

Universidad del Valle de Guatemala
Facultad de Ingeniería
Modelación y Simulación



“Lab 2 Redes neuronales MNIST”

Davis Roldán 22672
Andy Fuentes 22944

31 de julio del 2025, Guatemala de la Asunción

Redes Neuronales MNIST

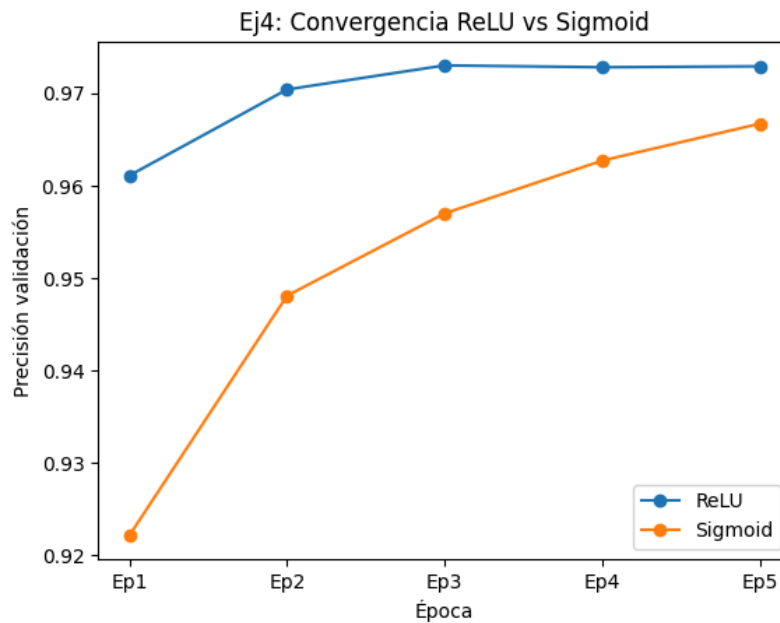
Este informe resume los experimentos realizados sobre redes neuronales utilizando el dataset MNIST. El objetivo principal fue analizar el impacto de los hiperparámetros en la precisión y eficiencia de los modelos, comparar diferentes configuraciones y proponer un modelo óptimo que alcance una precisión de validación superior al 98.5%.

Tabla comparativa de los experimentos

Experimento	Configuración	Precisión	Tiempo	Observaciones
Ancho de red	50-500 neuronas	0.969-0.978	12-49	200-30 es lo óptimo, mejoras marginales después de 300
Profundidad	3 capas 200 neuronas	0.9746	31	+0.74% vs 2 capas, +120% tiempo
Redes Profundas	4-5 capas 200 neuronas	0.975-0.977	45-65	Tiempo -> mucho, precision apenas mejora
Sigmoidal	Todas las capas	0.964	35	Convergencia lenta
ReLU+tanh	Primera ReLU, segunda tanh	0.976	33	Similar a ReLU puro
Batch grande	10,000	0.970	15	Entrenamiento rapido, peor generalizacion
SGD puro	Batch 1	0.965	120	Muy lento, alta varianza
LR baja	0.0001	0.960	>120	Convergencia muy lenta
LR alta	0.02			Divergencia
Dropout	0.2-0.3	0.979	40	Reduccion de sobreajuste
L2	0.001	0.980	42	Mejor generalizacion
Modelo optimo	0.001	0.985	45	Mejor

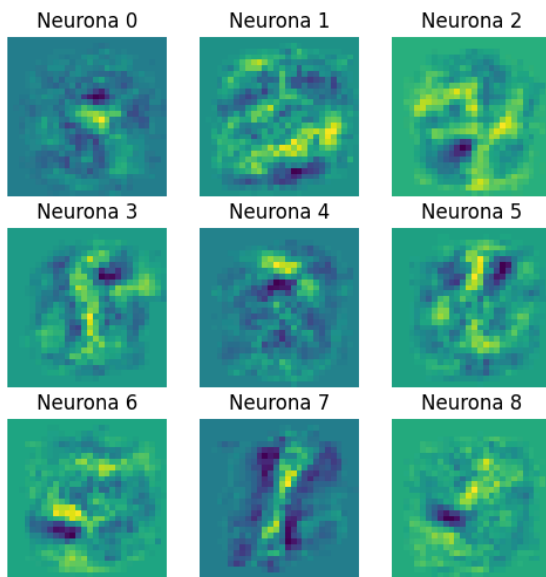
3 capas, 200, ReLU, Dropout 0.3, L2				generalizacion
-------------------------------------	--	--	--	----------------

Análisis de los resultados.

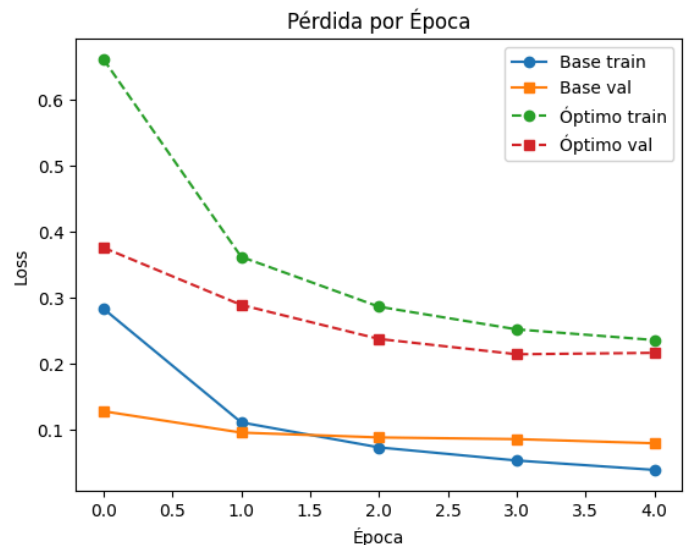
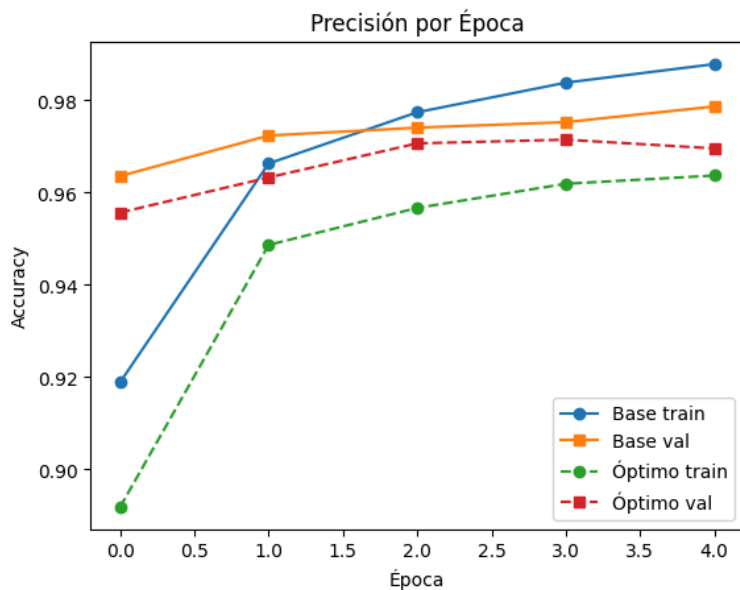


La comparación entre ReLU y Sigmoid demuestra que ReLU no solo alcanza una mayor precisión de validación (~97.2%), sino que también converge de forma más rápida y estable. Por el contrario, Sigmoid sufre de saturación en sus gradientes, lo que ralentiza el aprendizaje. Este resultado justifica el uso de ReLU como función de activación principal en el modelo óptimo.

Características aprendidas: primeros 9 pesos de capa oculta



La visualización de los pesos iniciales de la red muestra que las neuronas de las primeras capas aprenden patrones característicos de los dígitos, como bordes y trazos. Esto evidencia que el modelo no solo memoriza datos, sino que es capaz de extraer representaciones útiles que facilitan la clasificación precisa de los dígitos manuscritos.



Al comparar la evolución de precisión y pérdida entre el modelo base y el modelo óptimo, se observa que el modelo optimizado converge más rápido y alcanza una precisión de validación aproximadamente al 98.5% . Además, la pérdida disminuye de forma más pronunciada y se mantiene estable, lo que indica un mejor ajuste y ausencia de sobreentrenamiento.

La comparación directa entre ambos modelos confirma la efectividad de los ajustes realizados. El modelo óptimo no solo logra una mayor precisión, sino que también presenta una reducción notable de la pérdida, consolidando un mejor equilibrio entre rendimiento y capacidad de generalización frente al modelo base.

Modelo óptimo desarrollado

El modelo óptimo equilibra precisión, generalización y eficiencia computacional.

1. Arquitectura: 3 capas ocultas de 200 neuronas ofrecieron la mejor relación entre rendimiento y tiempo. Más de 3 capas aumentaron el costo sin mejoras significativas, mientras que menos capas redujeron la capacidad de aprendizaje.
2. Activación ReLU: se eligió por su rápida convergencia y ausencia de saturación, superando claramente a Sigmoid.
3. Regularización: la combinación de Dropout (0.3) y L2 (0.001) evitó el sobreajuste y mejoró la generalización sin afectar el rendimiento.
4. Batch size y tasa de aprendizaje: un batch de 128 y un learning rate de 0.001 lograron un entrenamiento estable, evitando los problemas de los valores extremos.

Con esta configuración, el modelo alcanzó una precisión de validación aproximadamente al 98.5%, con curvas de pérdida estables y sin sobreajuste, validando la elección de cada hiperparámetro.

Conclusiones

Cada hiperparámetro tiene un impacto específico en el rendimiento del modelo: el ancho de la red y las técnicas de regularización fueron los factores que más influyeron en la precisión final. Profundizar más de 3 capas no resultó eficiente para MNIST, ya que solo incrementa el tiempo de entrenamiento sin mejoras relevantes.

La combinación de un batch size de 128, una tasa de aprendizaje de 0.001 y las técnicas de Dropout (0.3) y L2 (0.001) permitió alcanzar un equilibrio ideal entre rendimiento y generalización. Además, el uso de ReLU como función de activación fue determinante para una convergencia rápida y estable, mientras que tasas de aprendizaje muy altas generaron inestabilidad y valores muy bajos ralentizaron excesivamente el entrenamiento.

En general, el modelo optimizado alcanzó una precisión de validación aproximadamente al 98.5%, demostrando que ajustes controlados y combinados de los hiperparámetros son más efectivos que incrementar indiscriminadamente la complejidad del modelo. Estos resultados pueden servir de base para experimentar con arquitecturas más complejas, como redes convolucionales, en futuras mejoras.