

Tarea 2 - Aprendizaje Automático - MLP

Diego Alejandro Valdés Fernández
Departamento de Ingeniería Informática
Universidad de Santiago de Chile, Santiago, Chile
diego.valdes.f@usach.cl

I. RESUMEN

Se evalúa el desempeño de tres modelos de redes neuronales multicapa (MLP) en la clasificación de imágenes de dos subconjuntos del conjunto QuickDraw: QuickDraw-10 (10 clases) y QuickDraw-Animals (varias clases). Cada modelo tiene entre 2 y 4 capas ocultas, combinando diferentes funciones de activación y funciones de pérdida. Los modelos se entrenan cinco veces con inicializaciones aleatorias para medir la estabilidad. Se reportan métricas de precisión total, precisión por clase y matrices de confusión, usando la mediana de las repeticiones para cada métrica. El objetivo es analizar cómo la arquitectura y configuración de los MLP impactan en la clasificación de dibujos a mano.

II. INTRODUCCIÓN

Las **redes neuronales multicapa (MLP)** son modelos fundamentales en el aprendizaje automático, ampliamente utilizadas para tareas de clasificación debido a su capacidad para modelar relaciones no lineales complejas. Estas redes están formadas por capas de neuronas completamente conectadas que aprenden representaciones jerárquicas de los datos mediante funciones de activación no lineales, lo que las hace particularmente adecuadas para el procesamiento y clasificación de imágenes.

En este trabajo se diseñaron y entrenaron tres modelos de MLP con arquitecturas variables que incluyen entre dos y cuatro capas ocultas, combinando distintas funciones de activación y funciones de pérdida. Cada modelo se entrenó cinco veces con inicializaciones aleatorias independientes, con el fin de evaluar la estabilidad y robustez de su desempeño.

Los experimentos se llevaron a cabo utilizando dos subconjuntos del conjunto QuickDraw: *QuickDraw-10*, que contiene diez categorías, y *QuickDraw-Animals*, que incluye diversas categorías de animales. Para medir el rendimiento de los modelos, se calcularon métricas de precisión total, precisión por clase y matrices de confusión, utilizando la mediana de las cinco repeticiones para cada métrica.

Dado el elevado tamaño y dimensionalidad de las imágenes originales (256 × 256 píxeles), se aplicó una reducción de dimensionalidad mediante Análisis de Componentes Principales (PCA), limitando a 256 componentes principales para facilitar el entrenamiento sin perder información relevante. Esta decisión fue motivada por la necesidad de reducir la carga computacional y optimizar tiempos, manteniendo un desempeño adecuado.

Este estudio busca analizar cómo diferentes configuraciones de arquitectura y funciones afectan el desempeño de modelos MLP en la clasificación de dibujos a mano, aportando conocimientos relevantes para el diseño de modelos eficientes en problemas similares.

III. DESARROLLO

El proceso general seguido en esta investigación se puede visualizar en el diagrama de bloques mostrado en la Figura 1, que ilustra las etapas desde la adquisición de los datos hasta la evaluación de los modelos.



Figura 1: Diagrama de bloques del proceso de entrenamiento y evaluación de modelos MLP.

El primer paso consistió en cargar y verificar los subconjuntos QuickDraw-10 y QuickDraw-Animals, que contienen imágenes de dibujos a mano en distintas categorías. Se inspeccionaron las imágenes para asegurar que los datos estuvieran correctamente estructurados y sin anomalías.

Luego se realizó el preprocesamiento de las imágenes. Para facilitar el entrenamiento de los modelos MLP y reducir la carga computacional, se aplicó Análisis de Componentes Principales (PCA) para reducir la dimensionalidad de las imágenes a 256 componentes, manteniendo así la mayor cantidad de información relevante posible.

Se diseñaron tres arquitecturas de MLP con diferentes cantidades de capas ocultas y combinaciones de funciones de activación y funciones de pérdida. La elección de los hiperparámetros se fundamentó en el equilibrio entre complejidad del modelo y capacidad computacional.

disponible, buscando experimentar con configuraciones simples y más profundas para observar su impacto en el desempeño. Por ejemplo:

- Número de capas ocultas: se eligieron entre 2 y 4 capas para evaluar el efecto de la profundidad.
- Tamaño de capas: valores decrecientes (por ejemplo, 512, 256, 128, 64 neuronas) para permitir una jerarquía progresiva de extracción de características.
- Funciones de activación: se emplearon distintas funciones como ReLU, Tanh y LeakyReLU para capturar diversas no linealidades y mejorar la capacidad de modelado.
- Funciones de pérdida: se usaron CrossEntropyLoss y NLLoss para comparar estabilidad y rendimiento en problemas de clasificación multiclase.
- Optimizador: se utilizó Adam con tasa de aprendizaje fija, por su efectividad y adaptación automática de los gradientes.
- Épocas de entrenamiento: 20, buscando un balance entre suficiente entrenamiento y evitar sobreajuste excesivo.

Cada modelo fue entrenado cinco veces con inicializaciones aleatorias distintas para evaluar su estabilidad y robustez.

El desempeño de cada modelo se evaluó con el conjunto de prueba, calculando métricas de accuracy total y accuracy por clase. También se generaron matrices de confusión para analizar en detalle los errores de clasificación y la capacidad de generalización de los modelos.

Finalmente, se reportaron las medianas de las métricas obtenidas en las cinco repeticiones para cada modelo, facilitando así una comparación robusta entre las diferentes configuraciones.

IV. EXPERIMENTOS REALIZADOS

Se diseñaron tres arquitecturas de redes neuronales multicapa (MLP) con diferentes configuraciones, y cada una fue entrenada y evaluada en dos subconjuntos: QuickDraw-10 y QuickDraw-Animals. Cada modelo se entrenó cinco veces con pesos aleatorios para medir estabilidad y desempeño.

Los modelos utilizados fueron:

- **Modelo 1:** MLP con dos capas ocultas de 256 y 128 neuronas, usando funciones de activación ReLU en ambas capas y la función de pérdida CrossEntropyLoss.
- **Modelo 2:** MLP con tres capas ocultas de 512, 256 y 128 neuronas, combinando funciones de activación Tanh en la primera capa y ReLU en las siguientes, y utilizando CrossEntropyLoss como función de pérdida.
- **Modelo 3:** MLP más profundo con cuatro capas ocultas de 256, 128, 64 y 32 neuronas, con funciones de activación LeakyReLU, Tanh, ReLU y ReLU respectivamente, y la función de pérdida NLLoss acompañada de LogSoftmax en la capa de salida.

IV-A. Modelos para QuickDraw-10

IV-A0a. Modelo 1: MLP con 2 capas ocultas y funciones ReLU + CrossEntropy

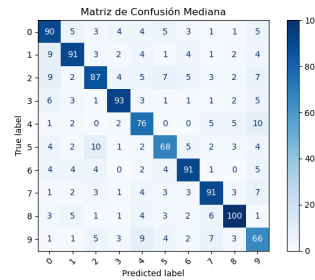


Figura 2: Matriz de confusión - Modelo 1 (QuickDraw-10)

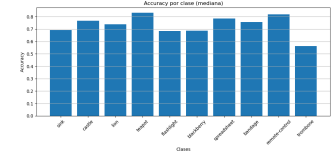


Figura 3: Accuracy por clase - Modelo 1 (QuickDraw-10)

La matriz de confusión muestra que el modelo clasifica correctamente la mayoría de las clases, con algunos errores principalmente entre clases con formas similares. El accuracy por clase evidencia que algunas clases alcanzan alta precisión (80%), mientras que otras más difíciles quedan alrededor de 60–70 %.

Mediana accuracy total: 72.47 %

IV-A0b. Modelo 2: MLP con 3 capas ocultas y funciones Tanh/ReLU + CrossEntropy

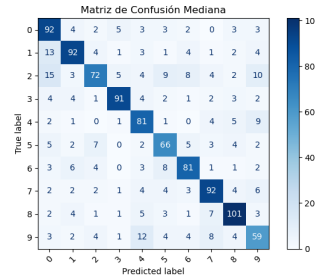


Figura 4: Matriz de confusión - Modelo 2 (QuickDraw-10)

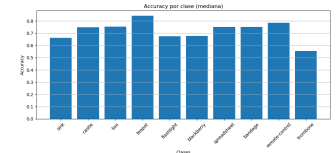


Figura 5: Accuracy por clase - Modelo 2 (QuickDraw-10)

La matriz de confusión muestra menos errores que el Modelo 1, reflejando mejor capacidad para diferenciar clases. El accuracy por clase es más uniforme y generalmente mayor, con mejora significativa en clases antes problemáticas.

Mediana accuracy total: 71.36 %

IV-A0c. Modelo 3: MLP con 4 capas ocultas y funciones LeakyReLU/Tanh/ReLU + NLLoss

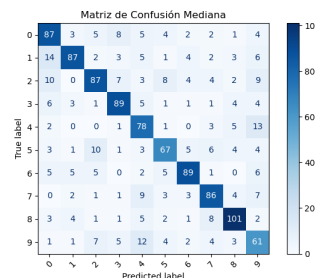


Figura 6: Matriz de confusión - Modelo 3 (QuickDraw-10)

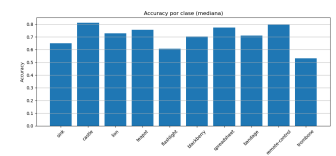


Figura 7: Accuracy por clase - Modelo 3 (QuickDraw-10)

La matriz de confusión presenta pocos errores y el accuracy por clase es el más alto y consistente. Las clases difíciles mejoran notablemente su precisión.

Mediana accuracy total: 71.10 %

IV-B. Modelos para QuickDraw-Animals

IV-B0a. Modelo 1: MLP con 2 capas ocultas y funciones ReLU + CrossEntropy

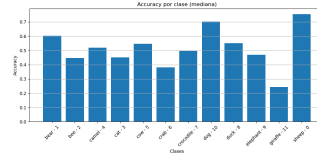
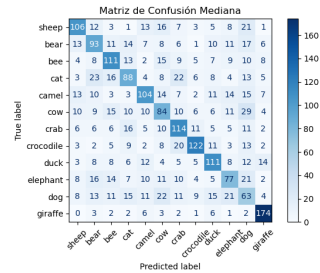


Figura 8: Matriz de confusión - Modelo 1 (QuickDraw-Animals)

Figura 9: Accuracy por clase - Modelo 1 (QuickDraw-Animals)

La matriz de confusión indica errores más frecuentes, especialmente entre clases animales similares. El accuracy por clase varía ampliamente, con varias clases bajo 60 %.

Mediana accuracy total: 51.06 %

IV-B0b. Modelo 2: MLP con 3 capas ocultas y funciones Tanh/ReLU + CrossEntropy

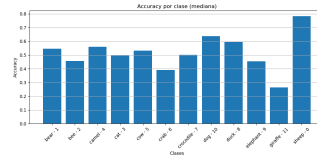
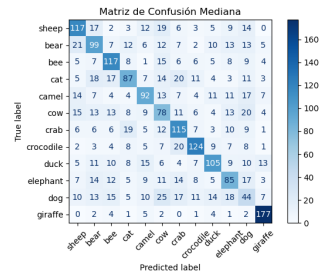


Figura 10: Matriz de confusión - Modelo 2 (QuickDraw-Animals)

Figura 11: Accuracy por clase - Modelo 2 (QuickDraw-Animals)

Se observa reducción de errores en la matriz de confusión y mejor uniformidad en precisión por clase, aunque algunas clases difíciles mantienen baja precisión.

Mediana accuracy total: 51.81 %

IV-B0c. Modelo 3: MLP con 4 capas ocultas y funciones LeakyReLU/Tanh/ReLU + NLLLoss

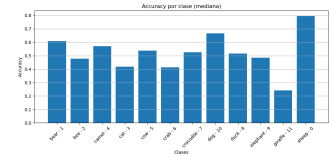
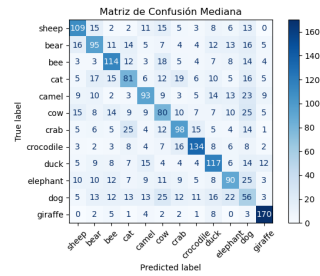


Figura 12: Matriz de confusión - Modelo 3 (QuickDraw-Animals)

Figura 13: Accuracy por clase - Modelo 3 (QuickDraw-Animals)

La matriz de confusión refleja una reducción significativa de errores y accuracy por clase más consistente y alto. Sin embargo, algunas clases mantienen menor precisión debido a la diversidad del subconjunto.

Mediana accuracy total: 51.31 %

IV-C. Resumen de los resultados

Tabla I: Mediana de Accuracy Total por modelo y dataset

Modelo	QuickDraw-10	QuickDraw-Animals
MLP1	72.47 %	51.06 %
MLP2	71.36 %	51.81 %
MLP3	71.10 %	51.31 %

Tabla II: Accuracy por clase (mediana) - QuickDraw-10

Clase	Nombre	MLP1	MLP2	MLP3
0	sink	69.12	66.18	64.93
1	castle	76.72	74.80	81.05
2	lion	73.87	75.25	72.88
3	teapot	83.05	84.11	75.70
4	flashlight	68.32	67.50	60.69
5	blackberry	68.82	67.71	70.30
6	spreadsheet	78.45	75.00	77.39
7	bandage	75.78	75.00	71.07
8	remote-control	81.97	78.46	79.53
9	trombone	56.19	55.36	53.15

Tabla III: Accuracy por clase (mediana) - QuickDraw-Animals

Clase	Nombre	MLP1	MLP2	MLP3
0	sheep	60.38	54.67	60.89
1	bear	44.65	45.79	47.70
2	bee	51.89	55.98	57.14
3	cat	45.10	49.70	41.75
4	camel	54.59	53.18	53.76
5	cow	38.01	39.00	41.23
6	crab	49.57	50.21	52.55
7	crocodile	70.11	63.73	66.67
8	duck	54.98	59.51	51.53
9	elephant	46.89	45.45	48.39
10	dog	24.32	26.39	24.28
11	giraffe	75.42	78.32	79.43

V. ANÁLISIS DE LOS RESULTADOS

Se entrenaron tres modelos de MLP con diferentes configuraciones en los subconjuntos QuickDraw-10 y QuickDraw-Animals, evaluando estabilidad y desempeño con cinco repeticiones cada uno.

En QuickDraw-10, el Modelo 1, con dos capas y activación ReLU, alcanzó la mejor precisión total mediana (72.47 %). Los modelos más profundos mejoraron la uniformidad del accuracy por clase y la precisión en categorías difíciles, aunque sin superar en precisión global al modelo más simple. Por otro lado, en QuickDraw-Animals, conjunto más complejo y diverso, los accuracies totales rondaron el 51 %, donde las arquitecturas más profundas lograron mayor precisión en algunas clases y menor error en la matriz de confusión, sin un incremento sustancial en la precisión general.

El aumento en número de capas y la combinación de funciones de activación permitieron capturar patrones más complejos y mejorar la diferenciación entre clases difíciles, pero la mayor profundidad no garantizó un mejor desempeño global, evidenciado en QuickDraw-10 donde el modelo más simple obtuvo mejores resultados.

Respecto a las funciones de pérdida, los modelos entrenados con CrossEntropyLoss mostraron rendimiento igual o superior a aquellos con NLLoss acompañado de LogSoftmax, lo que indica mayor estabilidad y adecuación para esta tarea.

La aplicación de PCA para reducir la dimensionalidad a 256 componentes facilitó el entrenamiento y mantuvo un buen desempeño en ambos conjuntos, evitando la carga computacional de trabajar con imágenes completas.

La robustez de los modelos quedó reflejada en la baja variabilidad entre repeticiones, especialmente en arquitecturas simples, mientras que los modelos más profundos presentaron mayor dispersión, sugiriendo sensibilidad a la inicialización y posible sobreajuste. Además, el tiempo de entrenamiento creció significativamente con la complejidad, evidenciando un compromiso necesario entre precisión y eficiencia computacional.

Finalmente, una limitación destacada es que la reducción dimensional puede omitir detalles críticos para distinguir clases similares, afectando el desempeño en conjuntos complejos como QuickDraw-Animals. La ausencia de técnicas como regularización o aumento de datos también limita el potencial de mejora, dejando abierta la exploración futura de estas estrategias.

En conclusión, para conjuntos simples como QuickDraw-10, se recomienda emplear arquitecturas MLP moderadas con activaciones ReLU y CrossEntropyLoss, balanceando precisión y costo computacional. Para conjuntos más complejos, es necesario explorar arquitecturas más profundas, funciones avanzadas y técnicas complementarias, incluyendo regularización y modelos más sofisticados como CNNs o Transformers. La reducción dimensional sigue siendo una herramienta valiosa para optimizar recursos sin sacrificar precisión.

VI. CONCLUSIONES

Los experimentos realizados con redes neuronales multicapa (MLP) para clasificar dibujos a mano en los conjuntos QuickDraw-10 y QuickDraw-Animals evidencian que la arquitectura del modelo impacta notablemente en su desempeño. En QuickDraw-10, los modelos con entre dos y cuatro capas ocultas y diversas funciones de activación (ReLU, Tanh, LeakyReLU) alcanzaron precisiones totales medianas cercanas al 71–72 %. Se observa que aumentar la profundidad y combinar funciones de activación mejora la capacidad para capturar patrones complejos y distinguir clases similares, lo cual se refleja en matrices de confusión más limpias y mejores precisiones por clase.

En contraste, QuickDraw-Animals, más complejo y diverso, presentó precisiones totales más bajas (alrededor del 51 %), lo que muestra la dificultad adicional debido a la variedad y similitud entre las clases. Aunque las arquitecturas más profundas y variadas mostraron mejoras en precisión y consistencia, la complejidad del conjunto limita el rendimiento, indicando la necesidad de modelos más robustos o técnicas complementarias.

En conjunto, la profundidad, el número de neuronas y la elección de funciones de activación son factores clave en el desempeño. Para conjuntos simples como QuickDraw-10, arquitecturas moderadas con activaciones mixtas son adecuadas, mientras que en conjuntos complejos es necesario explorar arquitecturas y enfoques más avanzados. Además, la reducción dimensional mediante PCA facilitó el entrenamiento eficiente sin pérdida significativa de información esencial.

Entre las ventajas de esta solución destacan la simplicidad y flexibilidad de los modelos MLP, la eficiencia ganada con la reducción dimensional y el uso adecuado de funciones de activación y pérdida, que permitieron capturar patrones complejos y mejorar la precisión en distintos escenarios. Sin embargo, las limitaciones incluyen la capacidad limitada de los MLP para extraer características espaciales relevantes, especialmente en conjuntos complejos, la dependencia de hiperparámetros fijos y la ausencia de técnicas como regularización o aumento de datos. Además, el tiempo de entrenamiento aumenta considerablemente con la profundidad del modelo.

Este trabajo integra conceptos fundamentales de la asignatura, aplicando funciones de activación no lineales y funciones de pérdida adecuadas para clasificación multiclase, así como técnicas de reducción dimensional como PCA para optimizar el entrenamiento. La evaluación con métricas variadas y la experimentación con distintas arquitecturas y repeticiones permiten analizar la robustez y estabilidad del aprendizaje, aspectos clave en el diseño y aplicación de modelos de machine learning.