

L U M B R E



Retropropagación

Grado: 7

Materia: Redes Neuronales Artificiales · Inteligencia Artificial

11/10/2026

Antes de empezar

Sin retropropagación, una red profunda sería inentrenable: no sabríamos qué peso mover y cuánto. Encadenando derivadas de atrás hacia delante, atribuye a cada peso su cuota de culpa en el error final.

QUÉ APRENDERÁS

- Explicar cómo la regla de la cadena distribuye el error de la salida hacia atrás para calcular el gradiente de cada peso, habilitando el descenso por gradiente en redes profundas.

Retropropagación

El error viaja hacia atrás

La red avanza hacia delante calculando salidas; el aprendizaje viaja hacia atrás. La retropropagación toma el error en la salida y lo reparte capa a capa, preguntando a cada derivada cuánto contribuyó.

VIDEO

¿Cómo funciona realmente la retropropagación?

Vídeo en español de 3Blue1Brown que muestra cómo el gradiente viaja hacia atrás por la red.



Escanea para verlo en Lumbre

https://www.youtube.com/watch?v=ScVpPS_CFYc

Pase hacia delante

Primero se propaga la entrada por las capas hasta la predicción y se mide la pérdida. Ese valor es la semilla de todo: sin pérdida no hay nada que retropropagar.

Gradiente de la pérdida

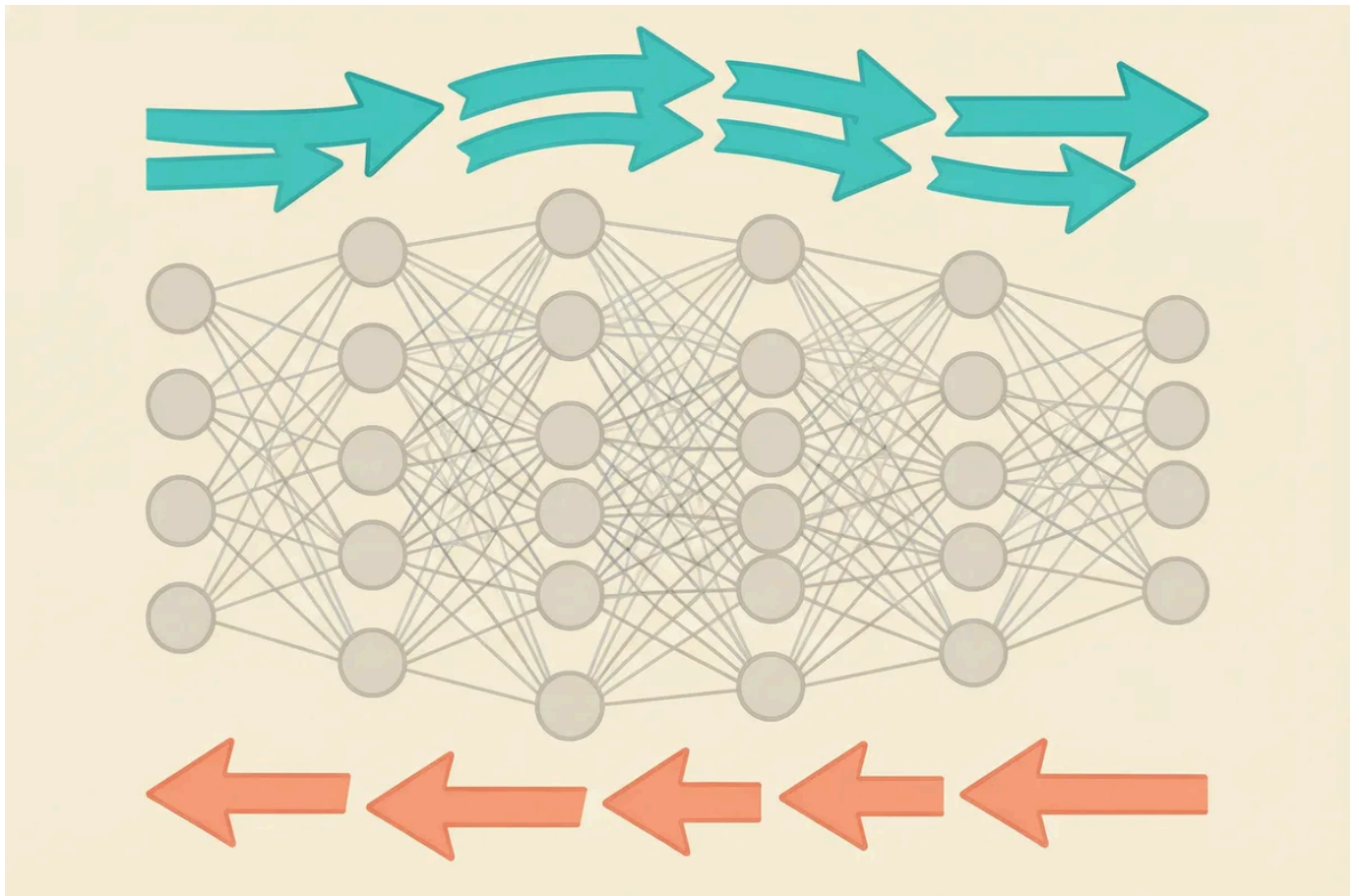
Queremos saber cómo cambia la pérdida al mover un peso. Ese cambio es la derivada, y mover el peso en sentido opuesto a su derivada reduce la pérdida: es el descenso por gradiente.

Regla de la cadena

- La derivada de una composición es el producto de las derivadas de cada eslabón.
- Cada capa deriva su transformación y multiplica lo que le llega de la siguiente.
- Así la señal de error se encadena desde la salida hasta la primera capa.

Propagación hacia atrás

Se calculan primero los gradientes de la capa de salida y se van transmitiendo hacia atrás, reutilizando lo ya derivado. Una sola pasada hacia atrás entrega el gradiente de todos los pesos, barato comparado con estimarlos a finite differences.



La señal avanza para predecir y el error retrocede capa a capa para asignar la culpa a cada peso.

Reutilizar el cálculo

El truco de eficiencia es que las derivadas intermedias se comparten: lo calculado en una capa alimenta a la anterior. Evita recalcular desde cero por cada peso, lo que haría la red prohibida.

Conexión con el desaparecimiento

Como el gradiente es un producto de muchas derivadas, si cada una es menor que uno el producto se apaga hacia atrás; si son mayores, explota. Esa es la raíz del desaparecimiento y la explosión del gradiente.

Autodiferenciación

Los frameworks construyen un grafo de operaciones y aplican la cadena automática, bien en modo cinta bien por compilación. El programador define el pase hacia delante; la máquina deriva sola hacia atrás.

Errores frecuentes

- Creer que la retropropagación aprende por ensayo y error perturbando pesos, cuando deriva de forma exacta con la cadena.
- Confundir el pase hacia delante con el entrenamiento: la red primero predice, y solo después se retropropaga.
- Olvidar que el gradiente es un producto de derivadas y que por eso puede apagarse o explotar con la profundidad.

Ejemplo resuelto: gradiente de un peso en la primera capa

Queremos saber cuánto mover un peso de la primera capa para reducir la pérdida final. Rastreamos su influencia hacia atrás:

1. Avanzar la entrada por la red hasta la predicción y medir la pérdida.
2. Derivar la pérdida respecto a la salida de la última capa.
3. Multiplicar por la derivada de cada capa intermedia, de atrás hacia delante.
4. Encadenar hasta llegar al peso de la primera capa.
5. Mover el peso en sentido contrario a ese gradiente acumulado.

¿Para qué sirve en la realidad?

Sin retropropagación no habría redes profundas: es lo que vuelve barato calcular, en una sola pasada, hacia dónde mover millones de pesos a la vez.

1. La retropropagación calcula el gradiente de todos los pesos aplicando principalmente:

- ☐ un barrido aleatorio de pesos
- ☐ la regla de la cadena propagando el error hacia atrás
- ☐ un ajuste a ciegas por ensayo y error
- ☐ la derivada de solo la capa de salida

2. Retropropagar el gradiente es más eficiente que estimarlo perturbando cada peso uno a uno.

- ☐ Verdadero
- ☐ Falso

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. Backprop



4. Ordena las fases de un paso de entrenamiento con retropropagación:

Elementos:

- propagar la entrada hasta la predicción
- medir la pérdida contra el objetivo
- derivadas por la cadena capa a capa atrás
- mover los pesos en sentido opuesto al gradiente

Delante	. . .
Perdida	. . .
Atras	. . .
Actualizar	. . .

Retropropagación <https://en.wikipedia.org/wiki/Backpropagation> Derivación y mecánica del algoritmo de retropropagación del gradiente.

5. Une cada término con su papel en el aprendizaje.

5.	medida de cuánto falla la predicción	-----
5.	derivada que dice cuánto mueve un peso la pérdida	-----
5.	regla que deriva composiciones multiplicando	-----
5.	dirección en que viaja la señal de error	-----

Opciones: pérdida · gradiente · cadena · atrás

El error viaja hacia atrás

Retropropagación

Hacia delante

propaga entrada

mide pérdida

Regla de cadena

producto de derivadas

culpa por peso

Hacia atrás

de salida a entrada

reutiliza cálculos

Riesgos

gradiente se apaga

o explota

El mapa mental interactivo es interactivo en la versión online.



Retropropagación en la Wikipedia <https://es.wikipedia.org/wiki/Retropropagaci%C3%B3n>
Formulación del algoritmo y su relación con la regla de la cadena.

6. Tu red profunda deja de mejorar y las primeras capas casi no reciben gradiente. Desde la óptica del encadenamiento de derivadas, ¿qué fenómeno sospechas y qué dos remedios probarías?

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-004-retropropagacion>

Material educativo generado desde Lumbre · lumbre.lat