

Estabilidad dinámica y selección de la tasa de aprendizaje en redes neuronales profundas mediante un indicador empírico

Dynamic Stability and Learning Rate Selection in Deep Neural Networks Using an Empirical Indicator

Julio Paredes Rojas



Universidad Abierta Interamericana (UAI), Argentina.

DOI <https://doi.org/10.59471/raia2026237>

Enviado: 09/03/2026. Aceptado: 01/06/2026. Publicado: octubre 2026

Como citar: Paredes Rojas, J. (2026). Estabilidad dinámica y selección de la tasa de aprendizaje en redes neuronales profundas mediante un indicador empírico, 10(1). <https://doi.org/10.59471/raia2026237>

Resumen

La selección de la tasa de aprendizaje en redes neuronales profundas suele realizarse mediante procedimientos empíricos que no siempre permiten comprender la dinámica del entrenamiento ni reducir eficientemente el espacio de búsqueda. En este trabajo se analiza dicha dinámica mediante un indicador empírico de estabilidad construido a partir de la evolución temporal de la función de pérdida, utilizado como proxy de sensibilidad del entrenamiento.

El estudio combina análisis dinámico, evaluación del desempeño predictivo y comparación entre selección tradicional y criterio dinámico. Para ello, se examina la relación entre tasa de aprendizaje, estabilidad relativa y rendimiento en redes multicapa sobre MNIST y FashionMNIST, y luego se extiende el análisis a arquitecturas de distinta profundidad, al conjunto CIFAR-10, a distintos optimizadores, a una evaluación multi-dataset con conjuntos tabulares y a una validación multi-seed.

Los resultados muestran que la tasa de aprendizaje modifica de manera sistemática la estabilidad relativa del entrenamiento y que los mejores desempeños tienden a concentrarse en regiones intermedias del indicador. En los experimentos multi-dataset, el criterio dinámico permitió reducir la región de búsqueda entre 30 % y 60 %, manteniendo desempeños finales comparables a los obtenidos mediante exploración exhaustiva. En conjunto, los hallazgos sugieren que el análisis dinámico puede complementar la selección de hiperparámetros y aportar una herramienta útil para estudiar la estabilidad del aprendizaje en redes neuronales profundas.

PALABRAS CLAVES: redes neuronales profundas; estabilidad dinámica; tasa de aprendizaje; optimización; selección de hiperparámetros..

Abstract

Learning rate selection in deep neural networks is often based on empirical procedures that do not always provide a clear understanding of training dynamics or allow the search space to be reduced efficiently. This study analyzes such dynamics through an empirical stability indicator built from the temporal evolution of the loss function, used as a proxy for training sensitivity.

The study combines dynamic analysis, predictive performance evaluation, and a comparison between traditional selection and a dynamic criterion. To this end, it examines the relationship among learning rate, relative stability, and performance in multilayer networks on MNIST and FashionMNIST, and then extends the analysis to architectures with varying depth, the CIFAR-10 dataset, different optimizers, a multi-dataset evaluation including tabular datasets, and a multi-seed validation.

The results show that the learning rate systematically modifies the relative stability of training and that the best predictive performances tend to concentrate in intermediate regions of the indicator. In the multi-dataset experiments, the dynamic criterion reduced the search region by 30% to 60% while maintaining final performances comparable to those obtained through exhaustive exploration. Overall, the findings suggest that dynamic analysis can complement hyperparameter selection and provide a useful tool for studying learning stability in deep neural networks.

KEYWORDS: deep neural networks; dynamic stability; learning rate; optimization; hyperparameter selection.

INTRODUCCIÓN

El aprendizaje profundo se ha consolidado como uno de los enfoques más influyentes de la inteligencia artificial contemporánea, con aplicaciones destacadas en visión por computadora, procesamiento del lenguaje natural y análisis de datos científicos. No obstante, este desarrollo convive con interrogantes todavía abiertos sobre la dinámica interna del entrenamiento, en particular en lo relativo a la estabilidad y sensibilidad de los algoritmos de optimización. En este contexto, diversos trabajos han señalado que el entrenamiento puede interpretarse como un sistema dinámico complejo, cuya evolución depende de la interacción entre la arquitectura del modelo, el optimizador y los hiperparámetros empleados (Bahri et al., 2020).

Entre estos hiperparámetros, la tasa de aprendizaje ocupa un lugar central, ya que regula la magnitud de las actualizaciones de los pesos y condiciona tanto la velocidad de convergencia como la regularidad de las trayectorias de entrenamiento. En la práctica, su selección suele apoyarse en procedimientos empíricos, como prueba y error o búsqueda exhaustiva. Aunque estas estrategias pueden ser efectivas, presentan limitaciones importantes: demandan un

costo computacional considerable y aportan escasa información sobre la dinámica subyacente del proceso de aprendizaje.

En los últimos años, el estudio del aprendizaje profundo desde la teoría de sistemas dinámicos ha cobrado creciente interés. Esta perspectiva permite interpretar el entrenamiento no solo como un problema de minimización, sino también como una trayectoria en el espacio de estados, sensible tanto a la geometría del paisaje de pérdida como a la configuración de los hiperparámetros (Li et al., 2018). Desde esta óptica, la estabilidad del entrenamiento adquiere interés analítico, ya que puede aportar información complementaria a la obtenida a partir de las métricas finales de desempeño. En esta línea, algunos antecedentes sugieren que ciertas configuraciones favorables pueden ubicarse en regiones intermedias entre dinámicas excesivamente rígidas y dinámicas altamente sensibles (Fort & Scherlis, 2019; Zhang et al., 2024).

A partir de esta motivación, el presente trabajo analiza la estabilidad dinámica del entrenamiento mediante un indicador empírico construido a partir de la evolución temporal de la función de pérdida, utilizado como proxy de sensibilidad relativa. El objetivo es examinar si esta información dinámica puede contribuir a identificar regiones del espacio de hiperparámetros asociadas con comportamientos favorables para el aprendizaje y, en particular, orientar la selección de la tasa de aprendizaje.

Sobre esta base, el estudio examina la relación entre tasa de aprendizaje, profundidad de red, optimizador y desempeño predictivo, e incorpora una comparación entre selección tradicional y criterio dinámico. De este modo, se busca aportar evidencia empírica sobre la utilidad del análisis dinámico como herramienta complementaria para estudiar el entrenamiento y orientar la selección de hiperparámetros.

El artículo se organiza del siguiente modo. La sección 2 presenta el marco teórico. La sección 3 formula el problema de investigación y explicita las hipótesis del estudio. La sección 4 describe la metodología experimental. La sección 5 expone los resultados. La sección 6 desarrolla la discusión de los hallazgos. Finalmente, la sección 7 presenta las conclusiones. .

Marco teórico

Entrenamiento de redes neuronales como sistema dinámico

El entrenamiento de redes neuronales profundas puede interpretarse como un proceso dinámico iterativo en el espacio de parámetros. En cada paso del algoritmo de optimización, los pesos del modelo se actualizan con el objetivo de minimizar una función de pérdida que cuantifica el error entre las predicciones y los datos observados. En el caso del descenso de gradiente, esta dinámica puede expresarse como: Metodología

$$w_{t+1} = w_t - \eta \nabla L(w_t)$$

donde $\mathbf{Wt}$ representa el vector de parámetros en la iteración $\mathbf{t}$, $L(\mathbf{w})$ es la función de pérdida y η corresponde a la tasa de aprendizaje.

Esta formulación permite describir el entrenamiento como un sistema dinámico discreto, en el que cada iteración define una transición en el espacio de estados. Desde esta perspectiva, la evolución del modelo puede analizarse en términos de estabilidad, convergencia y sensibilidad a perturbaciones, y no únicamente a partir del valor final de la pérdida. Li et al. (2018) destacan la utilidad de este enfoque para comprender con mayor profundidad el comportamiento del entrenamiento en redes neuronales profundas. En el presente trabajo, esta perspectiva constituye el punto de partida conceptual para examinar cómo la tasa de aprendizaje, la profundidad y el optimizador influyen sobre la dinámica observada durante el entrenamiento. Protocolo de transformación lingüística controlada

Geometría del paisaje de pérdida y estabilidad del entrenamiento

El comportamiento del entrenamiento depende, en parte, de la geometría local del paisaje de pérdida (loss landscape), es decir, de la estructura que adopta la función de pérdida en el espacio de parámetros. Aunque en redes neuronales modernas este paisaje se define en espacios de muy alta dimensión, su análisis puede abordarse mediante aproximaciones locales y proyecciones útiles para la interpretación empírica (Li et al., 2018).

Una herramienta central para describir esta geometría es la matriz Hessiana:

$$H = \nabla^2 L(\mathbf{w})$$

la cual informa sobre la curvatura local de la función de pérdida. En particular, el mayor autovalor de la Hessiana puede interpretarse como un indicador de sensibilidad local frente a pequeñas perturbaciones en los parámetros. Cuando la curvatura es elevada en relación con la tasa de aprendizaje, la dinámica del entrenamiento puede volverse más inestable, generando oscilaciones o incluso divergencia. De manera aproximada, una condición local de estabilidad puede expresarse como:

$$\eta \lambda_{\max}(H) < 2$$

donde $\lambda_{\max}(H)$ representa el mayor autovalor de la Hessiana.

Este vínculo entre curvatura y estabilidad ofrece una base teórica para examinar cómo la tasa de aprendizaje no solo incide sobre la velocidad de convergencia, sino también sobre el régimen dinámico en el que evoluciona el sistema.

Estabilidad, sensibilidad y generalización

Dentro del análisis del paisaje de pérdida, una distinción frecuente es la existente entre mínimos planos y mínimos agudos. Los mínimos planos corresponden a regiones en las que la pérdida varía suavemente ante pequeñas perturbaciones de los parámetros, mientras que

los mínimos agudos presentan cambios más abruptos frente a modificaciones de magnitud similar. Fort y Scherlis (2019) señalan que los modelos que convergen hacia regiones más planas tienden a exhibir mejor capacidad de generalización. Esta observación sugiere que la geometría local del paisaje no solo afecta la estabilidad de la optimización, sino que también puede guardar relación con el desempeño fuera de muestra.

Al mismo tiempo, el entrenamiento de redes neuronales no sigue una trayectoria estrictamente regular ni puramente determinista. Chaudhari y Soatto (2018) sostienen que el descenso de gradiente estocástico puede entenderse como una dinámica que combina componentes deterministas y estocásticos. En consecuencia, el comportamiento observado depende de la interacción entre datos, arquitectura, optimizador e hiperparámetros, lo que justifica estudiar la estabilidad del entrenamiento a partir de magnitudes observables durante su evolución.

Sensibilidad dinámica e indicadores empíricos

La teoría de sistemas dinámicos estudia la evolución de un sistema frente a perturbaciones y cambios en sus condiciones de operación. En este marco, una noción clásica para caracterizar sensibilidad es el exponente de Lyapunov, definido como la tasa promedio de divergencia entre trayectorias inicialmente cercanas:

$$\lambda = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \left| \frac{\delta x(t)}{\delta x(0)} \right|$$

Cuando $\lambda < 0$ las trayectorias tienden a converger; cuando $\lambda > 0$, la dinámica presenta divergencia sensible a perturbaciones iniciales. En aprendizaje profundo, esta intuición ha sido retomada para estudiar la estabilidad del entrenamiento y su posible relación con la reproducibilidad y el rendimiento del modelo. En esta línea, Ziyin et al. (2022) señalan que ciertos procesos de entrenamiento pueden mostrar sensibilidad marcada frente a perturbaciones y cambios en los hiperparámetros, afectando la consistencia de los resultados.

En este trabajo no se estima formalmente dicho exponente. En su lugar, se utiliza un indicador empírico de estabilidad construido a partir de las variaciones temporales de la función de pérdida, con el fin de aproximar de manera operativa la sensibilidad relativa del entrenamiento. El interés del enfoque no reside en demostrar comportamiento caótico, sino en examinar si un *proxy* dinámico computacionalmente accesible puede aportar información útil para caracterizar regímenes de entrenamiento y orientar la selección de hiperparámetros.

Delimitación conceptual del enfoque

Diversos trabajos sobre sistemas complejos sugieren que algunos procesos adaptativos pueden operar de manera eficiente en regiones intermedias entre dinámicas excesivamente rígidas y dinámicas altamente sensibles. En la literatura, esta intuición suele asociarse con la noción de *edge of chaos*. En el caso del entrenamiento de redes neuronales, esta idea

puede resultar útil como marco interpretativo para examinar si ciertos valores de la tasa de aprendizaje se vinculan con zonas intermedias particularmente favorables para el aprendizaje (Zhang et al., 2024).

No obstante, en el presente estudio esta noción se emplea únicamente como marco de lectura. El objetivo no consiste en demostrar formalmente que el entrenamiento de redes neuronales profundas opere en el borde del caos, sino en analizar si la estabilidad relativa observada durante la optimización permite distinguir regiones del espacio de hiperparámetros asociadas con comportamientos favorables. En este sentido, el trabajo adopta una perspectiva empírica y prudente, restringiendo sus conclusiones a la evidencia obtenida a partir de las trayectorias de entrenamiento analizada.

Formulación del problema

La selección de la tasa de aprendizaje en redes neuronales profundas suele realizarse mediante procedimientos empíricos que, aunque efectivos en muchos casos, implican un costo computacional elevado y ofrecen información limitada sobre la dinámica del entrenamiento. En consecuencia, persiste el problema de identificar criterios complementarios que permitan analizar la optimización no solo en función del rendimiento final, sino también de la estabilidad relativa con la que evoluciona el sistema.

En este marco, la pregunta de investigación que orienta el trabajo es la siguiente:

¿Es posible identificar regiones de la tasa de aprendizaje asociadas con cambios en la estabilidad dinámica del entrenamiento de redes neuronales profundas y utilizar esa información para orientar la selección de hiperparámetros?

Para abordar esta cuestión, el estudio utiliza un indicador empírico de estabilidad construido a partir de la evolución temporal de la función de pérdida y examina su relación con la tasa de aprendizaje, la arquitectura, el optimizador y el desempeño predictivo.

Hipótesis de investigación

A partir del marco teórico y del problema formulado, se plantean las siguientes hipótesis:

Hipótesis 1 (H1): El entrenamiento de redes neuronales profundas presenta variaciones sistemáticas en su estabilidad dinámica en función de la tasa de aprendizaje utilizada durante la optimización.

Hipótesis 2 (H2): Un indicador empírico de estabilidad dinámica permite caracterizar de manera útil la sensibilidad relativa del entrenamiento e identificar regiones relevantes dentro del espacio de hiperparámetros.

Hipótesis 3 (H3): Las regiones de estabilidad dinámica intermedia tienden a asociarse con desempeños predictivos competitivos y a contribuir a acotar la exploración del espacio de búsqueda de hiperparámetros.

Metodología experimental

Esta sección describe el diseño experimental adoptado para analizar la estabilidad dinámica del entrenamiento en redes neuronales profundas y su relación con la selección de hiperparámetros. En particular, se examina cómo la tasa de aprendizaje, la profundidad de la red y el optimizador influyen sobre la dinámica del proceso de optimización, así como la utilidad potencial de un indicador empírico de estabilidad para orientar la exploración del espacio de configuraciones. En línea con trabajos que destacan el valor de los experimentos controlados para aislar el efecto de los hiperparámetros sobre la evolución del sistema (Raghu et al., 2017), el diseño combina análisis dinámico, evaluación predictiva y comparación entre selección tradicional y criterio dinámico.

Los experimentos se implementaron en Python utilizando PyTorch y bibliotecas complementarias para procesamiento numérico, visualización y análisis de resultados. En todos los casos se trabajó con semillas controladas y con separación explícita entre las etapas de entrenamiento, validación y prueba cuando la tarea involucró selección de hiperparámetros. El conjunto de prueba se reservó exclusivamente para la evaluación final y no intervino ni en la definición del indicador ni en la selección de la tasa de aprendizaje. De este modo, las comparaciones entre estrategias se realizaron bajo un protocolo homogéneo y sin reutilización de información de test durante la etapa de ajuste.

Arquitecturas consideradas

Para el análisis experimental se emplearon arquitecturas neuronales simples, pero suficientemente expresivas para capturar comportamientos dinámicos relevantes durante el entrenamiento. En los experimentos con imágenes se utilizaron variantes de perceptrón multicapa (MLP) con capas completamente conectadas, adaptadas a la dimensión de entrada de cada conjunto de datos.

En MNIST se utilizó un MLP con cuatro capas ocultas de 512, 256, 128 y 64 neuronas, seguido por una capa de salida de 10 unidades. Para el análisis de profundidad variable se definió una arquitectura paramétrica (DynamicMLP) con entre 2 y 6 capas ocultas de 256 neuronas. En CIFAR-10 se empleó un MLP adaptado a imágenes RGB de 32×32 píxeles, con capas de 1024, 512, 256 y 128 unidades antes de la capa de salida. En FashionMNIST y en la evaluación multi-dataset se utilizaron arquitecturas más compactas, mientras que en los conjuntos tabulares se empleó una red con dos capas ocultas de 32 y 16 neuronas.

La elección de estas arquitecturas respondió a un criterio de control experimental: permiten observar variaciones en estabilidad, sensibilidad y rendimiento sin volver impracticables

los barridos sistemáticos y las repeticiones multi-seed. En consecuencia, el objetivo no fue comparar arquitecturas de última generación, sino estudiar regularidades dinámicas en escenarios controlados (Barrett & Dherin, 2021).

Conjuntos de datos

El estudio se desarrolló sobre múltiples conjuntos de datos estándar. En los experimentos principales de visión por computadora se emplearon MNIST, FashionMNIST y CIFAR-10. MNIST y FashionMNIST contienen imágenes en escala de grises de 28×28 píxeles, mientras que CIFAR-10 contiene imágenes RGB de 32×32 píxeles distribuidas en diez clases. Además, en la evaluación multi-dataset se incorporaron los conjuntos tabulares Wine, Breast Cancer y Digits, con el propósito de examinar si los patrones observados se sostenían en dominios distintos al de visión.

La selección de estos conjuntos respondió a tres criterios: su uso extendido en la literatura, la posibilidad de cubrir niveles crecientes de complejidad y la conveniencia de trabajar en escenarios computacionalmente controlables. MNIST y FashionMNIST permiten observar trayectorias de entrenamiento en problemas relativamente accesibles, CIFAR-10 introduce una tarea más exigente, y los conjuntos tabulares extienden la evaluación a un dominio diferente.

Para reducir el costo computacional, en algunos experimentos se utilizó submuestreo del conjunto de entrenamiento: 10.000 imágenes en MNIST y CIFAR-10, 8.000 en FashionMNIST y 6.000 en la evaluación multi-dataset de visión. En los experimentos que involucraron selección de hiperparámetros, el conjunto de entrenamiento se dividió adicionalmente en particiones de entrenamiento y validación, reservando el conjunto de prueba exclusivamente para la evaluación final. En los conjuntos tabulares, el escalado de variables se ajustó únicamente sobre los datos de entrenamiento de cada partición y luego se aplicó a validación y prueba, evitando filtración de información.

Barrido sistemático de la tasa de aprendizaje

Con el fin de estudiar el efecto de la tasa de aprendizaje sobre la dinámica del entrenamiento (Goodfellow et al., 2016), se realizaron barridos sistemáticos del hiperparámetro η en escala logarítmica. En el experimento principal con MNIST y en la extensión a CIFAR-10 se utilizó una grilla de 25 valores generados mediante $\text{np.logspace}(-4, 0.3, 25)$, lo que permitió explorar aproximadamente el intervalo $[10^{-4}, 2]$. En FashionMNIST se utilizaron 12 valores en $[10^{-4}, 1]$, mientras que en la evaluación multi-dataset y en la validación multi-seed se emplearon 10 valores en $[10^{-4}, 1]$.

La grilla de exploración se fijó antes del análisis comparativo y se mantuvo constante dentro de cada protocolo experimental. Esta decisión buscó evitar ajustes ad hoc del espacio de búsqueda en función de resultados observados ex post. El procedimiento permitió analizar el comportamiento del entrenamiento desde configuraciones conservadoras, asociadas con

trayectorias más lentas y regulares, hasta configuraciones más agresivas, en las que podían aparecer mayores niveles de sensibilidad o inestabilidad.

Registro de trayectorias dinámicas

Para caracterizar la dinámica del sistema durante el entrenamiento, se registró la evolución temporal de distintas variables a lo largo de las iteraciones del algoritmo de optimización. En el experimento principal sobre MNIST se almacenaron tres series temporales: la función de pérdida, la norma del vector de parámetros del modelo y la norma del gradiente. Estas magnitudes permitieron construir una representación empírica de la trayectoria del sistema y estudiar cómo cambia su comportamiento al modificar la tasa de aprendizaje.

En los experimentos posteriores, el análisis dinámico se apoyó principalmente en la serie temporal de la pérdida de entrenamiento, a partir de la cual se calculó el indicador empírico de estabilidad. Sobre esta base se elaboraron visualizaciones temporales, espacios de fase y trayectorias en el espacio de estados. Dado que el indicador se construye a partir de la pérdida observada durante la optimización, su interpretación se restringe a comparaciones realizadas dentro de un mismo protocolo experimental y no a comparaciones absolutas entre tareas con escalas de pérdida diferentes.

Construcción del indicador empírico de estabilidad dinámica

El indicador empírico de estabilidad dinámica se construyó a partir de la serie temporal de la función de pérdida registrada durante el entrenamiento. Dada una secuencia de pérdidas $\{L_t\}_{t=1}^T$, se calcularon primero las diferencias absolutas entre valores consecutivos:

$$\Delta_t = |L_{t+1} - L_t|$$

Luego se descartaron las diferencias menores o iguales a 10^{-8} , con el fin de evitar problemas numéricos en el cálculo del logaritmo. Si la cantidad de diferencias válidas resultante era menor que cinco, el indicador se consideró no definido y se registró como valor faltante. En caso contrario, el indicador se definió como:

$$S = \frac{1}{N} \sum_{t=1}^N \log(\Delta_t)$$

donde N representa el número de diferencias válidas conservadas tras el filtrado.

En términos computacionales, esta definición corresponde al promedio del logaritmo natural de las variaciones absolutas entre pérdidas consecutivas. El indicador no constituye una estimación formal del exponente de Lyapunov, sino un proxy empírico de la sensibilidad relativa del entrenamiento. En términos interpretativos, valores más negativos se asocian con trayectorias más regulares y menos oscilatorias, mientras que valores cercanos a 0 indican variaciones inter-iteración más marcadas y, por lo tanto, mayor sensibilidad dinámica. Si la cantidad de diferencias válidas resultaba insuficiente, el indicador se consideró no

informativo para el análisis dinámico, evitando forzar una estimación sobre trayectorias sin variación numérica suficiente.

Identificación de regiones dinámicas intermedias

A partir del comportamiento del indicador S para distintos valores de la tasa de aprendizaje, se buscaron regiones del espacio de hiperparámetros en las que el entrenamiento se aproximaba a una transición entre dinámicas relativamente estables y dinámicas más sensibles. En el análisis descriptivo general, estas transiciones se interpretaron a partir de la evolución del indicador y de su proximidad relativa a 0. En la extensión a CIFAR-10, además, se identificó para cada optimizador la tasa de aprendizaje cuyo valor absoluto del indicador resultó mínimo, utilizándola como referencia empírica de transición dinámica.

Sin embargo, en los experimentos de selección de hiperparámetros no se utilizó la cercanía a 0 como regla operativa principal. En su lugar, la región de interés se definió mediante la franja:

$$-4 < S < -2$$

Esta banda fue adoptada como criterio heurístico de trabajo para delimitar una región intermedia de estabilidad relativa. La elección se apoyó en la inspección exploratoria inicial de las trayectorias y en la necesidad de distinguir, con una regla simple y reproducible, configuraciones excesivamente rígidas de otras claramente más oscilatorias. En este sentido, valores muy negativos tienden a corresponder a trayectorias con variaciones inter-iteración muy pequeñas, mientras que valores próximos a 0 reflejan fluctuaciones más marcadas. La franja seleccionada busca, por lo tanto, capturar una zona intermedia de interés dinámico dentro del protocolo adoptado.

Con el fin de evitar que esta decisión se interprete como un ajuste ex post del criterio de selección, la banda se mantuvo fija en los experimentos comparativos una vez definida y no se recalibró según dataset, optimizador o semilla. En consecuencia, debe entenderse como una regla empírica y operativa, dependiente del diseño experimental, y no como un umbral teórico universal.

Comparación entre selección por rendimiento y criterio dinámico

Con el fin de explorar una aplicación práctica del análisis propuesto, se compararon dos estrategias de selección de tasa de aprendizaje.

La primera correspondió a la selección tradicional por rendimiento, definida como la identificación, dentro de la grilla explorada, de la configuración con mayor accuracy de validación. La segunda se basó en un criterio dinámico formulado como regla operativa explícita y reproducible.

En FashionMNIST, primero se restringió el análisis a configuraciones con accuracy de validación superior a 0,2. Este umbral se introdujo como filtro mínimo de desempeño para

excluir configuraciones claramente no informativas y evitar que el criterio dinámico privilegiara trayectorias relativamente regulares, pero irrelevantes en términos predictivos. Dado que se trata de una tarea de diez clases, el valor 0,2 equivale a un nivel claramente superior al azar y opera aquí como condición de elegibilidad, no como criterio de selección final. Luego, dentro de ese subconjunto elegible, se seleccionaron las configuraciones cuyo indicador cumplía la condición $-4 < S < -2$. Si existían varias configuraciones dentro de esa región, se eligió la de mayor accuracy de validación; si la región estaba vacía, se seleccionó la configuración válida con mayor valor del indicador.

En la evaluación multi-dataset se aplicó una regla análoga, pero sin introducir un umbral fijo único de accuracy entre tareas heterogéneas, con el fin de preservar la comparabilidad entre datasets con distinta dificultad y distinto número de clases. En ese caso, el criterio dinámico consistió en seleccionar la configuración con mayor accuracy de validación dentro de la región $-4 < S < -2$, o bien, si dicha región estaba vacía, elegir la configuración con mayor valor del indicador. Las configuraciones con indicador no definido no integraron la región dinámica intermedia utilizada en la selección y se interpretaron como casos no informativos para ese análisis específico.

En todos los casos, una vez seleccionados los learning rates, los modelos se reentrenaron y se evaluaron finalmente sobre el conjunto de prueba. El propósito de esta comparación no consistió en reemplazar la búsqueda tradicional, sino en examinar si el criterio dinámico podía contribuir a acotar la región de búsqueda manteniendo niveles de rendimiento final comparables.

Evaluación multi-dataset y validación multi-seed

Para reforzar la robustez del análisis, el estudio incluyó una evaluación sobre múltiples conjuntos de datos y una validación adicional con diferentes semillas aleatorias. En la validación multi-seed se utilizaron las semillas 7, 21, 42, 84 y 123. Esta etapa se aplicó sobre MNIST, FashionMNIST, Wine, BreastCancer y Digits, mientras que CIFAR-10 se utilizó como escenario de mayor complejidad en la extensión con múltiples optimizadores. Para cada semilla se repitió el experimento de exploración sobre la grilla de 10 tasas de aprendizaje, registrando estabilidad dinámica, accuracy de validación y desempeño final en test, y comparando selección por rendimiento y criterio dinámico.

Sobre esta base, se calcularon métricas de eficiencia exploratoria, como la Hyperparameter Search Efficiency (HSE), definida como:

$$HSE = \frac{Accuracy}{N_{configs}}$$

donde **Accuracy** representa el rendimiento final en test alcanzado por cada estrategia y $N_{configs}$ el número de configuraciones efectivamente retenidas o consideradas en la etapa de selección. Esta métrica no pretende estimar ahorro computacional en sentido estricto, sino resumir la relación entre desempeño final y tamaño de la región candidata identificada por

cada enfoque dentro del protocolo experimental adoptado. En este sentido, HSE permite comparar la eficiencia relativa entre la selección por rendimiento y el criterio dinámico como medida exploratoria de compactación del espacio de búsqueda.

Adicionalmente, se calculó la ganancia relativa de eficiencia exploratoria (HSE Gain), definida como el incremento porcentual de HSE del criterio dinámico respecto de la selección tradicional:

$$\text{HSE Gain} = \frac{\text{HSE}_{\text{dinámico}} - \text{HSE}_{\text{tradicional}}}{\text{HSE}_{\text{tradicional}}} \times 100$$

Esta medida resume en términos relativos la mejora de eficiencia obtenida al restringir la región candidata mediante el criterio dinámico. Su interpretación debe entenderse dentro del protocolo comparativo adoptado y no como una estimación directa del costo computacional absoluto de cada estrategia.

Reproducibilidad experimental

Con el fin de favorecer la trazabilidad y la reproducibilidad del estudio, el proceso experimental se implementó registrando resultados intermedios, tablas y figuras en formatos reutilizables. Asimismo, las corridas se realizaron con semillas aleatorias controladas y con separación explícita entre las etapas de entrenamiento, validación y prueba en los experimentos que involucraron selección de hiperparámetros.

El código y los experimentos utilizados en este trabajo se encuentran disponibles en un cuaderno de Google Colab, incorporado como material complementario del estudio¹

Resultados

Estabilidad dinámica del entrenamiento en función de la tasa de aprendizaje

El primer eje de resultados se orientó a examinar cómo varía la estabilidad dinámica del entrenamiento al modificar la tasa de aprendizaje. Para ello, se calculó el indicador empírico de estabilidad dinámica a partir de la evolución temporal de la función de pérdida, con el fin de sintetizar el grado relativo de regularidad o sensibilidad del sistema bajo distintas configuraciones del hiperparámetro.

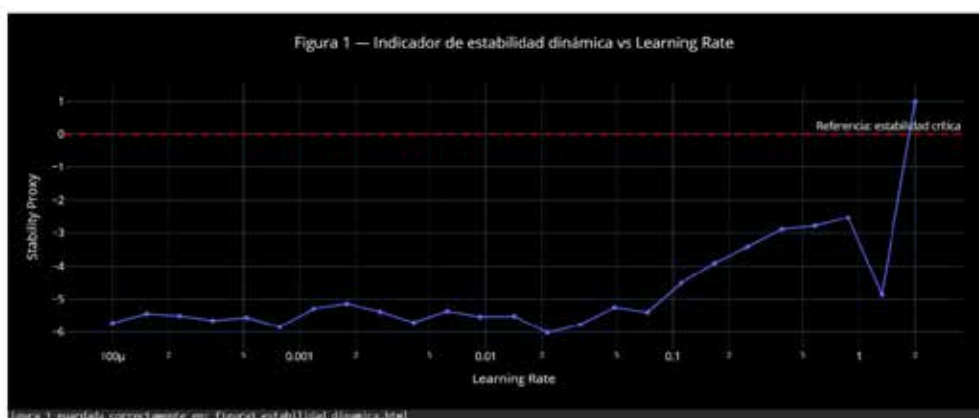
TABLA 1. INDICADOR EMPÍRICO DE ESTABILIDAD DINÁMICA PARA CADA TASA DE APRENDIZAJE

Learning Rate	Stability Proxy
0	0.000100
1	0.000151
2	0.000228
3	0.000345
4	0.000521
5	0.000787
6	0.001189
7	0.001786
8	0.002712
9	0.004097
10	0.006180
11	0.009350
12	0.014125
13	0.021339
14	0.032235
15	0.048607

Fuente: elaboración propia 2026

Los resultados muestran una tendencia sistemática. Para valores pequeños de la tasa de aprendizaje, el indicador adopta valores marcadamente negativos, compatibles con trayectorias relativamente regulares. A medida que la tasa de aprendizaje aumenta, el indicador se desplaza progresivamente hacia 0 y, en la configuración más extrema del barrido, alcanza incluso valores positivos. En términos empíricos, este comportamiento sugiere una disminución gradual de la estabilidad relativa del entrenamiento y una mayor sensibilidad inter-iteración.

Para apreciar con mayor claridad la forma global de esta transición, se incorpora una representación gráfica del indicador a lo largo del rango analizado.

FIGURA 1. INDICADOR EMPÍRICO DE ESTABILIDAD DINÁMICA PARA CADA TASA DE APRENDIZAJE


Fuente: elaboración propia 2026

La figura 1 resulta consistente con el patrón observado en la tabla: el incremento de la tasa de aprendizaje desplaza al sistema desde una región de mayor regularidad hacia otra de mayor sensibilidad dinámica. Esta transición puede interpretarse como compatible con un pasaje desde regímenes más estables hacia zonas dinámicamente más activas, sin que ello implique una demostración formal de caos.

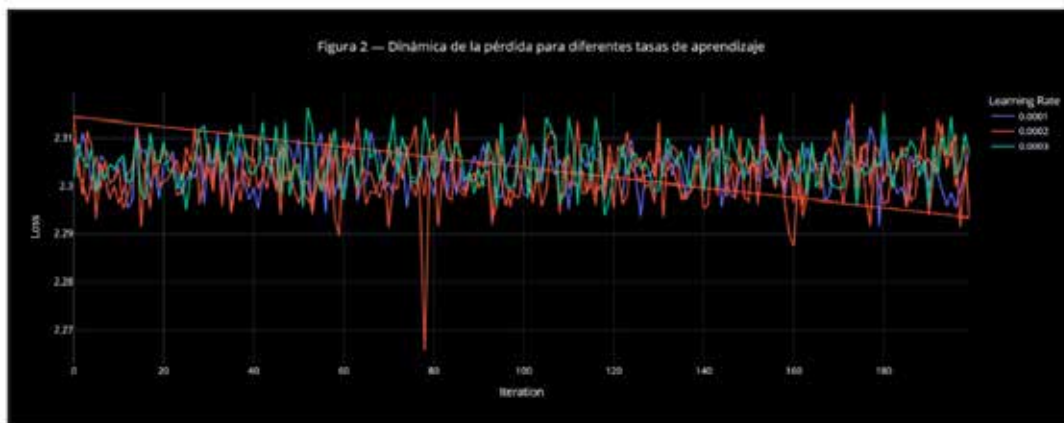
En conjunto, estos resultados sugieren que la tasa de aprendizaje no solo afecta la velocidad de optimización, sino también la dinámica del entrenamiento.

Dinámica temporal del entrenamiento

Una vez identificada la tendencia global del indicador de estabilidad, se analizó cómo esa transición se expresa en la evolución temporal de la función de pérdida. Mientras la subsección anterior ofreció una visión agregada del comportamiento dinámico, el análisis de trayectorias permite observar con mayor detalle la evolución del sistema durante las primeras iteraciones del entrenamiento.

Con este propósito, se incorporó una visualización comparativa de la pérdida para distintas tasas de aprendizaje.

FIGURA 2. DINÁMICA DE LA PÉRDIDA DURANTE EL ENTRENAMIENTO



Fuente: elaboración propia 2026

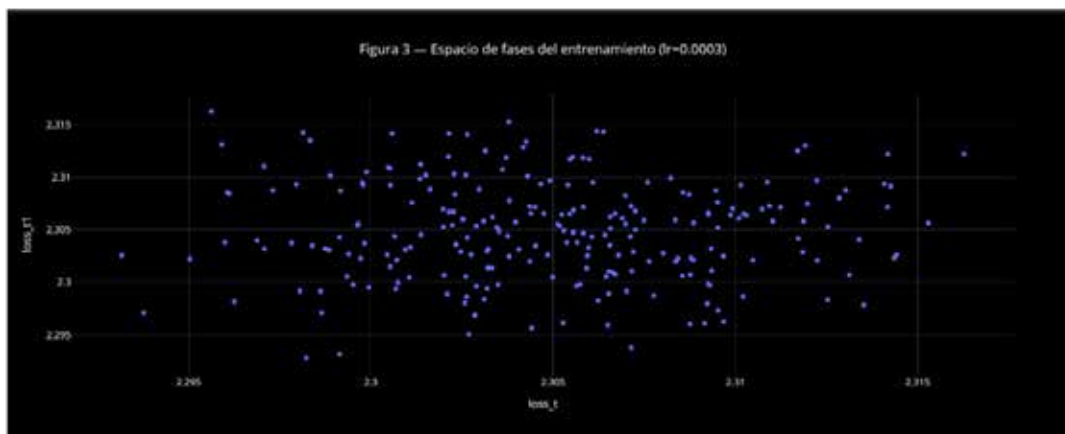
La figura 2 muestra que, en todos los casos considerados, las trayectorias permanecen acotadas alrededor de un valor cercano a 2.30, sin evidenciar comportamientos explosivos en esta etapa inicial. No obstante, las tasas de aprendizaje más pequeñas producen trayectorias más regulares, mientras que los valores mayores introducen oscilaciones más marcadas. En consecuencia, aun dentro de una región globalmente estable, la sensibilidad dinámica del sistema aumenta a medida que crece la tasa de aprendizaje.

Este patrón resulta consistente con lo observado en la tabla 1 y en la figura 1: allí donde el indicador se aproximaba a 0, aparecen trayectorias menos regulares y relativamente más oscilatorias.

Representación del entrenamiento en el espacio de estados

Con el propósito de profundizar la caracterización dinámica del entrenamiento, se construyó una representación bidimensional del espacio de fases a partir de la relación entre la pérdida en una iteración y la pérdida en la iteración siguiente. Este tipo de visualización permite examinar si la trayectoria del sistema permanece confinada en una región acotada o si, por el contrario, tiende a dispersarse hacia comportamientos menos regulares.

FIGURA 3. ESPACIO DE FASES DEL ENTRENAMIENTO



FUENTE: elaboración propia 2026

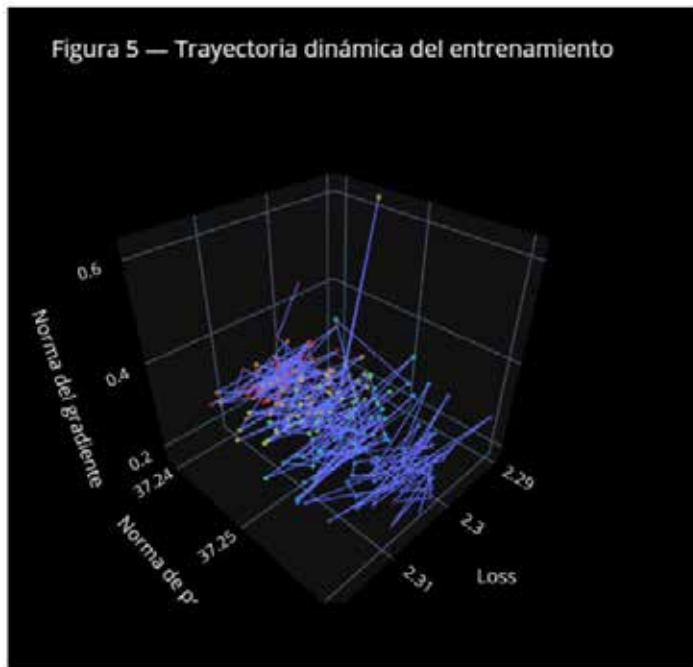
La figura 3 muestra una nube de puntos relativamente compacta, concentrada en una región limitada del plano. Esta disposición sugiere que, en el tramo analizado, la dinámica del entrenamiento se mantiene acotada y estructurada, con oscilaciones moderadas alrededor de una región central. En otros términos, la trayectoria no exhibe una dispersión arbitraria, sino una organización geométrica compatible con un régimen globalmente estable. En diálogo con la tabla 1 y la figura 2, esta representación resulta consistente con la lectura anterior: el entrenamiento se desarrolla dentro de una región acotada, aunque con sensibilidad creciente al aumentar la tasa de aprendizaje.

Relación entre estabilidad dinámica y rendimiento predictivo

Luego de analizar la estabilidad del sistema desde una perspectiva dinámica, se estudió su relación con el desempeño predictivo del modelo, medido mediante accuracy de validación. Esta relación resulta central, ya que permite examinar si las regiones de mayor estabilidad relativa coinciden, o no, con las zonas de mejor rendimiento durante la etapa de selección de hiperparámetros.

Además de la representación bidimensional del espacio de fases, se construyó una trayectoria tridimensional utilizando simultáneamente la pérdida, la norma de los pesos y la norma del gradiente. Esta visualización permite observar cómo evoluciona el sistema cuando se consideran de manera conjunta distintas magnitudes relevantes del entrenamiento.

FIGURA 5. TRAYECTORIA DINÁMICA DEL ENTRENAMIENTO EN EL ESPACIO DE ESTADOS



Fuente: elaboración propia 2026

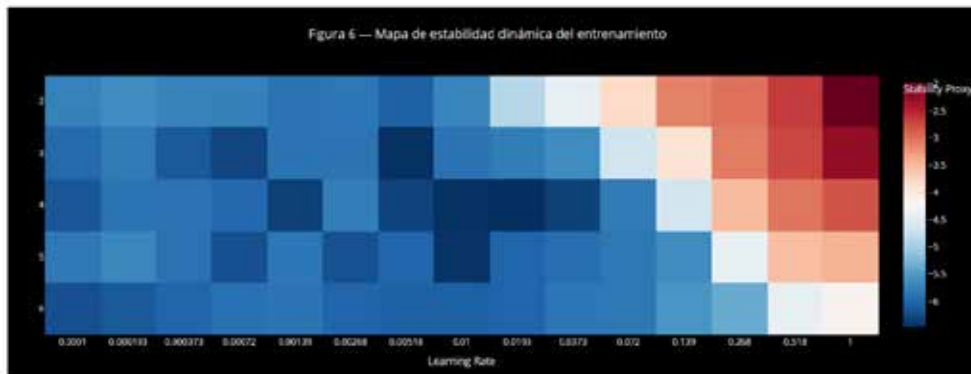
La figura 5 muestra que la trayectoria del sistema permanece confinada dentro de una región relativamente acotada del espacio tridimensional. En lugar de dispersarse de manera irregular, la evolución sigue un recorrido estructurado, lo que refuerza la evidencia de estabilidad relativa observada en las representaciones anteriores. La figura resulta útil porque no solo expone la evolución de la pérdida, sino que incorpora también la magnitud de los parámetros y la intensidad de las actualizaciones, ofreciendo una visión más completa del proceso de entrenamiento.

En consecuencia, la trayectoria tridimensional complementa la interpretación desarrollada hasta aquí: el entrenamiento no solo presenta regularidad temporal, sino también una organización geométrica observable en el espacio de estados.

Efecto de la profundidad de la red sobre la estabilidad dinámica

Hasta este punto, el análisis se concentró principalmente en el papel de la tasa de aprendizaje. Sin embargo, la estabilidad del entrenamiento también puede verse afectada por la arquitectura de la red. Con este fin, se construyó una arquitectura paramétrica con número variable de capas ocultas y se calculó el indicador de estabilidad dinámica para distintas combinaciones de profundidad y tasa de aprendizaje.

FIGURA 6. MAPA DE ESTABILIDAD DINÁMICA DEL ENTRENAMIENTO SEGÚN TASA DE APRENDIZAJE Y PROFUNDIDAD



Fuente: elaboración propia 2026

La figura 6 presenta un mapa de calor en el que cada celda resume el valor del indicador empírico de estabilidad dinámica para una combinación específica de profundidad y tasa de aprendizaje. En líneas generales, las tasas de aprendizaje pequeñas se asocian con regiones relativamente más estables, mientras que los valores mayores conducen a una dinámica más sensible. Al mismo tiempo, la profundidad introduce variaciones adicionales en la distribución de estas regiones, lo que indica que la estabilidad del sistema emerge de la interacción entre múltiples hiperparámetros y no exclusivamente del valor de la tasa de aprendizaje.

Este resultado amplía la interpretación anterior: la dinámica del entrenamiento no puede entenderse únicamente como una función de la tasa de aprendizaje, sino también en relación con la complejidad estructural del modelo.

Extensión del análisis a CIFAR-10 y múltiples optimizadores

Con el objetivo de evaluar la generalidad del enfoque, el análisis se extendió a un problema de mayor complejidad utilizando el conjunto de datos CIFAR-10 y tres algoritmos de optimización distintos: SGD, Adam y RMSProp. Esta extensión permite examinar si los patrones observados en un entorno más controlado se sostienen cuando aumenta la dificultad de la tarea y cuando cambia la regla de actualización de parámetros.

TABLA 2. RESULTADOS DE ESTABILIDAD DINÁMICA Y ACCURACY DE VALIDACIÓN EN CIFAR-10

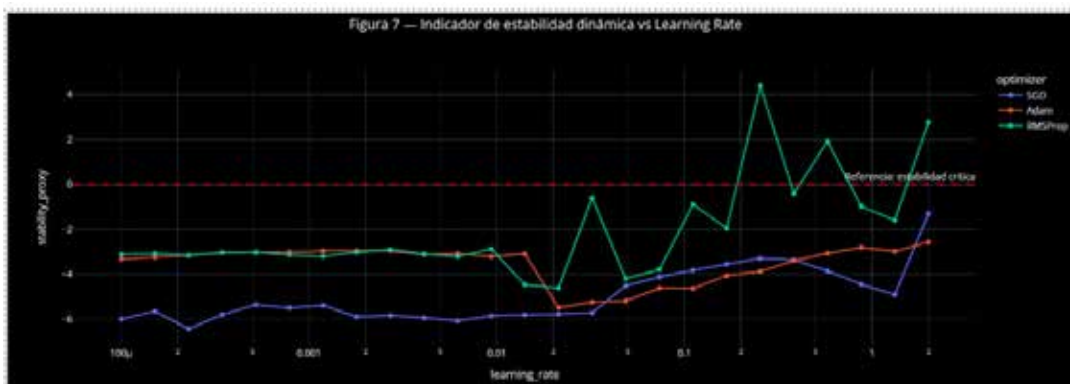
	optimizer	learning_rate	stability_proxy	val_accuracy
0	SGD	0.000100	-5.456190	0.1070
1	SGD	0.000151	-6.000757	0.1130
2	SGD	0.000228	-5.363980	0.1015
3	SGD	0.000345	-5.505495	0.1110
4	SGD	0.000521	-5.889977	0.1005

Fuente: elaboración propia 2026

La tabla 2 resume los resultados correspondientes a 75 configuraciones de entrenamiento, resultantes de la combinación de tres optimizadores y 25 tasas de aprendizaje. Para cada configuración se informa el optimizador, la tasa de aprendizaje, el indicador empírico de estabilidad dinámica y la accuracy de validación obtenida durante la exploración de hiperparámetros. En las primeras configuraciones se observan valores negativos del indicador y accuracies de validación cercanas a 0.10, lo que sugiere una dinámica relativamente estable, aunque con bajo rendimiento inicial. Este comportamiento resulta coherente con la mayor complejidad del problema, ya que CIFAR-10 exige un proceso de aprendizaje más rico para apartarse de los niveles iniciales de desempeño.

Sobre esta base, la representación gráfica permite examinar con mayor detalle cómo cambia la estabilidad según el optimizador considerado.

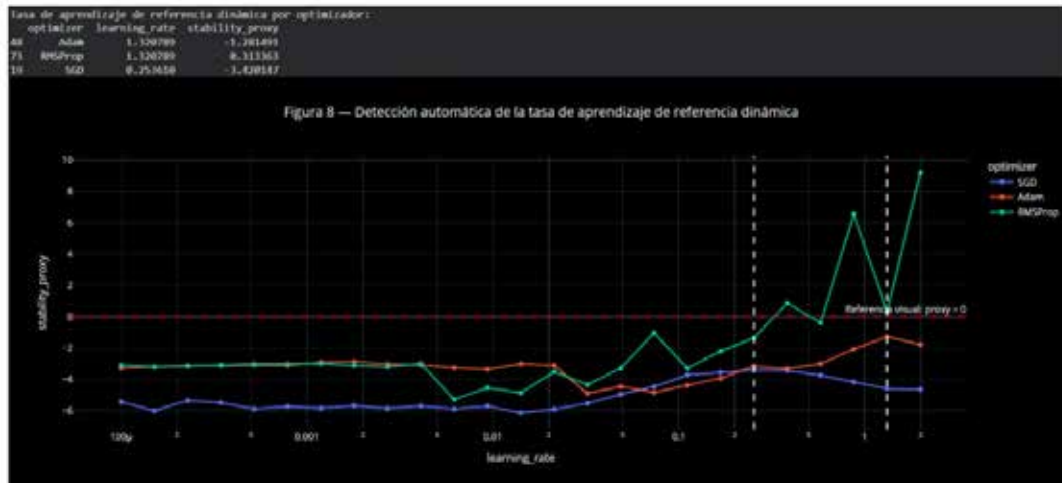
FIGURA 7. INDICADOR DE ESTABILIDAD DINÁMICA EN FUNCIÓN DE LA TASA DE APRENDIZAJE PARA DISTINTOS OPTIMIZADORES



Fuente: elaboración propia 2026

La figura 7 muestra que, para valores pequeños de la tasa de aprendizaje, el indicador permanece en valores negativos, compatibles con una dinámica relativamente estable. A medida que la tasa de aprendizaje aumenta, el indicador se aproxima a 0 y, en algunos optimizadores, incluso supera esa referencia, señalando regiones de mayor sensibilidad e inestabilidad. De este modo, los resultados sugieren que la estabilidad del entrenamiento depende no solo de la tasa de aprendizaje, sino también del algoritmo de optimización utilizado.

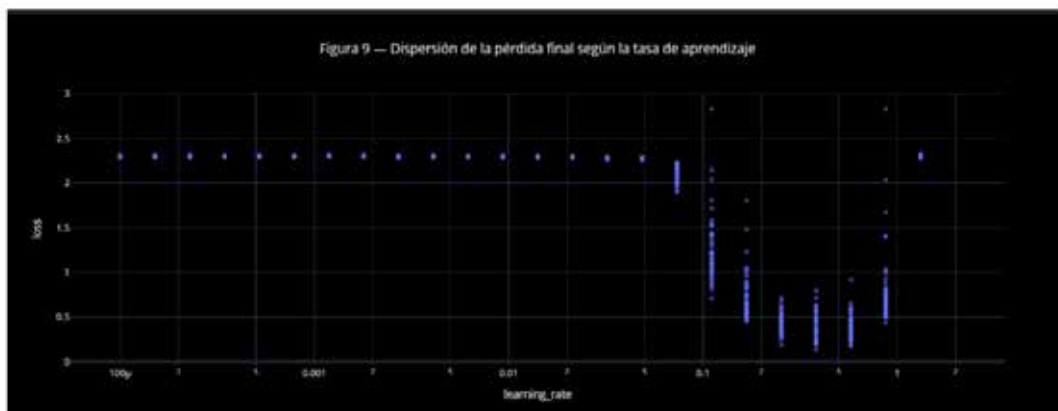
Para identificar empíricamente estas zonas de transición, se calculó la tasa de aprendizaje cuyo indicador se aproxima más a 0 para cada optimizador.

FIGURA 8. DETECCIÓN AUTOMÁTICA DE UNA TASA DE APRENDIZAJE DE REFERENCIA DINÁMICA


Fuente: elaboración propia 2026

La figura 8 muestra que la tasa de aprendizaje de referencia dinámica estimada varía según el optimizador. En este experimento, el valor detectado se ubica alrededor de 0.254 para SGD, con un indicador de estabilidad de -3.420, mientras que para Adam y RMSProp se localiza cerca de 1.321, con valores de -1.281 y 0.313, respectivamente. Estos puntos corresponden, en cada caso, a las configuraciones en las que el indicador empírico de estabilidad resulta más próximo a 0 dentro del rango explorado, por lo que pueden interpretarse como referencias empíricas de transición hacia regiones de mayor sensibilidad dinámica. En conjunto, la figura sugiere que la elección de la tasa de aprendizaje no depende solo del rendimiento del modelo, sino también de la dinámica interna del entrenamiento.

Como complemento, se examinó la dispersión de la pérdida en las últimas iteraciones del entrenamiento, con el propósito de visualizar de manera más directa cómo cambia el comportamiento del sistema en la fase final de la optimización.

FIGURA 9. DISPERSIÓN DE LA PÉRDIDA FINAL SEGÚN LA TASA DE APRENDIZAJE


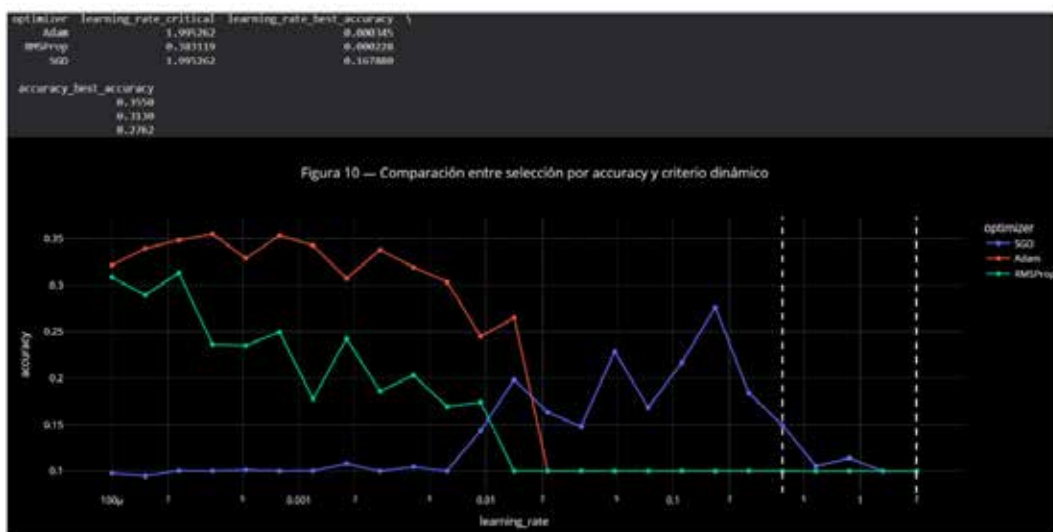
Fuente: elaboración propia 2026

La figura 9 muestra que, para tasas de aprendizaje pequeñas, la pérdida final permanece alta y concentrada, reflejando un entrenamiento estable pero poco eficiente. Para valores intermedios, la pérdida disminuye notablemente, lo que sugiere una región más favorable para el aprendizaje. En cambio, para tasas más altas, la dispersión aumenta y aparecen trayectorias más irregulares, indicando una mayor sensibilidad dinámica del entrenamiento. En conjunto, esta figura refuerza la idea de que existe una región intermedia donde el sistema combina una pérdida más baja con una variabilidad aún controlada. De esta manera, la extensión a CIFAR-10 confirma que la perspectiva dinámica conserva capacidad descriptiva incluso en un problema más exigente.

Comparación entre selección por rendimiento y criterio dinámico en CIFAR-10

Con el propósito de contrastar la información aportada por el análisis dinámico con la selección tradicional basada en desempeño, se comparó la tasa de aprendizaje seleccionada por el criterio de máxima accuracy de validación con la identificada a partir del criterio dinámico.

FIGURA 10. COMPARACIÓN ENTRE SELECCIÓN POR ACCURACY DE VALIDACIÓN Y CRITERIO DINÁMICO



Fuente: elaboración propia 2026

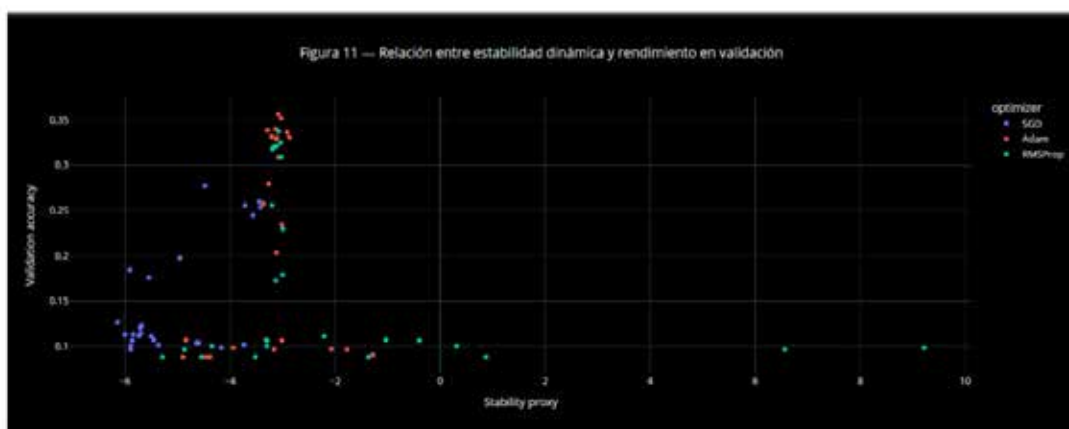
La figura 10 muestra que la selección por máxima accuracy de validación y la selección basada en criterio dinámico no siempre coinciden. Mientras el primer enfoque privilegia el mejor rendimiento predictivo observado durante la etapa de validación, el segundo identifica configuraciones ubicadas en una región dinámica intermedia del entrenamiento. En este sentido, ambos criterios deben entenderse como complementarios, ya que aportan información diferente sobre el comportamiento del sistema. Este resultado indica que la dinámica del

entrenamiento introduce una dimensión de análisis que no queda completamente reflejada en la métrica de validación utilizada para seleccionar hiperparámetros.

Relación entre estabilidad dinámica y rendimiento en CIFAR-10

A fin de vincular directamente el comportamiento dinámico del entrenamiento con el desempeño predictivo del modelo, se estudió la relación entre el indicador empírico de estabilidad dinámica y la accuracy de validación obtenida en CIFAR-10.

FIGURA 11. RELACIÓN ENTRE ESTABILIDAD DINÁMICA Y RENDIMIENTO EN VALIDACIÓN



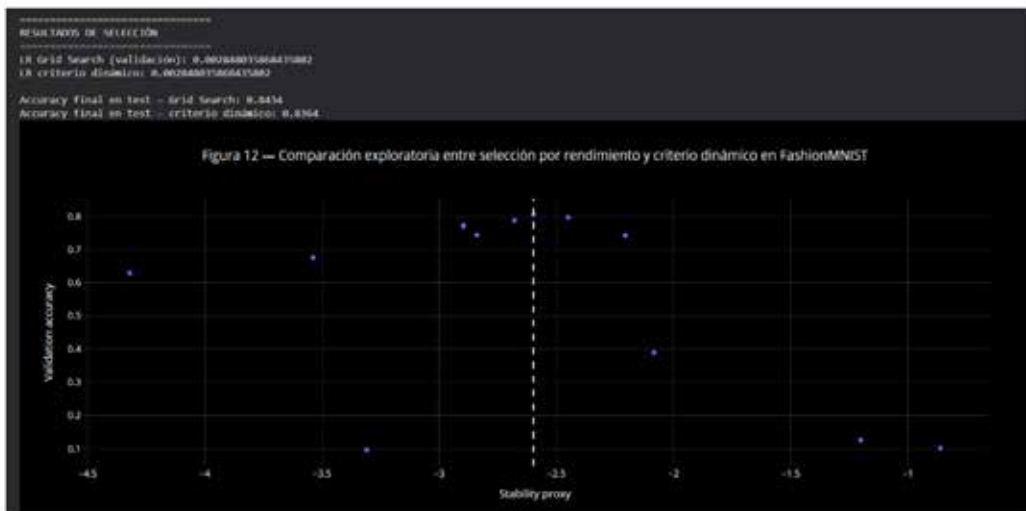
Fuente: elaboración propia 2026

La figura 11 muestra que los mayores valores de accuracy de validación tienden a concentrarse en una región intermedia del indicador empírico de estabilidad dinámica. En cambio, los valores muy negativos o los más próximos a 0 no se asocian de manera sistemática con mejor desempeño predictivo. En conjunto, este patrón sugiere que la relación entre estabilidad y rendimiento en validación no es lineal y que las configuraciones más favorables tienden a ubicarse en una zona de estabilidad relativa intermedia. Este hallazgo conecta de manera directa la parte descriptiva del trabajo con su dimensión aplicada.

Comparación exploratoria entre selección por rendimiento y criterio dinámico

Con el fin de evaluar una posible aplicación práctica del análisis dinámico, se comparó la selección tradicional de la tasa de aprendizaje basada en el máximo rendimiento en validación con una selección restringida por criterio dinámico en FashionMNIST.

FIGURA 12. COMPARACIÓN EXPLORATORIA ENTRE SELECCIÓN POR RENDIMIENTO Y CRITERIO DINÁMICO EN FASHIONMNIST



Fuente: elaboración propia 2026

La figura 12 muestra que, en este experimento sobre FashionMNIST, el criterio dinámico y la selección por rendimiento en validación conducen a una región prácticamente equivalente del espacio de learning rates. Ambos enfoques seleccionan un valor cercano a 0.00285 y, tras el reentrenamiento final, alcanzan desempeños muy similares en test (0.8434 para Grid Search y 0.8364 para el criterio dinámico). Este resultado sugiere que la estabilidad dinámica puede aportar una referencia útil para restringir la búsqueda de hiperparámetros sin alejarse de configuraciones competitivas.

Evaluación multi-dataset y eficiencia exploratoria

La comparación entre selección por rendimiento y criterio dinámico se extendió a múltiples conjuntos de datos, incluyendo problemas de visión por computadora y conjuntos tabulares. Esta ampliación permitió evaluar el alcance del enfoque en contextos experimentales heterogéneos.

TABLA 3. COMPARACIÓN MULTI-DATASET ENTRE SELECCIÓN POR RENDIMIENTO Y CRITERIO DINÁMICO

dataset	grid_configs	dynamic_configs	exploration_reduction_3	grid_lr	dynamic_lr	grid_val_accuracy	dynamic_val_accuracy	grid_test_accuracy	dynamic_test_accuracy
MNIST	10	6	40.0	0.005995	0.005995	0.9158	0.9158	0.9383	0.9354
fmnMIST	10	6	40.0	0.016681	0.016681	0.7883	0.7883	0.8292	0.8292
Wine	10	6	40.0	0.016681	0.016681	0.9600	0.9600	0.9815	0.9815
breastCancer	10	4	60.0	0.016681	0.016681	0.9790	0.9790	0.9849	0.9798
Digits	10	6	40.0	0.016681	0.016681	0.9365	0.9365	0.9444	0.9500

FUENTE: elaboración propia 2026

La tabla 3 resume los resultados obtenidos en MNIST, FashionMNIST, Wine, BreastCancer y Digits. Para cada conjunto se informa el número total de configuraciones exploradas, la cantidad de configuraciones retenidas por el criterio dinámico, la reducción relativa del espacio de búsqueda, las tasas de aprendizaje seleccionadas por máxima accuracy de validación y por criterio dinámico, así como las accuracies finales en test correspondientes. En esta

corrida, ambos enfoques seleccionaron la misma tasa de aprendizaje en los cinco conjuntos analizados, al tiempo que el criterio dinámico redujo el espacio de búsqueda entre 40 % y 60 %. Los desempeños finales en test resultaron muy próximos entre sí, lo que sugiere que la restricción dinámica permite acotar la exploración sin alejarse de configuraciones competitivas.

Sobre la base de esta comparación, se introdujo la métrica Hyperparameter Search Efficiency (HSE), definida como la relación entre la accuracy final alcanzada y el número de configuraciones consideradas durante el proceso de selección. Esta métrica permite resumir, en un único valor, la eficiencia exploratoria de cada enfoque dentro del protocolo adoptado.

TABLA 4. EFICIENCIA EXPLORATORIA DE BÚSQUEDA (HSE) EN MÚLTIPLES DATASETS

	dataset	grid_configs	dynamic_configs	grid_test_accuracy	dynamic_test_accuracy	grid_HSE	dynamic_HSE	HSE_gain_%
0	MNIST	10	6	0.9383	0.9354	0.0938	0.1559	66.1516
1	FashionMNIST	10	6	0.8292	0.8252	0.0829	0.1375	65.8627
2	Wine	10	6	0.9815	0.9815	0.0982	0.1636	66.6667
3	BreastCancer	10	4	0.9649	0.9708	0.0965	0.2427	151.5267
4	Digits	10	6	0.9444	0.9500	0.0944	0.1583	67.6549

FUENTE: elaboración propia 2026

La tabla 4 presenta los valores de la métrica Hyperparameter Search Efficiency (HSE) para la selección por rendimiento en validación y para el criterio dinámico en los distintos datasets evaluados. En todos los casos, el valor de HSE resulta mayor para el enfoque dinámico, lo que indica una mayor eficiencia exploratoria dentro del marco experimental considerado.

Este resultado se explica porque el criterio dinámico reduce sistemáticamente el número de configuraciones candidatas respecto de la búsqueda completa, mientras que las accuracies finales en test alcanzadas por ambos enfoques se mantienen muy próximas entre sí en los cinco datasets analizados. En consecuencia, la relación entre rendimiento final y tamaño de la región explorada favorece de manera consistente al enfoque dinámico.

En términos cuantitativos, la mejora relativa de eficiencia (HSE Gain) varía entre aproximadamente 65.9 % y 151.5 %, según el dataset. Los incrementos más elevados se observan en aquellos casos donde la reducción del espacio de búsqueda es más marcada, como BreastCancer, mientras que en MNIST, FashionMNIST, Wine y Digits también se registran mejoras sustantivas.

En conjunto, estos resultados sugieren que el criterio dinámico puede contribuir a acotar la exploración del espacio de hiperparámetros sin deteriorar de manera sustantiva el rendimiento predictivo final, al menos en los escenarios evaluados. Aunque esta evidencia debe interpretarse en términos exploratorios, refuerza la utilidad del análisis dinámico como herramienta complementaria para orientar la selección de hiperparámetros.

Validación multi-seed

Finalmente, con el objetivo de reforzar la robustez metodológica del análisis, se realizó una validación adicional repitiendo la evaluación multi-dataset con distintas semillas aleatorias. Esta etapa permite examinar si los resultados observados dependen de una única corrida o si, por el contrario, se mantienen con una consistencia razonable bajo distintas inicializaciones del proceso de entrenamiento. La validación multi-seed se aplicó sobre MNIST, FashionMNIST, Wine, BreastCancer y Digits, mientras que CIFAR-10 se reservó para la extensión de mayor complejidad con múltiples optimizadores. Los resultados resumidos y el detalle completo de las corridas por semilla se presentan en el material complementario del estudio, en particular en la Tabla 5 del cuaderno de Google Colab mencionado en la nota 1 del apartado de reproducibilidad experimental.

La validación multi-seed muestra que el comportamiento general del criterio dinámico se mantiene de forma razonablemente consistente al repetir los experimentos con distintas semillas aleatorias. En promedio, el enfoque dinámico reduce la cantidad de configuraciones candidatas respecto de la búsqueda completa en todos los datasets evaluados, con reducciones medias de 34 % en FashionMNIST, 30 % en Digits, 36 % en MNIST, 58 % en BreastCancer y 44 % en Wine. Estos resultados indican que la restricción dinámica permite acotar de manera apreciable la región de búsqueda aun bajo distintas inicializaciones aleatorias.

Al mismo tiempo, las accuracies medias de validación obtenidas por Grid Search y por el criterio dinámico resultan idénticas en MNIST, FashionMNIST y Digits, y muy próximas en BreastCancer y Wine. En cuanto al desempeño final en test, las diferencias entre ambos enfoques se mantienen en general acotadas, lo que sugiere que la reducción del espacio de búsqueda no implica, en la mayoría de los casos, una pérdida sustantiva de rendimiento predictivo.

Desde el punto de vista de la eficiencia exploratoria, los valores medios de HSE Gain son positivos en todos los datasets: aproximadamente 53.2 % en FashionMNIST, 44.9 % en Digits, 55.8 % en MNIST, 156.0 % en BreastCancer y 78.7 % en Wine. Esto indica que, dentro del marco experimental considerado, el criterio dinámico logra de manera sistemática una mejor relación entre rendimiento final y número de configuraciones evaluadas.

A partir de estos resultados, el análisis dinámico puede entenderse como una herramienta exploratoria útil para acotar la búsqueda de hiperparámetros manteniendo un rendimiento competitivo frente a la exploración exhaustiva.

Discusión

Los resultados obtenidos sugieren que la tasa de aprendizaje no solo afecta la rapidez de la optimización, sino también la sensibilidad relativa de las trayectorias de entrenamiento.

El desplazamiento del indicador empírico hacia valores menos negativos a medida que aumenta este hiperparámetro resulta consistente con una transición desde dinámicas más regulares hacia otras de mayor sensibilidad. Esta interpretación se alinea con la formulación del entrenamiento como sistema dinámico discreto y con lo señalado por Li et al. (2018) acerca de la relación entre paisaje de pérdida y comportamiento del entrenamiento.

Este patrón también dialoga con Ziyin et al. (2022), quienes destacan que ciertos procesos de entrenamiento pueden mostrar susceptibilidad frente a perturbaciones y cambios en los hiperparámetros. En el presente estudio, dicha susceptibilidad se observó como un gradiente de comportamientos, más que como una separación tajante entre estabilidad e inestabilidad. Las trayectorias temporales, las representaciones en el espacio de fases y las trayectorias en el espacio de estados resultaron consistentes con esa lectura.

Un hallazgo central del trabajo es que los mejores desempeños predictivos tienden a concentrarse en regiones intermedias del indicador. Esto sugiere que las configuraciones más favorables no se ubican necesariamente en los extremos del espacio explorado, sino en zonas donde la dinámica conserva regularidad sin volverse excesivamente rígida. En términos interpretativos, este resultado puede leerse a la luz de la noción de edge of chaos propuesta por Zhang et al. (2024), utilizada aquí solo como marco conceptual y no como demostración formal.

La evidencia también resulta compatible con la literatura sobre geometría del paisaje de pérdida. La relación entre curvatura local, magnitud de actualización y estabilidad ofrece un marco plausible para comprender por qué tasas de aprendizaje muy bajas favorecen trayectorias más regulares pero menos exploratorias, mientras que valores más altos incrementan la sensibilidad del proceso. En la misma línea, Fort y Scherlis (2019) señalan que la convergencia hacia regiones más planas puede asociarse con mejor generalización. Aunque el presente trabajo no evalúa directamente la planitud de los mínimos, los resultados sugieren una posible articulación entre trayectorias dinámicamente intermedias y desempeños competitivos.

Los experimentos con distinta profundidad y con múltiples optimizadores muestran, además, que la dinámica del entrenamiento no depende exclusivamente de la tasa de aprendizaje. La arquitectura, el optimizador y la estructura de los datos también intervienen en la forma en que evoluciona el sistema. Esta lectura resulta coherente con la perspectiva de Chaudhari y Soatto (2018), quienes entienden el entrenamiento como un proceso que combina componentes deterministas y estocásticos.

Desde el punto de vista metodológico, el estudio muestra que un indicador empírico de estabilidad puede utilizarse como apoyo para la selección de hiperparámetros. La comparación con la selección tradicional por rendimiento en validación sugiere que el criterio dinámico puede acotar la región de búsqueda sin deteriorar de manera sustantiva el desempeño final

en los escenarios analizados. En este sentido, el análisis dinámico no reemplaza estrategias clásicas como Grid Search, pero puede complementarlas.

El trabajo presenta, no obstante, algunas limitaciones. En primer lugar, el indicador utilizado es un proxy empírico basado en la evolución de la pérdida y no una estimación formal del exponente de Lyapunov; por ello, toda referencia al caos o al edge of chaos debe entenderse en sentido interpretativo. En segundo lugar, las arquitecturas consideradas fueron deliberadamente moderadas, lo que favorece el control experimental, pero limita la extrapolación inmediata a modelos de mayor escala. En tercer lugar, la banda utilizada para definir regiones dinámicas intermedias constituye una regla heurística de trabajo y no un umbral universal. Finalmente, aunque la evaluación multi-dataset y la validación multi-seed fortalecen la robustez del análisis, el alcance del estudio sigue siendo empírico y exploratorio.

Contrastación de hipótesis

En relación con la hipótesis 1, los resultados aportan evidencia favorable a la existencia de variaciones sistemáticas en la estabilidad dinámica del entrenamiento en función de la tasa de aprendizaje.

Respecto de la hipótesis 2, los hallazgos sugieren que el indicador empírico de estabilidad dinámica resulta útil para caracterizar la sensibilidad relativa del entrenamiento e identificar regiones relevantes dentro del espacio de hiperparámetros, aunque esta conclusión debe entenderse dentro de los límites del proxy utilizado.

En cuanto a la hipótesis 3, la evidencia respalda la idea de que regiones intermedias de estabilidad dinámica tienden a asociarse con desempeños predictivos competitivos y pueden contribuir a reducir la exploración del espacio de búsqueda.

En conjunto, la contrastación de hipótesis resulta favorable al enfoque propuesto.

Conclusiones

Este trabajo analizó la estabilidad dinámica del entrenamiento en redes neuronales profundas mediante un indicador empírico construido a partir de la evolución temporal de la pérdida. Los resultados muestran que la tasa de aprendizaje no solo afecta la rapidez de la optimización, sino también la sensibilidad relativa de las trayectorias, y que los desempeños más competitivos tienden a concentrarse en regiones intermedias del indicador.

En términos aplicados, la evidencia sugiere que el criterio dinámico puede contribuir a acotar la exploración del espacio de hiperparámetros sin deteriorar de forma sustantiva el rendimiento final en los escenarios analizados. De este modo, el estudio aporta evidencia empírica a favor del análisis dinámico como herramienta complementaria para interpretar el entrenamiento y orientar la selección de la tasa de aprendizaje.

A futuro, el enfoque podría extenderse a arquitecturas de mayor complejidad, a indicadores más refinados de sensibilidad dinámica y a esquemas automáticos de ajuste de hiperparámetros.

Referencias bibliográficas

- Bahri, Y., Kadmon, J., Pennington, J., Schoenholz, S. S., Sohl-Dickstein, J., & Ganguli, S. (2020). Statistical mechanics of deep learning. *Annual Review of Condensed Matter Physics*, 11, 501–528.
- Barrett, D. G. T., & Dherin, B. (2021). Implicit gradient regularization. In *International Conference on Learning Representations*.
- Chaudhari, P., & Soatto, S. (2018). Stochastic gradient descent performs variational inference, converges to limit cycles for deep networks. In *International Conference on Learning Representations*.
- Fort, S., & Scherlis, L. (2019). The Goldilocks zone: Towards better understanding of neural network loss landscapes. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1), 357–364.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Krizhevsky, A. (2009). *Learning multiple layers of features from tiny images* (Technical Report). University of Toronto.
- Li, H., Xu, Z., Taylor, G., Studer, C., & Goldstein, T. (2018). Visualizing the loss landscape of neural nets. In *Advances in Neural Information Processing Systems* (Vol. 31).
- Raghu, M., Gilmer, J., Yosinski, J., & Sohl-Dickstein, J. (2017). SVCCA: Singular vector canonical correlation analysis for deep learning dynamics and interpretability. In *Advances in Neural Information Processing Systems* (Vol. 30).
- Zhang, L., Feng, L., Chen, K., & Lai, C. H. (2024). Asymptotic edge of chaos as guiding principle for neural network training. *International Journal of Artificial Intelligence and Robotics Research*, 1(1), 2350001.
- Ziyin, L., Li, B., Simon, J. B., & Ueda, M. (2022). SGD with a constant large learning rate can converge to local maxima. In *International Conference on Learning Representations*.