30 algoritmos de machine learning (ML) y deep learning (DL) a través de cuatro categorías principales: aprendizaje supervisado, no supervisado, por refuerzo y deep learning. La discusión se centró en la descripción de cada algoritmo, la aplicación de fundamentos matemáticos en su implementación y su relevancia en diversas aplicaciones prácticas. A través de este análisis, se destacó la importancia crítica de los conceptos matemáticos y estadísticos subyacentes, como la optimización, la teoría de la probabilidad, y la geometría, en el desarrollo de modelos de ML y DL. Una conclusión clave del trabajo es la diversidad y la adaptabilidad de los algoritmos de ML y DL en una amplia gama de campos, incluyendo la visión por computadora, el procesamiento del lenguaje natural, la robótica y la medicina. Este análisis subraya la rapidez con la que avanza el campo, marcado por la evolución de modelos complejos que han transformado la capacidad de aprendizaje y adaptación de las máquinas. Sin embargo, el trabajo también reconoce los desafíos y limitaciones existentes en el diseño y la implementación de estos algoritmos, incluyendo el sobreajuste, la interpretabilidad y el consumo computacional. Estos desafíos subrayan la necesidad de investigación y desarrollo continuos para optimizar y crear nuevas técnicas que superen estas barreras. Mirando hacia el futuro, el trabajo sugiere un enfoque en el desarrollo de modelos aún más generalizables y eficientes, con un aprendizaje más autónomo y menos dependiente de la intervención humana. La investigación en aprendizaje no supervisado y aprendizaje por refuerzo, en particular, promete avances significativos hacia sistemas de inteligencia artificial más completos y adaptativos. Este análisis integral de los algoritmos de ML y DL ilustra el papel fundamental de las matemáticas en el avance de la inteligencia artificial, al tiempo que enfatiza la importancia de la colaboración interdisciplinaria para superar desafíos actuales y desbloquear el potencial completo de estas tecnologías para resolver problemas complejos del mundo real.

Palabras clave: Matemática, machine learning, algoritmos.

I. INTRODUCCIÓN

La revolución digital y tecnológica en la que nos encontramos inmersos ha transformado radicalmente la forma en que comprendemos y aplicamos el conocimiento matemático, especialmente en el campo del aprendizaje automático (Machine Learning, ML). El aprendizaje automático, una rama esencial de la inteligencia artificial (IA), se ha convertido en una herramienta indispensable para el análisis de datos, la predicción de eventos y la toma de decisiones basada en datos. Este artículo pretende explorar los fundamentos matemáticos que sustentan el ML, destacando

cómo conceptos de álgebra lineal, cálculo, estadística y teoría de la probabilidad forman la columna vertebral de esta disciplina revolucionaria.

La importancia del álgebra lineal en el ML no puede ser subestimada, ya que proporciona las herramientas para manipular los conjuntos de datos representados en matrices y vectores, facilitando operaciones como transformaciones lineales, descomposiciones matriciales, y soluciones de sistemas lineales, que son críticos para algoritmos de ML como la regresión lineal y la descomposición en valores singulares (SVD) [1]. Por otro lado, el cálculo, especialmente el cálculo diferencial, juega un papel crucial en la optimización de funciones de coste y la retropropagación en redes neuronales, permitiendo el ajuste de parámetros para minimizar el error de predicción [2].

La estadística y la teoría de la probabilidad son fundamentales para entender la inferencia en el ML. Permiten comprender la variabilidad de los datos, evaluar la incertidumbre de las predicciones y tomar decisiones informadas a pesar de la incertidumbre [3]. Estas herramientas matemáticas no solo permiten desarrollar nuevos algoritmos de ML, sino también mejorar la interpretación de los resultados obtenidos de los modelos de ML.

El Machine Learning se asienta sobre una base matemática sólida que no solo es esencial para el desarrollo de algoritmos eficientes, sino también para garantizar que estos algoritmos sean interpretables y fiables. Este artículo se adentra en estas áreas matemáticas, proporcionando una comprensión detallada de cómo cada una contribuye al campo del ML, con el objetivo de ofrecer a los lectores una base sólida sobre la cual construir o mejorar su comprensión del aprendizaje automático.

II. MATEMÁTICA EN LOS ALGORITMOS

Los algoritmos de Machine Learning (ML) se pueden clasificar ampliamente en tres categorías principales: supervisado, no supervisado y aprendizaje por refuerzo. Cada tipo aborda diferentes aspectos y desafíos de aprendizaje, desde la predicción basada en datos etiquetados hasta el descubrimiento de estructuras ocultas en datos no etiquetados y la optimización de acciones en un entorno dado.

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

Aprendizaje Supervisado: Los algoritmos de aprendizaje supervisado utilizan un conjunto de datos etiquetados para entrenar el modelo, lo que significa que cada entrada de datos tiene una salida correspondiente. El objetivo es aprender una función que, dada una entrada, produzca la salida esperada. Este tipo de aprendizaje es útil para tareas de regresión y clasificación. Ejemplos comunes incluyen regresión lineal y logística, máquinas de vectores de soporte (SVM), y redes neuronales [1].

Aprendizaje No Supervisado: A diferencia del aprendizaje supervisado, los algoritmos de aprendizaje no supervisado trabajan con datos que no están etiquetados, buscando patrones, agrupaciones o correlaciones intrínsecas. Estos algoritmos intentan organizar la información de manera que un humano o un sistema supervisado pueda entenderla mejor. Las aplicaciones comunes incluyen el análisis de agrupamiento y la reducción de dimensionalidad, con ejemplos como el algoritmo K-means y el análisis de componentes principales (PCA) [4].

Aprendizaje por Refuerzo: El aprendizaje por refuerzo se basa en la idea de agentes que toman decisiones y aprenden a través de la retroalimentación de las consecuencias de sus acciones en un entorno. No se trabaja con datos etiquetados, sino con recompensas o penalizaciones basadas en las acciones realizadas. Este tipo de aprendizaje se utiliza para desarrollar sistemas que toman decisiones secuenciales, como en juegos, navegación de robots, y sistemas de recomendación [5].

Matemática en algoritmos de Aprendizaje Supervisado

Los algoritmos de aprendizaje supervisado son fundamentales en el campo del machine learning, permitiendo a los modelos aprender a partir de datos etiquetados para predecir resultados en datos nuevos. En este trabajo analizamos los 10 algoritmos de aprendizaje supervisado más utilizados, en cada uno hemos identificado los fundamentos matemáticos que utilizan:

Regresión Lineal: Utiliza la relación lineal entre las variables independientes y la variable dependiente para predecir el valor de esta última. Matemáticamente, se busca minimizar la suma de los cuadrados de las diferencias entre los valores observados y los predichos por el modelo lineal [1].

Regresión Logística: Aunque se llama regresión, se utiliza para clasificación binaria. Se basa en la función logística para predecir la probabilidad de que una entrada pertenezca a una categoría. Matemáticamente, se utiliza la función logística para modelar la probabilidad y se aplica el método de máxima verosimilitud para estimar los parámetros [4].

Máquinas de Vectores de Soporte (SVM): Busca el hiperplano que mejor separa las clases en el espacio de características. Los fundamentos matemáticos incluyen conceptos de geometría

vectorial y optimización cuadrática para maximizar el margen entre las clases [5].

K-Nearest Neighbors (K-NN): Clasifica las entradas basándose en las 'k' observaciones más cercanas en el espacio de características. La distancia entre los puntos se calcula utilizando métricas como la distancia euclidiana, lo cual se fundamenta en la geometría matemática [4].

Árboles de Decisión: Modelan las decisiones y sus posibles consecuencias como un árbol. Utilizan algoritmos como ID3 o C4.5, que aplican conceptos de teoría de la información y entropía para seleccionar las características que mejor dividen el conjunto de datos [5].

Bosques Aleatorios (Random Forests): Consiste en un conjunto de árboles de decisión, donde cada árbol se entrena con una muestra aleatoria de los datos. La decisión final se toma por mayoría de votos. Utiliza conceptos de estadística y probabilidad para estimar la importancia de las características y para la decisión final [1].

Boosting (AdaBoost): Combina múltiples modelos débiles para crear un modelo fuerte. Matemáticamente, se enfoca en ajustar los pesos de los clasificadores débiles y de las instancias de entrenamiento para minimizar el error de predicción del modelo combinado [4].

Redes Neuronales Artificiales: Inspiradas en las redes neuronales biológicas, estas redes aprenden a realizar tareas considerando ejemplos, sin estar programadas con reglas específicas. Utilizan el cálculo diferencial para la optimización de los pesos de la red mediante el algoritmo de retropropagación (Backpropagation) [5].

Naive Bayes: Un clasificador probabilístico basado en el teorema de Bayes con la suposición de independencia entre las características. Aplica conceptos de probabilidad y estadística para calcular la probabilidad posterior de cada clase, dadas las características observadas [1].

Gradient Boosting Machines (GBM): Construye de forma iterativa modelos de árboles de decisión en etapas, donde cada árbol nuevo se crea para corregir los errores cometidos por los árboles anteriores. Se basa en el descenso del gradiente, un método de optimización en el espacio de las funciones de error, para minimizar el error de predicción [4].

Matemática en algoritmos de Aprendizaje No Supervisado

El aprendizaje no supervisado permite descubrir patrones ocultos en datos sin etiquetar, facilitando la comprensión de la estructura subyacente o la distribución de un conjunto de datos. A continuación, se detallan los 10 algoritmos de aprendizaje no

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

supervisado más utilizados, enfocándose en cómo los fundamentos matemáticos son aplicados en su implementación:

K-Means Clustering: Agrupa los datos en 'k' número de clusters basándose en la proximidad de los puntos de datos a los centroides del cluster. Utiliza la distancia euclidiana para medir la proximidad, minimizando la suma de las distancias cuadradas entre los puntos de datos y el centroide de su cluster [6].

Análisis de Componentes Principales (PCA): Reduce la dimensionalidad del conjunto de datos proyectándolo en un número menor de dimensiones (componentes principales) que explican la mayor varianza de los datos. Matemáticamente, PCA utiliza la descomposición en valores singulares (SVD) para encontrar las direcciones de máxima varianza en los datos [7].

Análisis de Cluster Jerárquico: Construye un dendrograma (árbol) que representa las relaciones de proximidad entre los datos. Puede utilizar diferentes métricas de distancia (como la euclidiana o la de Manhattan) y métodos de enlace (como el enlace completo o el enlace promedio) para calcular la proximidad entre clusters [8].

DBSCAN (Density-Based Spatial Clustering of Applications with Noise): Identifica clusters basándose en la densidad de los datos, agrupando puntos que están estrechamente juntos y marcando como outliers los puntos en regiones de baja densidad. Este algoritmo utiliza un concepto matemático de densidad que requiere dos parámetros: 'eps' (distancia máxima entre dos puntos para ser considerados vecinos) y 'minPts' (número mínimo de puntos para formar un cluster denso) [9].

t-Distributed Stochastic Neighbor Embedding (t-SNE): Reduce la dimensionalidad de los datos para facilitar su visualización, manteniendo la similitud entre los puntos. Matemáticamente, convierte las similitudes entre los puntos de datos en probabilidades conjuntas y minimiza la divergencia de Kullback-Leibler entre las probabilidades conjuntas de los datos de alta dimensión y los datos de baja dimensión [10].

Algoritmo de Expectation-Maximization (EM) para Gaussian Mixture Models (GMM): Un modelo de mezcla gaussiana representa una distribución de subpoblaciones dentro de una población total. EM encuentra los parámetros que maximizan la verosimilitud de la distribución de mezcla, asignando probabilísticamente cada dato a una subpoblación [11].

Algoritmo Apriori: Utilizado para análisis de asociación, identifica conjuntos de ítems que ocurren frecuentemente juntos en transacciones. Matemáticamente, se basa en la propiedad de que todos los subconjuntos de un conjunto de ítems frecuente también son frecuentes [12].

Autoencoders: Son redes neuronales utilizadas para reducción de dimensionalidad y aprendizaje de características. A través de la codificación y decodificación de los datos, un autoencoder aprende una representación comprimida, donde la función de pérdida minimiza la diferencia entre los datos originales y su reconstrucción [13].

Isomap (Isometric Mapping): Es una técnica de reducción de dimensionalidad que mantiene las distancias geodésicas entre todos los pares de puntos. Matemáticamente, Isomap utiliza técnicas de grafos para estimar las distancias geodésicas y luego aplica MDS (Multidimensional Scaling) para encontrar la representación en baja dimensión que mejor preserva estas distancias [14].

Local Linear Embedding (LLE): Reduce la dimensionalidad manteniendo las relaciones lineales locales entre los puntos. LLE reconstruye cada punto como una combinación lineal de sus vecinos más cercanos y luego busca una representación en baja dimensión que preserva estas relaciones locales [15].

Matemática en Algoritmos de Aprendizaje por Refuerzo

El aprendizaje por refuerzo (Reinforcement Learning, RL) se centra en enseñar a un agente a tomar decisiones optimizando recompensas a través de interacciones con un entorno. A continuación, se describen los 10 algoritmos de RL más utilizados, destacando cómo se aplican los fundamentos matemáticos en su implementación:

Q-Learning: Este es un método de diferencias temporales que aprende la función de valor de acción (Q-function) que estima la recompensa total esperada de tomar una acción en un estado particular. Matemáticamente, utiliza la ecuación de Bellman para actualizar iterativamente las estimaciones de Q-values basándose en las recompensas recibidas [16].

SARSA (State-Action-Reward-State-Action): Similar a Q-learning, SARSA es un algoritmo de control on-policy que actualiza su función Q basándose en la acción tomada y la política actual. Utiliza la ecuación de Bellman para las políticas on-policy para realizar sus actualizaciones [17].

Deep Q-Network (DQN): Combina Q-learning con redes neuronales profundas para aproximar la función Q. Introduce técnicas como la repetición de experiencia y los Q-values objetivo para estabilizar el entrenamiento. La red se entrena minimizando la diferencia entre el Q-value predicho y el Q-value objetivo, usando descenso de gradiente [18].

Policy Gradient Methods: Estos métodos aprenden directamente la política de acciones que maximiza las recompensas esperadas. Matemáticamente, ajustan los parámetros de la política en la dirección del gradiente

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

ascendente del rendimiento esperado, usando el teorema del gradiente de política para calcular gradientes [19].

Actor-Critic Methods: Combinan las ideas de los métodos de gradientes de política (actor) con las de aprender la función de valor (crítico). El crítico actualiza la función de valor, y el actor actualiza la política en la dirección sugerida por el crítico, usando el gradiente de la función de valor como aproximación del gradiente de política [20].

Proximal Policy Optimization (PPO): Optimiza la política de acciones tomando pasos de tamaño pequeño para evitar cambios bruscos en la política, lo que mejora la estabilidad del entrenamiento. Matemáticamente, utiliza una función objetivo recortada para mantener la actualización de la política dentro de un rango determinado [21].

Trust Region Policy Optimization (TRPO): Garantiza mejoras monótonas en la política mediante la optimización de una función objetivo sujeta a una restricción de región de confianza, lo que asegura que las actualizaciones de la política no se desvíen demasiado de la política anterior [22].

Asynchronous Advantage Actor-Critic (A3C): Permite que múltiples instancias de agentes trabajen en paralelo en múltiples entornos para acelerar el entrenamiento. Utiliza un enfoque actor-crítico donde el crítico estima la función de ventaja, que indica cuánto mejor es una acción comparada con la política actual [23].

Soft Actor-Critic (SAC): Es un algoritmo actor-crítico que optimiza una política estocástica en un marco de RL de entropía máxima, promoviendo la exploración mediante la maximización de la entropía además de la recompensa [24].

Monte Carlo Tree Search (MCTS): Utilizado en juegos de suma cero con información perfecta, como el Go, explora posibles movimientos de forma selectiva para construir un árbol de búsqueda y utiliza la simulación para estimar el valor de cada movimiento. Matemáticamente, equilibra la exploración y explotación utilizando la fórmula UCB (Upper Confidence Bound) para seleccionar movimientos [25].

Una de las ramas del Machine Learning que ha evolucionado mucho en los últimos años es el Deep Learning o Aprendizaje Profundo. El aprendizaje profundo ha revolucionado la forma en que se abordan los problemas de visión por computadora, procesamiento de lenguaje natural, y muchos otros campos, gracias a su capacidad para aprender representaciones complejas de grandes volúmenes de datos. A continuación, se describen los 10 algoritmos de aprendizaje profundo más utilizados, resaltando cómo se aplican los fundamentos matemáticos en su implementación:

Redes Neuronales Convolucionales (CNNs): Especializadas en procesar datos con una topología de cuadrícula, como imágenes. Las CNNs utilizan la operación matemática de convolución en lugar de la multiplicación general de matrices. Esto reduce el número de parámetros, aprovechando la localidad espacial de los datos [26].

Redes Neuronales Recurrentes (RNNs): Diseñadas para procesar secuencias de datos, como el lenguaje natural. Las RNNs utilizan bucles para pasar información de un paso del tiempo al siguiente, modelando temporalmente las dependencias [27].

Long Short-Term Memory Networks (LSTMs): Una variante de las RNNs que puede aprender dependencias a largo plazo. Las LSTMs utilizan puertas de olvido, entrada y salida para regular el flujo de información, evitando el problema del desvanecimiento del gradiente mediante el diseño de una estructura de celda que permite mantener el estado a lo largo del tiempo [27].

Autoencoders: Redes neuronales utilizadas para aprendizaje no supervisado de codificaciones eficientes. Los autoencoders se entrenan para minimizar la reconstrucción de errores, aprendiendo una representación comprimida de los datos de entrada, generalmente a través de una función de pérdida como el error cuadrático medio [13].

Redes Generativas Adversarias (GANs): Consisten en dos redes, un generador y un discriminador, que se entrenan simultáneamente mediante un juego adversario. El generador intenta producir datos falsos indistinguibles de los reales, mientras que el discriminador intenta distinguir entre datos reales y falsos. Matemáticamente, esto se formula como un juego de suma cero o un problema de minimax [28].

Transformers: Modelo basado en la atención que ha revolucionado el procesamiento del lenguaje natural. Los Transformers utilizan mecanismos de atención para ponderar la importancia relativa de diferentes palabras en una secuencia. La atención se calcula utilizando productos escalares de vectores de consulta, clave y valor [29].

U-Net: Especialmente diseñada para la segmentación de imágenes biomédicas. La U-Net utiliza una arquitectura de codificador-decodificador con conexiones de salto para preservar la información de contexto y permitir una segmentación precisa. La efectividad de la U-Net radica en su capacidad para capturar tanto el contexto como los detalles locales a través de su arquitectura simétrica [30].

ResNet (Redes Residuales): Introduce conexiones residuales que permiten a las señales saltarse ciertas capas, facilitando el entrenamiento de redes muy profundas al mitigar el problema del desvanecimiento del gradiente. Las ResNets implementan

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

estas conexiones residuales sumando la entrada de una capa a su salida, facilitando el flujo de gradientes durante el entrenamiento [31].

Capsule Networks (CapsNets): Propone una arquitectura que utiliza cápsulas para modelar jerarquías de características en espacios de entidades. Las CapsNets buscan mejorar la representación espacial de las CNNs mediante el uso de vectores de cápsulas y la "routing by agreement" para enviar información solo a las cápsulas de nivel superior con las que están de acuerdo [32].

BERT (Bidirectional Encoder Representations from Transformers): Modelo preentrenado de procesamiento de lenguaje natural que utiliza transformadores para entender el contexto de las palabras en textos basándose en ambos lados de una palabra dentro de una oración. BERT se entrena con un conjunto de tareas, como la predicción de palabras enmascaradas y la predicción de la siguiente oración, para aprender representaciones ricas en contexto [33].

A lo largo de este trabajo, hemos explorado una amplia variedad de algoritmos de machine learning y deep learning, los cuales se fundamentan en numerosos métodos, teoremas y conceptos matemáticos. A continuación, se presenta una lista de estos fundamentos matemáticos mencionados en el análisis de los algoritmos:

Optimización: Incluye la minimización o maximización de funciones, utilizada en la mayoría de los algoritmos para ajustar los parámetros del modelo.

Teoría de la Probabilidad: Fundamental para modelar la incertidumbre y calcular probabilidades en algoritmos como Naive Bayes y Gaussian Mixture Models.

Estadística: Incluye conceptos como la media, varianza, desviación estándar, y más, cruciales para el análisis de datos y la inferencia estadística.

Álgebra Lineal: Engloba matrices, vectores, descomposición en valores singulares (SVD), y otras operaciones matriciales esenciales para el procesamiento de datos y la estructura de muchos algoritmos.

Geometría: Utilizada en el análisis y visualización de datos, especialmente en la reducción de dimensionalidad y en la construcción de modelos geométricos de datos.

Cálculo Diferencial e Integral: Aplicado en la derivación de funciones de coste y en el entrenamiento de modelos mediante el descenso de gradiente.

Ecuación de Bellman: Fundamental en el aprendizaje por refuerzo, utilizada para describir la relación entre el valor actual y los valores futuros esperados.

Teorema del Gradiente de Política: Permite calcular gradientes de políticas en métodos de gradiente de política para el aprendizaje por refuerzo.

Convolución: Operación matemática clave en las redes neuronales convolucionales para el procesamiento de imágenes.

Funciones de Activación: Incluyen la sigmoide, tanh, ReLU, y otras, esenciales para introducir no linealidades en los modelos de red neuronal.

Entropía y Entropía Cruzada: Conceptos de la teoría de la información utilizados para medir la incertidumbre y el error en las predicciones de los modelos.

Divergencia de Kullback-Leibler: Mide cuánto se diferencia una distribución de probabilidad de otra, utilizada en métodos como t-SNE y en la regularización de modelos.

Métodos de Montecarlo: Técnicas de simulación para la estimación numérica, importantes en Monte Carlo Tree Search y en la evaluación de políticas en aprendizaje por refuerzo.

Backpropagation: Método de optimización basado en la regla de la cadena del cálculo para entrenar redes neuronales.

Producto Escalar y Normas Vectoriales: Utilizados en el cálculo de distancias y similitudes entre vectores, fundamentales en algoritmos como K-NN y en la arquitectura de los Transformers.

Región de Confianza y Algoritmos de Optimización: Técnicas avanzadas de optimización como TRPO y PPO en aprendizaje por refuerzo utilizan estas para asegurar mejoras estables en la política de decisiones.

Esta lista resume los conceptos matemáticos clave que subrayan la complejidad y el poder de los algoritmos de ML y DL, demostrando la importancia de las matemáticas en la fundación y evolución de la inteligencia artificial.

CONCLUSIONES

Este trabajo ha explorado una amplia gama de algoritmos en los dominios del aprendizaje supervisado, no supervisado, por refuerzo, y profundo, destacando la diversidad y profundidad de las técnicas disponibles en el campo del machine learning (ML) y deep learning (DL). A través del análisis de más de 30 algoritmos, hemos identificado aspectos clave que subrayan la importancia de los fundamentos matemáticos y estadísticos en el desarrollo e implementación de estos métodos.

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

La implementación efectiva de algoritmos de ML y DL se basa en una comprensión sólida de conceptos matemáticos y estadísticos, como la optimización, la teoría de la probabilidad, la geometría, y el álgebra lineal. Estos fundamentos no solo permiten la creación de modelos precisos y eficientes, sino que también facilitan la interpretación y la mejora de estos.

Los algoritmos analizados demuestran la versatilidad del ML y DL, encontrando aplicaciones en una multitud de campos, desde la visión por computadora y el procesamiento del lenguaje natural hasta la robótica y la medicina. Esta diversidad refleja la capacidad de los modelos de ML y DL para adaptarse y extraer insights valiosos de diferentes tipos de datos.

La evolución de los algoritmos, especialmente en el área del aprendizaje profundo, subraya la rapidez con la que avanza el campo. Modelos como las redes neuronales convolucionales, LSTMs, GANs, y Transformers, han marcado hitos significativos, abriendo nuevas vías para la investigación y aplicaciones prácticas que eran inviables hace tan solo una década.

A pesar de los avances, existen desafíos inherentes en el diseño y la implementación de estos algoritmos, como el sobreajuste, la interpretabilidad, y el consumo computacional. Estos desafíos requieren un enfoque continuo en la investigación y desarrollo de nuevas técnicas y estrategias de optimización.

El futuro del ML y DL se proyecta hacia el desarrollo de modelos aún más generalizables y eficientes, que puedan aprender de manera más autónoma y con menos intervención humana. La investigación en áreas como el aprendizaje no supervisado y el aprendizaje por refuerzo promete avances significativos hacia sistemas de inteligencia artificial más avanzados y adaptativos.

En conclusión, este trabajo subraya el papel crucial de los fundamentos matemáticos en el ML y DL, así como la importancia de la investigación continua para superar desafíos existentes y explorar el potencial ilimitado de estas tecnologías. La colaboración interdisciplinaria entre matemáticos, estadísticos, ingenieros, y científicos de datos seguirá siendo esencial para impulsar la innovación y aplicar estos algoritmos de manera efectiva en la resolución de problemas complejos del mundo real.

REFERENCIAS

- [1] G. Kumar, R. Banerjee, D. Singh, N. Choubey y Arnaw, «Mathematics for Machine Learning,» *Journal of Mathematical Sciences & Computational Mathematics*, 2020.

- [2] M. Belkin, «Fit without fear: remarkable mathematical phenomena of deep learning through the prism of interpolation,» *Acta Numerica*, vol. 30, pp. 203-248, 2021.
- [3] F. Cucker y S. Smale, «On the mathematical foundations of learning,» *Bulletin of the American Mathematical Society*, vol. 39, n° 1, pp. 1-49, 2001.
- [4] A. Subasi, «Machine learning techniques,» de *Practical Machine Learning for Data Analysis Using Python*, Academic Press, 2020, pp. 91-202.
- [5] P. Atzberger, Importance of the Mathematical Foundations of Machine Learning Methods for Scientific and Engineering Applications, ArXiv, 2018.
- [6] A. Jain, «Data clustering: 50 years beyond K-means,» *Pattern Recognition Letters*, vol. 31, n° 8, pp. 651-666, 2010.
- [7] I. Jolliffe, Principal component analysis, Springer Series in Statistics, 2002.
- [8] F. Murtagh y P. Contreras, «Algorithms for hierarchical clustering: An overview,» *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, n° 1, pp. 86-97, 2012.
- [9] M. Ester, H. Kriegel, J. Sander y X. Xu, «A density-based algorithm for discovering clusters in large spatial databases with noise,» de *Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, KDD'96.*, 1996.
- [10] L. van der Maaten y G. Hinton, «Visualizing data using t-SNE,» *Journal of Machine Learning Research*, vol. 9, n° Nov, pp. 2579-2605, 2008.
- [11] A. P. Dempster, N. M. Laird y D. B. Rubin, «Maximum likelihood from incomplete data via the EM algorithm,» *Journal of the Royal Statistical Society, Series B, n° Methodological*, pp. 1-38, 1977.
- [12] R. Agrawal y R. Srikant, «Fast algorithms for mining association rules,» de *Proc. 20th Int. Conf. Very Large Data Bases, VLDB, 487-499.*, 1994.
- [13] G. E. Hinton y R. R. Salakhutdinov, «Reducing the dimensionality of data with neural networks,» *Science*, vol. 313, n° 5786, pp. 504-507, 2006.
- [14] J. B. Tenenbaum, V. de Silva y J. C. Langford, «A global geometric framework for nonlinear dimensionality reduction,» *Science*, vol. 290, n° 5500, pp. 2319-2323, 2000.
- [15] S. T. Roweis y L. K. Saul, «Nonlinear dimensionality reduction by locally linear embedding,» *Science*, vol. 290, n° 5500, pp. 2323-2326, 2000.
- [16] C. J. C. H. Watkins, «Learning from Delayed Rewards. PhD Thesis,» University of Cambridge, 1989.
- [17] G. A. Rummery y M. Niranjan, «On-line Q-learning using connectionist systems,» University of Cambridge, Department of Engineering, 1994.

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

- [18] V. e. a. Mnih, «Human-level control through deep reinforcement learning,» *Nature*, vol. 518, n° 7540, pp. 529-533, 2015.
- [19] R. S. e. a. Sutton, «Policy Gradient Methods for Reinforcement Learning with Function Approximation,» de *Advances in Neural Information Processing Systems*, 2000.
- [20] V. R. Konda y J. N. Tsitsiklis, «Actor-Critic Algorithms,» de *Advances in Neural Information Processing Systems 13.*, 2000.
- [21] J. e. a. Schulman, Proximal Policy Optimization Algorithms, arXiv, 2017.
- [22] J. e. a. Schulman, «Trust Region Policy Optimization.,» de *International Conference on Machine Learning*, 2015.
- [23] V. e. a. Mnih, «Asynchronous Methods for Deep Reinforcement Learning.,» de *International Conference on Machine Learning*, 2016.
- [24] T. e. a. Haarnoja, Soft Actor-Critic Algorithms and Applications, arXiv, 2018.
- [25] C. e. a. Browne, «A Survey of Monte Carlo Tree Search Methods.,» *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, n° 1, pp. 1-43, 2012.
- [26] Y. LeCun, L. Bottou, Y. Bengio y P. Haffner, «Gradient-Based Learning Applied to Document Recognition,» *Proceedings of the IEEE*, vol. 86, n° 11, pp. 2278-2324, 1998.
- [27] S. Hochreiter y J. Schmidhuber, «Long Short-Term Memory,» *Neural Computation*, vol. 9, n° 8, pp. 1735-1780, 1997.
- [28] I. P.-A. J. M. M. X. B. W.-F. D. O. S. C. A. & B. Y. Goodfellow, «Generative Adversarial Nets,» de *Advances in Neural Information Processing Systems 27*, 2014.
- [29] A. S. N. P. N. U. J. J. L. G. A. N. K. Ł. & P. I. Vaswani, «Attention Is All You Need.,» de *Advances in Neural Information Processing Systems 30*, 2017.
- [30] O. F. P. & B. T. Ronneberger, «U-Net: Convolutional Networks for Biomedical Image Segmentation,» de *Medical Image Computing and Computer-Assisted Intervention MICCAI*, 2015.
- [31] K. Z. X. R. S. & S. J. He, «Deep Residual Learning for Image Recognition,» de *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [32] S. F. N. & H. G. E. Sabour, «Dynamic Routing Between Capsules,» de *Advances in Neural Information Processing Systems 30*, 2017.
- [33] J. C. M. W. L. K. & T. K. Devlin, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv, 2018.

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE