

Reporte de laboratorio: clasificación tabular con MLP

Nombre: Alejandro Lembke Barrientos

Fecha: 25 de agosto de 2026

Repositorio o archivo entregado:

<https://gitea.p-lao.com/aleleba/Procesamiento-de-Imagenes-y-Vision-Par-Computadora>

Propósito. Documentar una comparación reproducible entre un MLP convencional y un MLP residual con Batch Normalization, usando una división estratificada 800/200 y un conjunto de entrenamiento desbalanceado. Los valores de este reporte provienen de la ejecución real de `notebooks/laboratorio_estudiante.ipynb`.

1. Datos, división y control de fuga

Elemento	Valor	Evidencia o justificación	Estado
Archivo público	train_1000_desbalanceado.csv	1,000 ejemplos	Real
División	800 train / 200 dev	Estratificada, semilla 31	Real
Clases en train / dev	600/200 clase 0 y 1 en train; 150/50 en dev	Proporción 75/25 conservada en ambas particiones	Real
Estandarización	Media y desviación de train	Sin usar dev ni test	Real
Test privado	No usado durante selección	Sólo evaluación final (a cargo del instructor)	Real

Respuesta. Describa una fuente posible de fuga de información y cómo la evitó:

La fuente de fuga más directa habría sido ajustar la media y la desviación estándar del estandarizador (`ajustar_estandarizador`) usando todo el conjunto de 1,000 ejemplos antes de dividir en train/dev, o recalcularlas después de ver las métricas de desarrollo. Se evitó ajustando el estandarizador exclusivamente con `x[indice_train]` (800 ejemplos) y aplicando esa misma transformación a dev y, al final, al artefacto exportado; el umbral y la arquitectura también se eligieron mirando solo el costo en desarrollo, sin usar en ningún momento `test_200_balanceado.csv`.

2. Estrategia ante el desbalance

Decisión	Valor elegido	Justificación
Pérdida	BCE ponderada; peso positivo = 3.0	Compensa la relación 600/200 de train.
Selección de modelo	Costo esperado en dev	Falso negativo cuesta 5; falso positivo cuesta 1.
Métricas reportadas	Precision, recall, F1, especificidad y accuracy	Accuracy sola oculta el desempeño de la clase minoritaria.
Umbral	Se elige sólo en dev	Se congela antes del test privado.

Respuesta. Declare los costos y explique por qué el umbral 0.5 puede no ser adecuado:

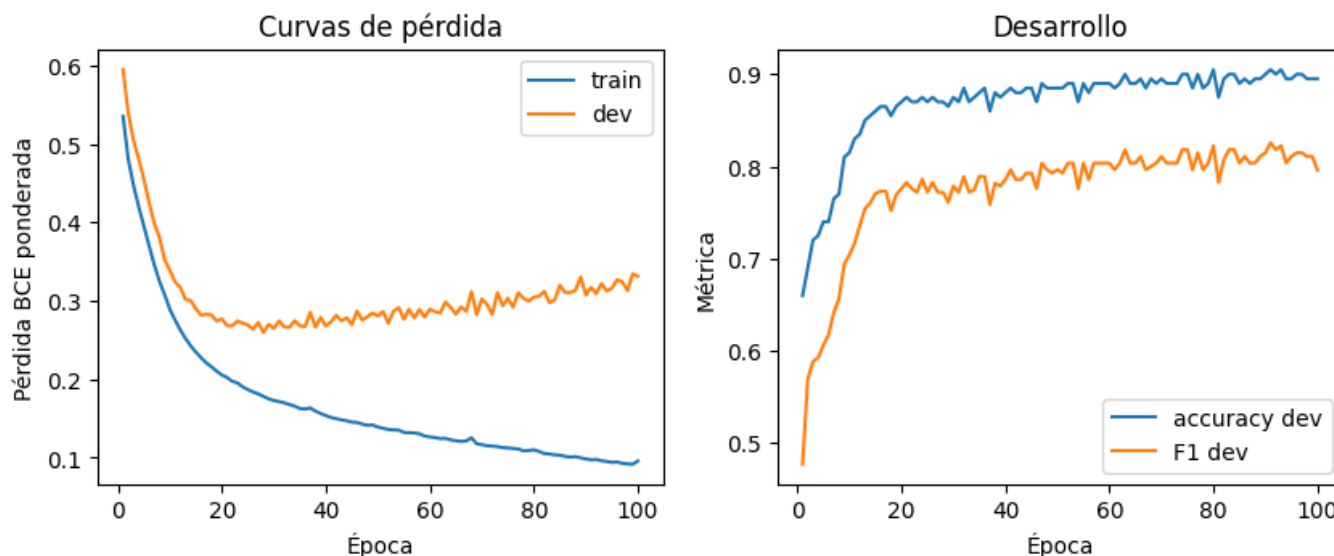
Se asumió costo de falso negativo = 5 y costo de falso positivo = 1 (dejar pasar un caso positivo real es cinco veces más costoso que una falsa alarma). Con esos costos, el umbral que minimiza el costo esperado en desarrollo resultó ser 0.315, por debajo de 0.5: un umbral de 0.5 favorece la precisión pero deja escapar más falsos negativos, que en este problema son mucho más caros que los falsos positivos adicionales que introduce un umbral más bajo.

3. Búsqueda de hiperparámetros

ID	Arquitectura	Capas/ancho	LR	Épocas	Costo dev	F1 dev	Recall dev	Umbral
A	MLP	[24,16]	0.003	100	43	0.715	0.980	0.050
B	MLP	[48,24]	0.002	100	43	0.796	0.900	0.315
C	Residual + BN	24	0.002	100	55	0.686	0.940	0.060
D	Residual + BN	40	0.0015	100	56	0.804	0.820	0.400

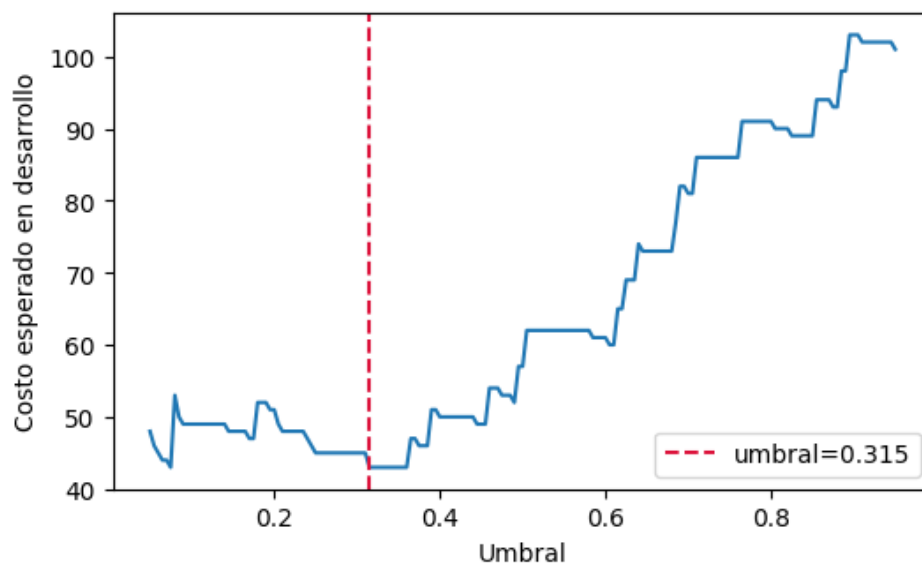
Decisión. Se elige la configuración de menor costo esperado en desarrollo; A y B empataron en costo (43), por lo que se desempató por mayor F1 en desarrollo, lo que favoreció a B (MLP, capas [48,24], learning rate 0.002, 100 épocas, umbral 0.315).

4. Curvas de entrenamiento y desarrollo



Interpretación. Indique la época o checkpoint seleccionado y explique si observa sobreajuste, subajuste o estabilidad: Se entrenaron las 100 épocas completas y se usó el modelo final (época 100). La pérdida de train baja de forma monótona hasta $\approx 0,09$, mientras que la pérdida de dev baja rápido hasta $\approx 0,27$ (época ≈ 20) y luego se mantiene estable con una leve subida hacia el final ($\approx 0,30$ – $0,33$ después de la época 70), señal de un sobreajuste leve. Accuracy y F1 en dev crecen rápido y se estabilizan alrededor de la época 20–30 (accuracy $\approx 0,88$ – $0,91$, F1 $\approx 0,78$ – $0,82$), con ruido pero sin caída sostenida, por lo que la configuración es razonablemente estable y no sería indispensable un checkpoint anterior.

5. Selección de umbral y reporte de métricas



Conjunto	Accuracy	Precision	Recall	F1	Especificidad
Desarrollo, umbral 0.315	0.885	0.714	0.900	0.796	0.880
Test privado, umbral congelado	_____	_____	_____	_____	_____

La fila de test privado se deja en blanco a propósito: `instructor_privado/test_200_balanceado.csv` nunca se usa ni se lee durante el desarrollo del laboratorio; esa evaluación final es exclusiva del instructor.

Matriz de confusión en desarrollo (umbral 0.315).

	Pred. 0	Pred. 1
Real 0	132	18
Real 1	5	45

6. Conclusión reproducible

1. Arquitectura final, pesos y ruta del artefacto: MLP con capas ocultas [48, 24], entrada de dimensión 4; pesos guardados en `entrega/modelo_elegido.npz`.
2. Umbral congelado y costos asumidos: umbral 0.315, elegido minimizando el costo esperado en desarrollo con costo de falso negativo = 5 y costo de falso positivo = 1.
3. Cambio que haría antes de desplegar: repetir la búsqueda de hiperparámetros con más de una semilla por configuración para verificar que la elección de arquitectura y umbral no dependa de una sola inicialización, y monitorear en producción la distribución de `x1..x4` para detectar drift respecto a los datos de entrenamiento.