Variedades estadísticas y funciones de pre-contraste: una revisión sistemática según las directrices PRISMA

Marilyn Argenis Vallejo Castillo

Asesor

Isaac Esteban Camargo Freile

Universidad Nacional Abierta y a Distancia – UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería (ECBTI)

Especialización en Ciencia de Datos y Analítica

2025

Resumen

En la actualidad, la ciencia de datos y el machine learning se han integrado en diversas disciplinas del conocimiento, gracias a las características innovadoras y los beneficios significativos que ofrecen. Estas tecnologías permiten abordar problemas complejos con mayor precisión y eficiencia mediante el uso de algoritmos avanzados. Dichos algoritmos son fundamentales para procesos críticos como la limpieza y el procesamiento de grandes volúmenes de datos, los cuales son esenciales para obtener información valiosa y tomar decisiones informadas. Sin embargo, el procesamiento de datos a gran escala requiere una cantidad considerable de recursos computacionales, lo que ha generado una necesidad de optimización de algoritmos para mejorar la eficiencia y reducir los costos.

En este contexto, la geometría diferencial emerge como un campo matemático de gran relevancia para la ciencia de datos y el machine learning. La geometría diferencial ofrece herramientas teóricas que facilitan la optimización de algoritmos a través de la implementación de variedades estadísticas específicas. Estas variedades permiten modelar y analizar de manera efectiva los espacios de parámetros y las estructuras de los datos, lo que contribuye a mejorar el rendimiento de los algoritmos en términos de velocidad y precisión. La utilización de estas herramientas no solo optimiza los procesos, sino que también abre nuevas posibilidades en el desarrollo de algoritmos más robustos.

La presente monografía tiene como objetivo principal realizar una revisión sistemática de la literatura existente sobre la aplicación de variedades estadísticas y funciones de precontraste en la optimización de algoritmos dentro del ámbito del machine learning y la ciencia de datos. Para ello, se seguirán las directrices establecidas por Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), un marco metodológico ampliamente reconocido en la

investigación científica que garantiza una revisión exhaustiva y transparente de la literatura, se analizarán trabajos que aborden tanto el desarrollo teórico como las aplicaciones prácticas de estas herramientas matemáticas en diversos contextos de la ciencia de datos. El proceso de revisión incluirá la selección de estudios relevantes, la extracción de datos clave y la síntesis de los hallazgos para proporcionar una visión comprensiva del estado actual del conocimiento en este campo.

Un aspecto importante de esta revisión será la evaluación de cómo las funciones de precontraste, un concepto fundamental en el análisis estadístico y la geometría diferencial, contribuyen a la mejora de algoritmos. Las funciones de precontraste permiten identificar estructuras subyacentes en los datos y optimizar el proceso de aprendizaje, lo que resulta en algoritmos más efectivos y eficientes. A través de esta monografía, se pretende no solo resumir los avances actuales, sino también identificar áreas donde se requiere más investigación, ofreciendo así una base sólida para futuras exploraciones en el área de la geometría diferencial y la ciencia de datos.

En resumen, la monografía proporcionará una recopilación detallada y sistemática de los resultados obtenidos de estudios anteriores, enfocándose en cómo las variedades estadísticas y las funciones de precontraste pueden ser aprovechadas para optimizar algoritmos en el campo del machine learning. Este enfoque permitirá una comprensión más profunda de las técnicas actuales y destacará las oportunidades para innovaciones futuras en la optimización de algoritmos.

***Palabras clave*:** Ciencia de datos, Machine learning, Geometría diferencial, Variedades estadísticas, Funciones de pre-contraste.

Abstract

Nowadays, data science and machine learning have been integrated into various fields of knowledge, thanks to their innovative features and significant benefits. These technologies allow for solving complex problems with greater precision and efficiency through the use of advanced algorithms. These algorithms are fundamental to critical processes such as cleaning and processing large volumes of data, which are essential for obtaining valuable insights and making informed decisions. However, large-scale data processing requires substantial computational resources, leading to the need for algorithm optimization to improve efficiency and reduce costs.

In this context, differential geometry emerges as a highly relevant mathematical field for data science and machine learning. Differential geometry provides theoretical tools that facilitate algorithm optimization through the implementation of specific statistical manifolds. These manifolds effectively model and analyze parameter spaces and data structures, contributing to improved algorithm performance in terms of speed and accuracy. The use of these tools not only optimizes processes but also opens new possibilities for developing more robust algorithms.

This monograph aims to conduct a systematic review of the existing literature on the application of statistical manifolds and precontrast functions in the optimization of algorithms within the field of machine learning and data science. To this end, it will follow the guidelines established by Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), a widely recognized methodological framework in scientific research that ensures a comprehensive and transparent literature review. Both theoretical developments and practical applications of these mathematical tools in various contexts of data science will be analyzed. The review process will include the selection of relevant studies, the extraction of key data, and the

synthesis of findings to provide a comprehensive overview of the current state of knowledge in this field.

An important aspect of this review will be evaluating how precontrast functions, a fundamental concept in statistical analysis and differential geometry, contribute to improving algorithms. Precontrast functions help identify underlying structures in data and optimize the learning process, resulting in more effective and efficient algorithms. This monograph aims not only to summarize current advancements but also to identify areas where further research is needed, thus offering a solid foundation for future explorations in differential geometry and data science.

In summary, the monograph will provide a detailed and systematic compilation of results from previous studies, focusing on how statistical manifolds and precontrast functions can be leveraged to optimize algorithms in the field of machine learning. This approach will enable a deeper understanding of current techniques and highlight opportunities for future innovations in algorithm optimization.

Keywords: Data science, Machine learning, Differential geometry, Statistical manifolds, Precontrast functions.

Tabla de Contenido

Introducción 9

Descripción del Problema 10

Planteamiento del Problema 10

Justificación 12

Objetivos 13

Objetivo General 13

Objetivos Específicos 13

Marco de Referencial 14

Marco Conceptual 14

Marco Teórico 14

Resultados 20

Tópicos más Usados 20

Descripción de los Artículos Incluidos 21

Gráficos y Visualizaciones 23

Discusión 25

Análisis de Tendencias y Lagunas 27

Impacto en el Campo de Machine Learning 27

Conclusiones 29

Recomendaciones 30

Referencias Bibliográficas 31

Lista de Tablas

Tabla 1 *Palabras más Frecuentes en los Artículos*.. 20

Tabla 2 *Artículos Seleccionados*.. 21

Tabla 3 *Conceptos Clave* 25

Lista de Figuras

Figura 1 *Diagrama de Flujo PRISMA* 17

Figura 2 *Nube de Palabras más Frecuentes* 20

Figura 3 *Artículos Publicados por Año* 23

Figura 4 *Distribución de Artículos Publicados por País* 24

Figura 5 *Distribución de Artículos Teóricos y Computaciones* 24

Introducción

En el campo de la ciencia de datos y el machine learning, la optimización de algoritmos es fundamental para mejorar la eficiencia y precisión en la resolución de problemas complejos. Un enfoque reciente en este campo es el uso de conceptos matemáticos avanzados como las variedades estadísticas y la geometría diferencial, herramientas matemáticas que permiten modelar y analizar estructuras de datos de alta complejidad. Este enfoque geométrico ofrece una nueva manera de comprender las características matemáticas y estructurales de los modelos, lo que facilita la identificación de nuevas oportunidades para optimizar los algoritmos en áreas como el aprendizaje automático y el aprendizaje profundo.

Mediante una revisión sistemática basada en las directrices PRISMA, este estudio explorará las aplicaciones variedades estadísticas, métricas riemannianas y otras métricas relevantes en la optimización de algoritmos, lo que se busca es tener un panorama claro de los casos en los que el uso de variedades estadísticas ha logrado avances importantes en la eficiencia computacional orientada a la integración y optimización en algoritmos de machine learning y en determinado caso, establecer posibles temas de investigación.

Descripción del Problema

Planteamiento del Problema

El problema específico identificado radica en el conocimiento limitado sobre la aplicación de las variedades estadísticas y las funciones de pre-contraste en el campo del machine learning. Esta falta de comprensión y aplicación puede limitar el desarrollo y la optimización de algoritmos, ya que los métodos estadísticos clásicos a menudo resultan ineficientes frente a grandes volúmenes de datos y procesos complejos (Sarmashghi et al., 2022).

Hasta la fecha, se han realizado estudios sobre técnicas de machine learning y estadísticas de manera independiente, pero la integración específica de variedades estadísticas y funciones de pre-contraste en este contexto aún no ha sido explorada en profundidad. Esta laguna en la investigación es crítica, ya que el machine learning se basa en gran medida en la estadística computacional, cuyo núcleo es la aplicación de la estadística clásica en diversas áreas. Ejemplos de esto incluyen la detección de fallas en motores mediante indicadores estadísticos (García-Basurto et al., 2021b), el análisis de relaciones genómicas en el ADN mediante estadística multivariable (Reyes-Alemán et al., 2013), y la creación de modelos de riesgo financiero (Peña, 1993).

Dado este contexto, surge la necesidad de examinar más a fondo el impacto y el papel de las variedades estadísticas y las funciones de pre-contraste en el desarrollo actual del machine learning. La monografía propuesta buscará abordar esta laguna mediante una exploración detallada y sistemática de estas herramientas, con el objetivo de mejorar la eficiencia y precisión de los algoritmos de machine learning.

Pregunta de Investigación:

¿Cuál es el impacto y el papel de las variedades estadísticas y funciones de pre-contraste en la aplicación de algoritmos de machine learning?

Justificación

El problema que se aborda en esta investigación es la aplicación limitada de variedades estadísticas y funciones de pre-contraste en el contexto del machine learning, lo cual impacta de forma negativa en la eficiencia y precisión de los algoritmos. Un ejemplo de esto es cuando se modela un problema semántico, dada la complejidad es natural la lentitud e ineficiencia del algoritmo (Pasquinelli et al., 2022). Es importante investigar este problema debido a que por el momento, solo se han utilizado algunas variedades estadísticas lineales en algoritmos de machine learning y la implementación de algoritmos de geometría diferencial genera secuencias de código extensas y complejas (Kühnel & Sommer, 2017).

La literatura existente revela importantes lagunas en relación a este tema. A pesar de los avances en el campo del machine learning, la integración específica de variedades estadísticas y funciones de pre-contraste no ha sido muy explorada, incluso la optimización geométrica no se ha aplicado a todos los métodos de aprendizaje profundo (Fei et al., 2023).

Esta investigación busca llenar estas lagunas y contribuir al campo de estudio proporcionando más información sobre la aplicación de variedades estadísticas y funciones de pre-contraste en machine learning. Se espera que los hallazgos tengan un impacto positivo.

La importancia de esta investigación radica en el avance constante de la tecnología, y en los aportes significativos recientes en el área de machine learning, que han tenido un impacto en diversas áreas del conocimiento (Forero-Corba & Bennásar, 2023). Esto ha generado una alta demanda por nuevas aplicaciones y conocimientos en este campo. Los beneficiarios de esta investigación incluyen a investigadores y académicos interesados en ampliar su comprensión del machine learning, así como a cualquier persona interesada en explorar las potencialidades y aplicaciones de esta tecnología.

Objetivos

Objetivo General

Analizar el impacto y el papel de las variedades estadísticas y funciones de pre-contraste en la aplicación de algoritmos de machine learning.

Objetivos Específicos

Indagar en la literatura existente el uso de variedades estadísticas en modelos de aprendizaje automático.

Recopilar información acerca del uso de las variedades estadísticas y funciones de pre-contraste en modelos de machine learning aplicados.

Identificar tendencias, patrones, lagunas en la investigación y áreas de interés que puedan orientar investigaciones futuras.

Marco de Referencial

Marco Conceptual

Machine Learning (Aprendizaje Automático): "El aprendizaje automático se define como el campo que estudia los algoritmos computacionales que mejoran a partir de la experiencia" (Mitchell, 1997, p. 4), agrupa al conjunto de algoritmos que son capaces de aprender a partir de los datos (Ruiz & Velásquez, 2023).

Variedades Estadísticas: En el contexto del análisis de datos, las variedades estadísticas se refieren a un tipo especial de variedad de Riemann (Henmi & Matsuzoe, 2018), la que incluye, métodos y técnicas para modelar y analizar datos estadísticos, por medio de estructuras geométricas.

Funciones de Pre-Contraste: Son funciones utilizadas para preparar los datos antes de aplicar la función de contraste, encargada de generar la estructura geométrica conocida como variedad (Henmi & Matsuzoe, 2018).

Geometría de la Información: La geometría de la información es un método para explorar el mundo de la información mediante la geometría. Por el momento, se han utilizado métodos algebraicos, lógicos, analíticos y probabilísticos; desde la geometría, se puede estudiar relaciones entre elementos como distancia y curvatura, por lo tanto, otorga a las ciencias de la información nuevas herramientas y métodos (Amari, 2016).

Marco Teórico

El método Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) es un protocolo utilizado en la investigación científica que se utiliza para realizar revisiones sistemáticas y meta-análisis de estudios primarios. El método PRISMA se desarrolló en 2009 y proporciona pautas detalladas para la identificación, selección, evaluación y síntesis de evidencia científica de manera rigurosa y transparente. Este enfoque sistemático garantiza la reproducibilidad y la calidad de los resultados obtenidos en la investigación, lo que lo convierte en una herramienta fundamental en la síntesis de la evidencia científica actual, como lo indica Page et al. (2021)

El método PRISMA consta de varias etapas bien definidas que guían el proceso de revisión sistemática:

Identificación: En esta etapa, se lleva a cabo una búsqueda exhaustiva de la literatura científica relevante en múltiples bases de datos y otros recursos pertinentes. Se utilizan términos de búsqueda específicos y se aplican criterios de inclusión y exclusión para seleccionar los estudios pertinentes.

Selección: Los estudios identificados se someten a un proceso de selección riguroso, en el cual se aplican criterios predefinidos para determinar su relevancia y calidad. Este proceso implica la revisión independiente de los títulos y resúmenes de los estudios, seguida de la evaluación completa de los textos completos de aquellos que cumplen con los criterios de inclusión.

Evaluación de la calidad y el riesgo de sesgo: En esta etapa, se evalúa la calidad metodológica de los estudios incluidos y se identifican posibles fuentes de sesgo. Se utilizan herramientas estandarizadas, como la escala de Newcastle-Ottawa para estudios observacionales, para evaluar la calidad de los estudios y la herramienta ROBINS-I para estudios de intervención.

Extracción de datos: Se extraen datos relevantes de los estudios incluidos de manera sistemática y estructurada. Esto puede incluir información sobre las características de los participantes, los métodos utilizados, los resultados principales y cualquier otro dato relevante para los objetivos de la revisión.

Síntesis de los resultados: Los datos extraídos se analizan y se presentan de manera sistemática, utilizando métodos estadísticos si es apropiado. Se pueden realizar meta-análisis para combinar los resultados de estudios similares y estimar medidas de efecto globales.

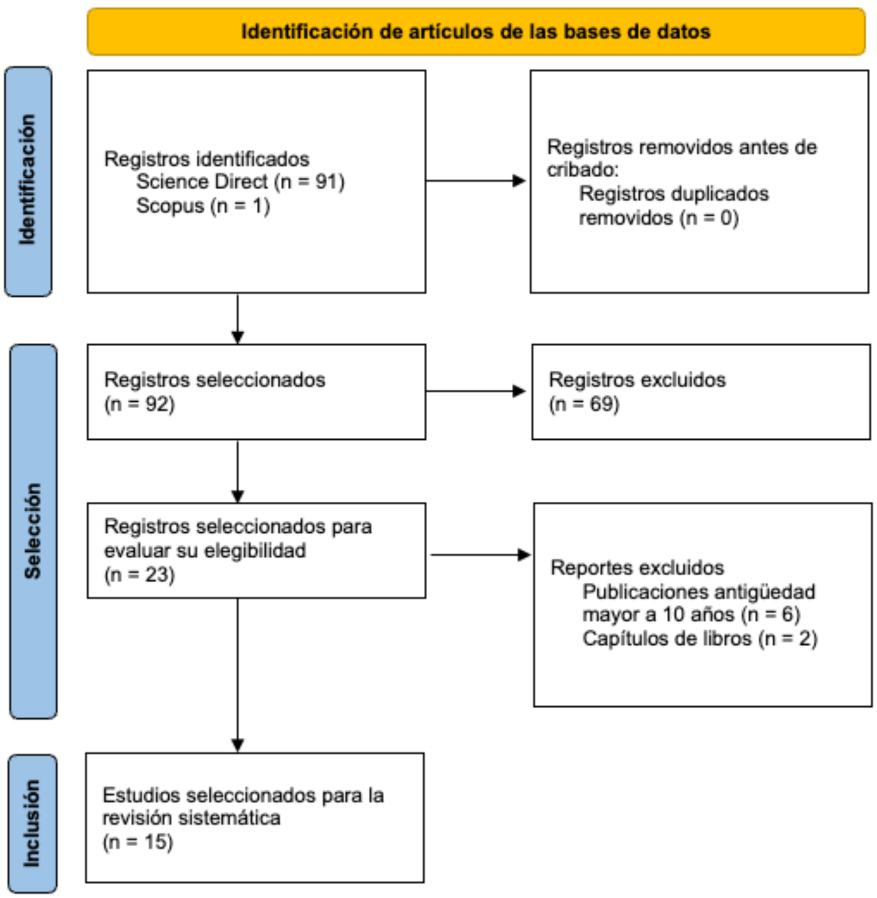
Interpretación y presentación: Finalmente, los resultados de la revisión se sintetizan por medio de una caracterización, la cual permite sintetizar, interpretar y presentar los resultados.

Metodología

Esta revisión se llevó a cabo siguiendo las directrices del modelo PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), con el fin de garantizar la exhaustividad, transparencia y consistencia en el análisis de la literatura sobre variedades estadísticas y funciones de pre-contraste en machine learning.

Figura 1

Diagrama de Flujo PRISMA



Para este propósito, se siguieron las siguientes etapas:

Fuentes de búsqueda: Se consultaron las bases de datos Science Direct y Scopus.

Estrategias de búsqueda: Para construir la cadena de búsqueda, se elaboró un listado con los términos clave. Posteriormente, se definieron los términos relevantes que representan la población de interés en la revisión, los cuales se separan utilizando el operador OR, y las restricciones, que se separan por el operador AND. La cadena resultante fue: contrast function OR precontrast function OR pre-contrast function AND statistical manifold AND riemannian AND hessian AND connection.

Criterios de inclusión y exclusión: Se definieron criterios específicos para la selección de los artículos incluidos en la revisión. Como criterios de inclusión, se consideraron aquellos artículos que estudien el uso de variedades estadísticas y métricas riemannianas, especialmente en la optimización de algoritmos de machine learning. También se incluyeron artículos publicados en inglés.

Por otro lado, se excluyeron los estudios que no se enfocaran en aplicaciones de las variedades estadísticas, además, se descartaron artículos con más de 10 años de antigüedad, para garantizar la actualidad de los resultados.

Proceso de selección: Después de aplicar la cadena de búsqueda en la consulta de las bases de datos, se identificaron 91 registros en Science Direct y 1 en Scopus, se revisaron las referencias en busca de duplicados y no se encontró ninguno, posteriormente se realizó la lectura rápida, es decir, se tomó en cuenta solo títulos y abstract con lo cual se excluyeron 69 ítems quedando 23 registros seleccionados, de estos, se eliminaron aquellos que cumplían los criterios de exclusión: 6 con antigüedad mayor a 10 años y 2 que pertenecen a capítulos de libro; finalmente se obtuvieron 15 artículos para revisión.

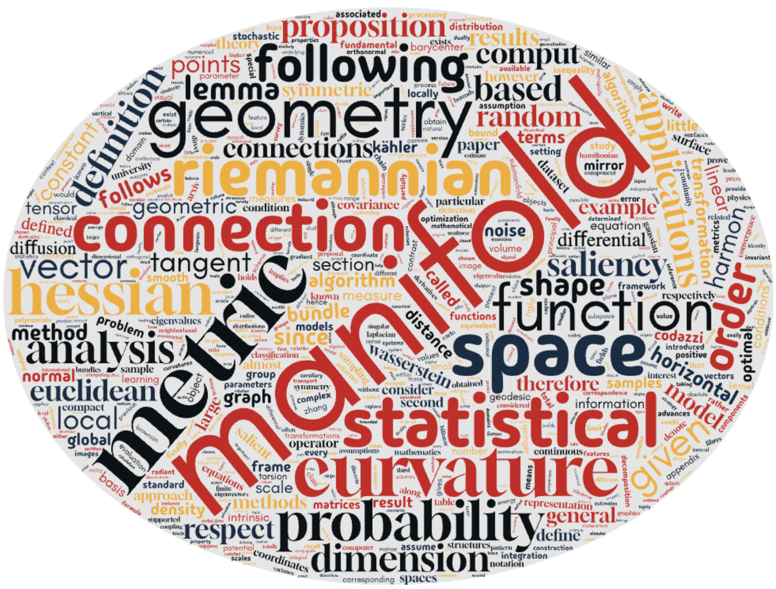
Extracción de datos: Con el fin de obtener la caracterización de los artículos, se realizó la lectura de los artículos con especial atención a la métrica utilizada para el cálculo de distancias, si se trata de un ejemplo aplicado o no, y en caso de tener cálculos computacionales, el lenguaje o herramienta utilizada, los datos se registraron de forma individual en la tabla 2.

Resultados

Tópicos más Usados

Figura 2

Nube de Palabras más Frecuentes



En la nube de palabras se puede apreciar que los artículos mencionan las variedades estadísticas y la métrica riemanniana como temas principales, de igual manera las palabras: métrica, teorema, curvatura y geometría en menor medida, sin embargo, el término machine learning no está presente lo que representa una baja frecuencia.

Para realizar una descripción cuantitativa, en la tabla 1 se presenta el top 10 de palabras más usadas en los artículos.

Tabla 1

Palabras más Frecuentes en los Artículos

Número	Palabra	Frecuencia
1	manifold	1045
2	metric	369

Número	Palabra	Frecuencia
3	riemannian	368
4	function	347
5	geometry	327
6	space	320
7	curvature	292
8	statistical	290
9	connection	277
10	hessian	241

Nota. Lista de palabras más frecuentes en los artículos

Descripción de los Artículos Incluidos

Con la revisión realizada, se identificaron 15 artículos en los cuales se utilizan las variedades estadísticas para el modelado de problemas y la solución mediante cálculos computacionales y/o teóricos. Se evidencia que en su gran mayoría, los cálculos realizados son teóricos debido a la complejidad de dimensiones utilizadas, por otra parte la métrica que más se usa es la riemanniana, debido a la característica intrinseca con respecto a las curvas, finalmente se encontró que ha sido aplicado en reconocimiento de imágenes, estadística multiescala, pero solo un artículo aborda un algoritmo de clasificación el cual entra en el ámbito de machine learning.

Tabla 2

Artículos Seleccionados

No	Artículo	Año	País	Cálculo Computacional	Cálculo Teórico	Métrica	Ejemplo aplicado	Lenguaje de programación
1	Curvature of Hessian manifolds	2014	Japón		X	Hessian, riemanniana	No	No
2	Discriminative subspace learning via optimization on Riemannian manifold	2023	China	X		Riemanniana	Reconocimiento de imágenes	Matlab
3	Gradient-based MCMC samplers	2015	Londres		X	Riemanniana	Neuroimagen	No

No	Artículo	Año	País	Cálculo Computacional	Cálculo Teórico	Métrica	Ejemplo aplicado	Lenguaje de programación
	for dynamic causal modelling							
4	Local eigenvalue decomposition for embedded Riemannian manifolds	2020	Alemania		X	Riemanniana	No	No
5	Locally conformally Hessian and statistical manifolds	2023	Rusia		X	Hessian, riemanniana	No	No
6	Multiscale geometric methods for data sets I_Multiscale SVD, noise, and curvature	2017	Estados Unidos	X		Hessian, riemanniana	- Redes multiescala ('M-Net') - Estadísticas multiescala ('M-Stats')	Matlab
7	Riemannian Gaussian distributions, random matrix ensembles and diffusion kernels	2021	Portugal		X	Riemanniana, RaoFisher	No	No
8	Rigid motion invariant statistical shape modeling based on discrete fundamental forms	2021	Alemania		X	Riemanniana	Algoritmos de clasificación	
9	Statistical manifold with degenerate metric	2023	China		X	Pseudo-riemanniana	No	No
10	Shape context based mesh saliency detection and its applications_A survey	2016	China		X	Riemanniana	No	No
11	Statistical mirror symmetry	2020	Estados Unidos		X	Pseudo-riemanniana	No	No
12	Stochastic dynamical systems developed on Riemannian manifolds	2022	Londres		X	Riemanniana	No	No
13	The diffusion geometry of fibre bundles_Horizontal diffusion maps	2019	Estados Unidos		X	Riemanniana	Si	No
14	Transformations and coupling relations for affine connections	2016	Estados Unidos		X	Riemanniana	No	No

No	Artículo	Año	País	Cálculo Computacional	Cálculo Teórico	Métrica	Ejemplo aplicado	Lenguaje de programación
15	Wasserstein barycenters over Riemannian manifolds	2016	Canadá		X	Métrica de Wasserstein	No	No

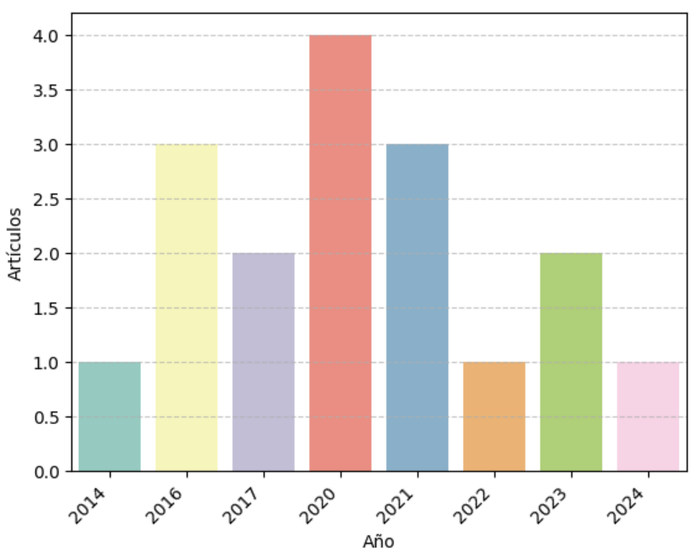
Nota. Lista de artículos seleccionados para el análisis.

Gráficos y Visualizaciones

Artículos realizados por año: A partir de la caracterización realizada, se presenta la distribución de los artículos teniendo en cuenta el año de publicación, aunque la publicación dentro de estas líneas de investigación ha sido constante, solo en el año 2021 se encuentra un artículo que combina machine learning con geometría diferencial. La distribución se presenta en la figura 3.

Figura 3

Artículos Publicados por Año

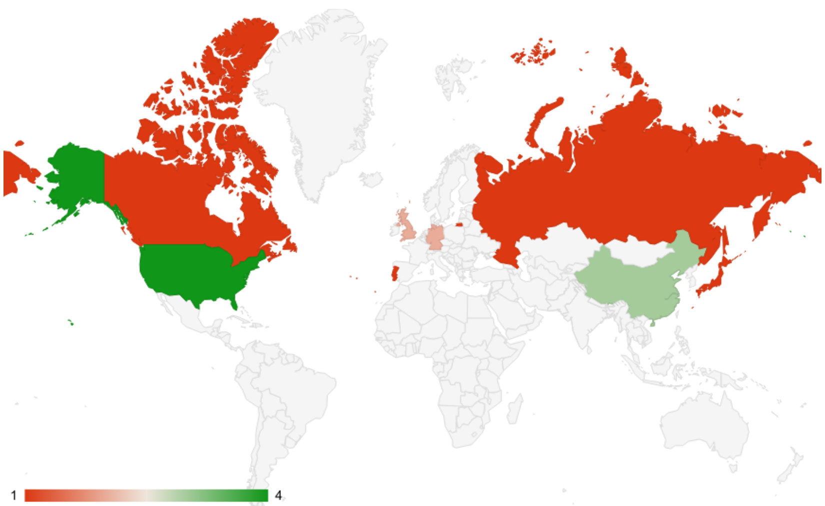


Distribución de Artículos por Ubicación Geográfica: La Figura 4, muestra la distribución de las publicaciones, demostrando que los países con más publicaciones son: Estados Unidos con

4 artículos, seguido por China con 3 artículos. Los países como: Japón, Rusia, Canadá y Portugal tienen la menor cantidad de publicaciones, con 1 artículo cada uno. Es necesario recalcar que Latinoamérica no presenta publicaciones de esta índole.

Figura 4

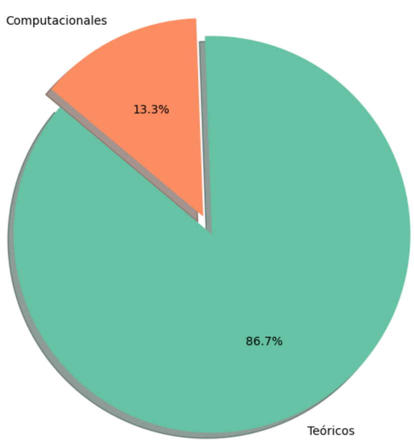
Distribución de Artículos Publicados por País



Tipos de Investigación Realizada: Se encontró un 86,76% de artículos que corresponden a estudios teóricos, mientras que el 13,33% corresponde a cálculos computacionales.

Figura 5

Distribución de Artículos Teóricos y Computaciones



Discusión

Tabla 3

Conceptos Clave

Número	Concepto
1	La geometría diferencial es la maquinaria matemática para realizar cálculos sobre una hipersuperficie de forma arbitraria en cualquier dimensión, digamos R, y puede verse como una generalización útil del cálculo estándar en el contexto euclidiano. La desviación de la configuración euclidiana se captura específicamente a través de ciertos tensores de incompatibilidad, por ejemplo, el tensor de curvatura en la geometría de Riemann. (Mamajiwala y Roy, 2022)
2	La estructura estadística de una variedad M se basa en un tipo especial de acoplamiento entre la métrica riemanniana g y una conexión afín libre de torsión en TM, de modo que g es totalmente simétrica, formando por definición un par de Codazzi. (Tao y Zhang, 2016)
3	Una métrica riemanniana g se denomina métrica hessiana si existen coordenadas locales tales que g se pueda escribir como la hessiana de alguna función potencial convexa. (Amari y Armstrong, 2014)
4	Una variedad estadística (MDg) es una variedad M dotada de una conexión libre de torsión D y una métrica riemanniana g tal que el tensor Dg es totalmente simétrico. Si D es plano, entonces M,g,D es una variedad hessiana. (Osipov, 2023)
5	El análisis de componentes principales (o descomposición en valores singulares) es una de las herramientas más básicas y, sin embargo, más utilizadas en estadística y análisis de datos. (Little et al. 2017)
6	En el mundo físico, una variedad riemanniana ofrece un marco ideal para estudiar evoluciones sobre o de objetos generalmente curvos. A diferencia del espacio euclidiano, dos espacios tangentes a dos puntos cualesquiera en una variedad riemanniana curva no son canónicamente isomorfos. (Mamajiwala y Roy, 2022)

Nota. Esta tabla muestra los conceptos clave presentados en algunos artículos.

Se extrajeron los conceptos relevantes presentados en algunos artículos, lo que permitió elaborar la tabla 3: Conceptos clave. Se observa que varios autores coinciden en definir las variedades estadísticas como una rama de la geometría diferencial, destacando su capacidad para modelar superficies multivariables respetando la curvatura que los datos representan. A partir de esta perspectiva, se presentan diferentes herramientas que permiten encontrar la distancia entre 2 puntos, similar a la distancia euclidiana pero teniendo en cuenta la curvatura, la más conocida es la métrica riemanniana.

Los modelos basados en variedades estadísticas, las métricas riemannianas, hessianas y demás, han sido usadas en diferentes áreas, tales como:

El análisis armónico, la teoría de números, el análisis matricial, la geometría de la información y la estadística multivariable (Santilli y Tierz, 2021)

La optimización y teoría de cuerdas (Amari y Armstrong, 2014).

El análisis de imágenes médicas con una amplia gama de aplicaciones, como segmentación de estructuras anatómicas, diagnóstico asistido por computadora y planificación de terapias (Ambellan et al., 2021).

La saliencia de malla y gráficos por computadora reducción de malla, segmentación de malla, coincidencia de autosimilitud, integración de escaneo, renderizado de volumen, impresión 3D (Liu et al., 2016)

DSL (Aprendizaje subespacial discriminativo) aplicado a visión por computadora, procesamiento de señales e ingeniería biométrica (Yin et al., 2023).

La aplicación en diferentes áreas del conocimiento, se explica debido a la gran versatilidad que tiene un modelo geométrico en la representación de datos de multivariables, así mismo, estos modelos abordan problemas complejos, como los siguientes:

Algoritmos de clasificación: Los algoritmos de aprendizaje generalmente aprenden a discriminar diferentes clases calculando la distancia o similitud entre las características extraídas de los datos de entrada (Yin et al., 2023), la geometría de la información proporciona una caracterización geométrica de familias paramétricas de distribuciones de probabilidad o modelos estadísticos (Zhang y Khan, 2020), los SSM pueden aprender de una base de datos de instancias parametrizadas consistentemente de la clase de objeto bajo estudio (Ambellan et al., 2021).

Disminución de costo computacional: Cuando hay muchos parámetros, el muestreo a partir de la densidad posterior es computacionalmente intratable (Sengupta et al., 2016), se han desarrollado enfoques no lineales basados en conceptos geométricos y físicos, como la distancia de Hausdorff, elasticidad y flujos viscosos. En general, estos métodos carecen de robustez numérica y de tasas de respuestas rápidas. (Ambellan et al., 2021).

Métricas imprecisas: Las variedades, permiten el análisis de datos que no se encuentran en espacios euclidianos ordinarios (Santilli y Tierz, 2021), En un entorno riemanniano consistente, permiten construir un marco que maneja de forma fiable grandes deformaciones (Ambellan et al., 2021), los métodos lineales suelen ser inadecuados para capturar la alta variabilidad en las formas biológicas (Ambellan et al., 2021).

Análisis de Tendencias y Lagunas

Se observó una tendencia alta hacia el uso de variedades riemannianas para la representación de datos en espacios con propiedades geométricas curvas. Sin embargo, se identificaron lagunas en la integración de funciones de pre-contraste en modelos de machine learning, lo que resalta la necesidad de realizar investigaciones más profundas en este ámbito.

Otra línea interesante de trabajo futuro es explorar la viabilidad de diagnósticos asistidos por computadora totalmente automáticos basados en aprendizaje automático avanzado (Ambellan et al., 2021).

Impacto en el Campo de Machine Learning

La adquisición de conjuntos de datos complejos, masivos y, a menudo, de alta dimensión, se ha convertido en una práctica común en muchos campos de la ciencia. Si bien estos conjuntos de datos son inspiradores y estimulantes, pueden resultar difíciles de analizar o comprender de manera eficiente (Gao, 2021), generalmente se trabajan desde machine learning, se sabe que los modelos resultantes proporcionan una forma previa que se puede utilizar para limitar los problemas de síntesis y análisis (Ambellan et al., 2021), si bien es cierto que los métodos tradicionales en su mayoría están basados en espacio euclidiano (Yin et al., 2023) y que el rendimiento de estos algoritmos depende en gran medida de la estrategia para buscar el subespacio discriminativo en el espacio euclidiano, y que la mayoría de los algoritmos supervisados existentes para el aprendizaje del subespacio son lineales. (Yin et al., 2023), las variedades superan estos inconvenientes, puesto que de forma intrínseca, su espacio de parámetros proporciona una representación compacta que es adecuada para algoritmos de aprendizaje (clasificación o agrupamiento) (Ambellan et al., 2021).

Conclusiones

Hasta donde la exploración lo permitió, se evidencia una limitada aplicación de variedades estadísticas y funciones de pre-contraste en el ámbito de machine learning, lo cual representa una oportunidad de estudio enfocada en la integración de métricas riemannianas en herramientas de machine learning.

La distribución geográfica de publicaciones nos permite identificar que el campo de variedades estadísticas aplicado en machine learning, no ha sido explorado ampliamente y mucho menos, existen publicaciones latinoamericanas, lo que significa una oportunidad para futuras investigaciones.

La revisión nos indica que la tendencia más frecuente en el campo de variedades estadísticas, es el uso de métricas riemannianas y/o propuestas de métricas para optimizar y mejorar las métricas en curvaturas, esta preferencia se explica, debido a la precisión y robustez en comparación a otro tipo de mediciones.

Recomendaciones

Optimizar las revisiones sistemáticas con la ayuda de herramientas de automatización como systematic-reviewpy, que es un proyecto orientado a revisiones sistemáticas usando la metodología PRISMA.

Desde la universidad, se podría impulsar la creación de grupos de investigación multidisciplinarios para integrar distintas áreas del conocimiento. Esto permitiría identificar aplicaciones innovadoras y optimizaciones en diversos campos.

Dada la relevancia y potencial que presentan las variedades estadísticas en modelos de machine learning, se sugiere profundizar este tema en una futura investigación.

Referencias Bibliográficas

Amari, S. (2016). *Information Geometry and its Applications*. Springer

Amari, S., & Armstrong, J. (2014). *Curvature of hessian manifolds*. Differential Geometry and its Applications, 33, 1–12. https://doi.org/https://doi.org/10.1016/j.difgeo.2014.01.001

Ambellan, F., Zachow, S., & von Tycowicz, C. (2021). *Rigid motion invariant statistical shape modeling based on discrete fundamental forms: Data from the osteoarthritis initiative and the alzheimer's disease neuroimaging initiative.* Medical Image Analysis, 73, 102178. https://doi.org/https://doi.org/10.1016/j.media.2021.102178

ECBTI (2016). *Líneas de Investigación ECBTI Universidad Nacional Abierta y a Distancia (UNAD)*. Universidad Nacional Abierta y a Distancia (UNAD). https://academia.unad.edu.co/investigacion-y-productividad-ecbti/lineas

Fei, Y., Wang, X., Liu, Y., Li, Z., & Chen, M. (2023). *A Survey of Geometric Optimization for Deep Learning: From Euclidean Space to Riemannian Manifold*. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2302.08210

Forero-Corba, W., & Bennásar, F. N. (2023). *Técnicas y aplicaciones del Machine Learning e Inteligencia Artificial en educación: una revisión sistemática*. RIED: Revista Iberoamericana de Educación A Distancia, 27(1), 209-253. https://doi.org/10.5944/ried.27.1.37491

Gao, T. (2021). *The diffusion geometry of fibre bundles: Horizontal diffusion maps*. Applied and Computational Harmonic Analysis, 50, 147–215. https://doi.org/https://doi.org/10.1016/j.acha.2019.08.001

García-Basurto, A., Saucedo-Dorantes, J. J., Pérez-Cruz, Á., & Osornio-Ríos, R. A. (2021b). *Análisis de falla de encendido en motores de combustión utilizando señales de vibración* *basado en el cálculo y reducción de indicadores estadísticos*. Científica, 25(1), 83-95. https://doi.org/10.46842/ipn.cien.v25n1a07

Henmi, M., & Matsuzoe, H. (2018). *Statistical Manifolds Admitting Torsion and Partially Flat Spaces*. En Signals and communication technology (pp. 37-50). https://doi.org/10.1007/978-3-030-02520-5_3

Kühnel, L., & Sommer, S. (2017). *Computational Anatomy in Theano*. En Lecture Notes in Computer Science (pp. 164-176). https://doi.org/10.1007/978-3-319-67675-3_15

Little, A. V., Maggioni, M., & Rosasco, L. (2017). *Multiscale geometric methods for data sets i: Multiscale svd, noise and curvature*. Applied and Computational Harmonic Analysis, 43(3), 504–567. https://doi.org/https://doi.org/10.1016/j.acha.2015.09.009

Liu, X., Liu, L., Song, W., Liu, Y., & Ma, L. (2016). *Shape context based mesh saliency detection and its applications: A survey*. Computers Graphics, 57, 12–30. https://doi.org/https://doi.org/10.1016/j.cag.2016.03.001

Mamajiwala, M., & Roy, D. (2022). *Stochastic dynamical systems developed on Riemannian manifolds.* Probabilistic Engineering Mechanics, 67, 103179. https://doi.org/https://doi.org/10.1016/j.probengmech.2021.103179

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill Science/Engineering/Math

Osipov, P. (2023). *Locally conformally hessian and statistical manifolds*. Journal of Geometry and Physics, 193, 104989. https://doi.org/https://doi.org/10.1016/j.geomphys.2023.104989

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., Alonso-Fernández, S. (2021). *Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas*. Revista Española de Cardiología/Revista Española de Cardiología, 74(9), 790-799. https://doi.org/10.1016/j.recesp.2021.06.016

Pasquinelli, M., Cafassi, E., Monti, C., Peckaitis, H., & Zarauza, G. (2022). *Cómo una máquina aprende y falla – una gramática del error para la inteligencia artificial*. Hipertextos, 10(17), 13-29. https://doi.org/10.24215/23143924e054

Peña, J. (1993). *Medidas de volatilidad en mercados financieros*. Revista española de financiación y contabilidad. 77. 937-948. https://www.researchgate.net/publication/28216735_Medidas_de_volatilidad_en_mercados_financieros

Reyes-Alemán, J. C., Valadez-Moctezuma, E., Simuta-Velázco, L., Barrientos-Priego, A. F., & Gallegos-Vázquez, C. (2013). *Distinción de especies del género Persea mediante RAPD e ISSR de ADN 1*. Revista Mexicana de Ciencias Agrícolas 4(4), 517-529. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2007-09342013000400003&lng=es&tlng=es

Ruiz & Velásquez, 2023 = Ruiz, R. B., & Velásquez, J. D. (2023). *Inteligencia artificial al servicio de la salud del futuro*. Revista MéDica ClíNica las Condes, 34(1), 84-91. https://doi.org/10.1016/j.rmclc.2022.12.001

Santilli, L., & Tierz, M. (2021). *Riemannian gaussian distributions, random matrix ensembles and diffusion kernels*. Nuclear Physics B, 973, 115582. https://doi.org/https://doi.org/10.1016/j.nuclphysb.2021.115582

Sarmashghi, M., Jadhav, S. P., & Eden, U. T. (2022). *Integrating Statistical and Machine Learning Approaches for Neural Classification*. IEEE Access, 10, 119106-119118. https://doi.org/10.1109/access.2022.3221436

Sengupta, B., Friston, K. J., & Penny, W. D. (2016). *Gradient-based mcmc samplers for dynamic causal modelling*. NeuroImage, 125, 1107–1118. https://doi.org/https://doi.org/10.1016/j.neuroimage.2015.07.043

Tao, J., & Zhang, J. (2016). Tr*ansformations and coupling relations for affine connections*. Differential Geometry and its Applications, 49, 111–130. https://doi.org/https://doi.org/10.1016/j.difgeo.2016.06.006

Yin, W., Ma, Z., & Liu, Q. (2023). *Discriminative subspace learning via optimization on riemannian manifold*. Pattern Recognition, 139, 109450. https://doi.org/10.1016/j.patcog.2023.109450

Zhang, J., & Khan, G. (2020). *Statistical mirror symmetry*. Differential Geometry and its Applications, 73, 101678. https://doi.org/https://doi.org/10.1016/j.difgeo.2020.101678