

L U M B R E



Inicialización y normalización

Grado: 7

Materia: Redes Neuronales Artificiales · Inteligencia Artificial

11/10/2026

Antes de empezar

Antes de la primera actualización, el entrenamiento se juega en cómo arrancan los pesos y cómo se escalan las activaciones. Una buena inicialización y un lote normalizado evitan que la señal se apague o explote desde el inicio.

QUÉ APRENDERÁS

- Aplicar inicialización de Xavier y He y normalización por lotes para mantener la varianza de activaciones y gradientes y estabilizar el entrenamiento profundo.

Inicialización y normalización

Empezar bien es medio entrenar

Poner los pesos a cero deja todas las neuronas idénticas y simétricas; ponerlos muy grandes hace explotar la señal. La inicialización correcta rompe la simetría y preserva la varianza, y la normalización la cuida durante toda la marcha.

VIDEO

Batch normalization: qué es y para qué sirve

Vídeo en español que explica cómo la normalización por lotes estabiliza y acelera el entrenamiento.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=qefch39DHoM>

Por qué importa el arranque

Si la varianza de la activación se mantiene constante capa a capa, la señal ni se desvanece ni revienta. Una inicialización pensada para ello evita el desaparecimiento y la explosión antes de la primera pasada.

Ni ceros ni demasiado grandes

- Cero iguala a todas las neuronas: ninguna aprende distinto, la simetría no se rompe.
- Valores enormes saturan las activaciones y disparan el gradiente.
- Aleatoriedad calibrada rompe la simetría con la escala justa.

Xavier o Glorot

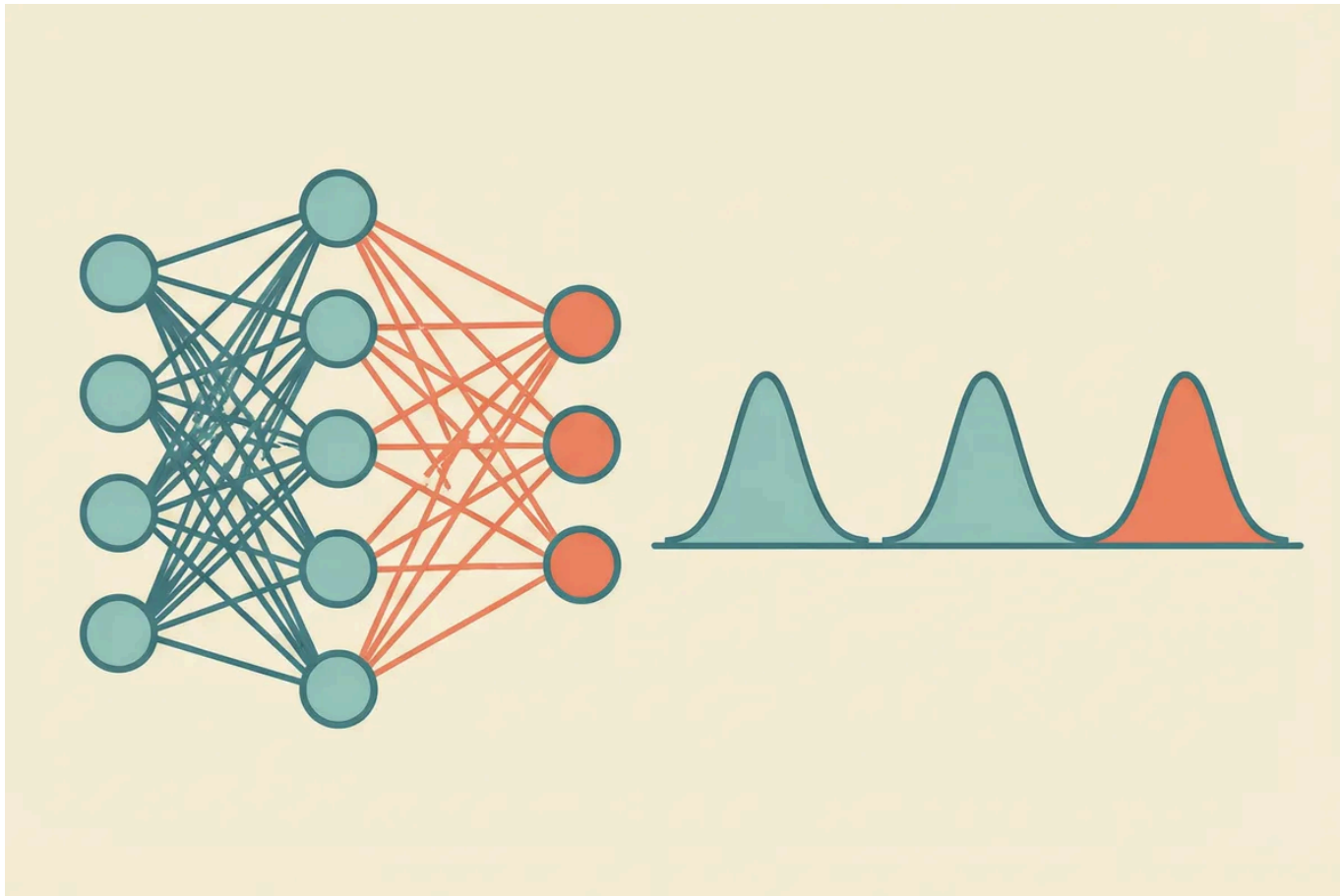
Para activaciones simétricas como sigmoide o tanh, Xavier fija la varianza de los pesos en función del número de entradas y salidas de la capa, repartiendo señal equilibrada hacia delante y hacia atrás.

He para ReLU

Como ReLU anula la mitad de la señal, He duplica la varianza respecto a Xavier para compensar esa pérdida. Usar la inicialización acorde a la activación mantiene la varianza bajo control.

Normalización por lotes

Batch norm estandariza las entradas de cada capa dentro del lote —media cero, varianza uno— y luego las reajusta con parámetros aprendibles. Estabiliza la distribución interna y acelera el entrenamiento.



Una buena inicialización reparte la varianza y la normalización mantiene las activaciones centradas durante el entrenamiento.

Efectos de la normalización

Al domesticar la escala de las activaciones, permite tasas de aprendizaje mayores, suaviza el paisaje de error y aporta un poco de regularización por el ruido del lote. También exige tratar bien el modo de inferencia.

Inferencia frente a entrenamiento

En inferencia no hay lote que promediar, así que batch norm usa estadísticas acumuladas del entrenamiento. Olvidar cambiar de modo corrompe las predicciones: la media del lote no existe al servir.

Errores frecuentes

- Inicializar los pesos en cero o con valores idénticos, dejando las neuronas simétricas e incapaces de especializarse.
- Usar Xavier con ReLU o He con tanh, descalibrando la varianza respecto a la activación de la capa.
- Olvidar el modo de inferencia en batch norm y servir con estadísticas del lote en vez de las acumuladas.

Ejemplo resuelto: por qué se dispara la señal al inicio

Una red profunda arranca con pesos demasiado grandes y la pérdida revienta en las primeras épocas. Revisamos el arranque:

1. Comprobar que los pesos iniciales tienen una varianza excesiva para el tamaño de la capa.
2. Recordar que la varianza de la activación se multiplica capa a capa y explota.
3. Aplicar Xavier o He según la activación para fijar la varianza correcta.
4. Añadir batch norm para recentrar cada capa dentro del lote.
5. Subir la tasa de aprendizaje con confianza y ver que la señal deja de dispararse.

¿Para qué sirve en la realidad?

Inicializar y normalizar bien es la diferencia entre un entrenamiento que despegua y uno que se apaga o explota antes de la primera época: cuesta poco y ahorra horas de ajuste.

1. Inicializar todos los pesos en cero es problemático porque:

- ☐ los gradientes se vuelven infinitos
- ☐ las neuronas de una capa quedan simétricas y aprenden igual
- ☐ la red no puede tener activaciones lineales
- ☐ el sesgo desaparece

2. La inicialización de He está calibrada para activaciones ReLU, que recortan la mitad de la señal.

- ☐ Verdadero
- ☐ Falso

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. Inicialización



4. Ordena el razonamiento para inicializar y normalizar una red profunda:

Elementos:

- mirar qué activación usará la capa
- fijar la varianza de pesos acorde a ella
- añadir normalización por lotes si conviene
- gestionar estadísticas al pasar a inferencia

Activacion	. . .
Escala	. . .
Normalizar	. . .
Modo	. . .

Normalización por lotes https://en.wikipedia.org/wiki/Batch_normalization Mecanismo de batch norm y su efecto sobre el entrenamiento.

5. Une cada técnica con lo que resuelve.

5.	rompe la simetría con la varianza justa para sigmoide y tanh	-----
5.	duplica la varianza para compensar que ReLU recorta negativos	-----
5.	estandariza las activaciones de cada capa dentro del lote	-----
5.	usa estadísticas acumuladas al servir el modelo	-----

Opciones: xavier · he · batch norm · modo inferencia

Empezar bien es medio entrenar

Inicialización y normalización

Arranque

ni ceros ni enormes

romper simetría

Xavier y He

según la activación

preservan varianza

Batch norm

centra por capa

permite tasa mayor

Inferencia

estadísticas acumuladas

cambiar de modo

El mapa mental interactivo es interactivo en la versión online.



Inicialización de pesos https://en.wikipedia.org/wiki/Weight_initialization Estrategias de inicialización y su efecto sobre la varianza.

6. Cambias de tanh a ReLU en una red profunda y el entrenamiento se descontrola. ¿Qué dos ajustes de inicialización y normalización tocarías primero y por qué?

Referencia curricular

Grado:	7
Materia:	Redes Neuronales Artificiales · Inteligencia Artificial

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-redn-006-inicializacion-y-normalizacion>

Material educativo generado desde Lumbre · lumbre.lat