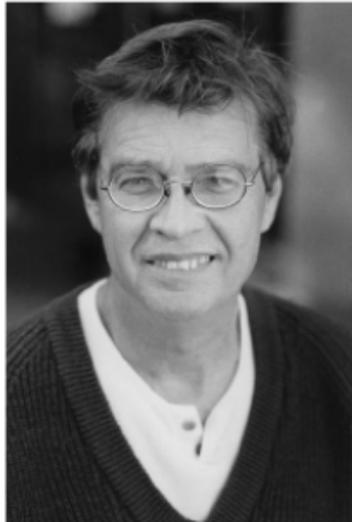


Lección 7

BACKPROPAGATION

Backpropagation

La retropropagación es un algoritmo fundamental en el entrenamiento de redes neuronales, utilizado para ajustar los pesos de la red con el objetivo de minimizar la función de pérdida.



Rumelhart y otros autores, en 1986, presentan la

"Regla Delta Generalizada"

Para adaptar los pesos propagando los errores hacia atrás, es decir, propagar los errores hacia las capas ocultas inferiores.



Haz clic para
más info.



Consigue trabajar con múltiples capas y al incluir capas intermedias con conexiones hacia adelante (feedforward) y funciones de activación continuas (lineales o sigmoidales) demuestra que el Perceptrón multicapa es un aproximador universal.

Los algoritmos de propagación hacia atrás son una familia de métodos utilizados para entrenar de manera eficiente redes neuronales artificiales (ANN), siguiendo un enfoque de descenso de gradiente que explota la regla de la cadena.



Proceso de Backpropagation



1. Inicialización: los pesos de la red se inicializan aleatoriamente.

2. **Paso hacia adelante (Forward Pass):** se realiza una propagación hacia adelante para obtener las salidas predichas, cada neurona aplica la función de activación a la suma ponderada de sus entradas.

3. **Cálculo de Pérdida:** se calcula la diferencia entre las predicciones y las etiquetas reales utilizando la función de pérdida.

4. **Paso hacia atrás (Backward Pass):** se calcula el gradiente de la función de pérdida con respecto a los pesos mediante la regla de la cadena. El gradiente se propaga hacia atrás a través de la red.

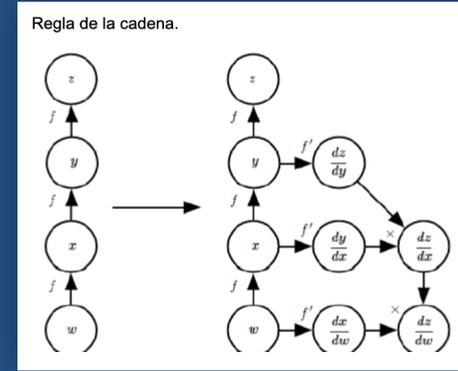
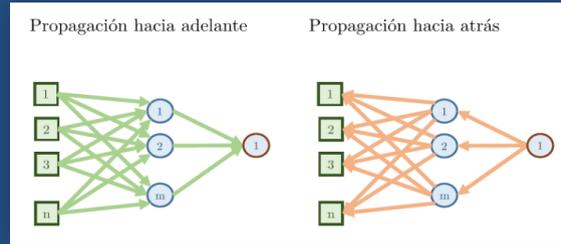
5. **Actualización de Pesos:** se actualizan los pesos utilizando el gradiente descendente. El tamaño del paso (tasa de aprendizaje) controla la magnitud de las actualizaciones.

Iteración: los pasos 2-5 se repiten iterativamente para múltiples lotes de datos o épocas.

Proceso de Backpropagation



Haz clic para ver cada imagen.



$$C(a^L(z^L))$$

$\frac{\partial C}{\partial w^L}$	Costo $\frac{\partial C}{\partial a^L}$	Activación $\frac{\partial a^L}{\partial z^L}$	Sumatoria $\frac{\partial z^L}{\partial w^L}$
$\frac{\partial C}{\partial b^L}$	$\frac{\partial C}{\partial a^L}$	$\frac{\partial a^L}{\partial z^L}$	$\frac{\partial z^L}{\partial b^L}$

$z^L = W^L a^{L-1} + b^L$



$$\frac{\partial C}{\partial w^L} = \frac{\partial C}{\partial a^L} \cdot \frac{\partial a^L}{\partial z^L} \cdot \frac{\partial z^L}{\partial w^L} = \delta^L a_i^{L-1}$$

$$\frac{\partial C}{\partial b^L} = \frac{\partial C}{\partial a^L} \cdot \frac{\partial a^L}{\partial z^L} \cdot \frac{\partial z^L}{\partial b^L} = \delta^L$$

δ^L

$$\frac{\partial C}{\partial w^{L-1}} = \frac{\partial C}{\partial a^L} \cdot \frac{\partial a^L}{\partial z^L} \cdot \frac{\partial z^L}{\partial a^{L-1}} \cdot \frac{\partial a^{L-1}}{\partial z^{L-1}} \cdot \frac{\partial z^{L-1}}{\partial w^{L-1}}$$

$$\frac{\partial C}{\partial b^{L-1}} = \frac{\partial C}{\partial a^L} \cdot \frac{\partial a^L}{\partial z^L} \cdot \frac{\partial z^L}{\partial a^{L-1}} \cdot \frac{\partial a^{L-1}}{\partial z^{L-1}} \cdot \frac{\partial z^{L-1}}{\partial b^{L-1}}$$

δ^L



$$\frac{\partial \mathcal{C}}{\partial w^{L-1}} = \frac{\partial \mathcal{C}}{\partial a^L} \cdot \frac{\partial a^L}{\partial z^L} \cdot \frac{\partial z^L}{\partial a^{L-1}} \cdot \frac{\partial a^{L-1}}{\partial z^{L-1}} \cdot \frac{\partial z^{L-1}}{\partial w^{L-1}}$$
$$\frac{\partial \mathcal{C}}{\partial w^{L-1}} = \frac{\partial \mathcal{C}}{\partial a^L} \cdot \frac{\partial a^L}{\partial z^L} \cdot \frac{\partial z^L}{\partial a^{L-1}} \cdot \frac{\partial a^{L-1}}{\partial z^{L-1}} \cdot \frac{\partial z^{L-1}}{\partial b^{L-1}}$$

δ^L



$$\frac{\partial C}{\partial w^L} = \frac{\partial C}{\partial a^L} \cdot \frac{\partial a^L}{\partial z^L} \cdot \frac{\partial z^L}{\partial w^L} = \delta^L a_i^{L-1}$$
$$\frac{\partial C}{\partial b^L} = \frac{\partial C}{\partial a^L} \cdot \frac{\partial a^L}{\partial z^L} \cdot \frac{\partial z^L}{\partial b^L} = \delta^L$$

$$\delta^L$$



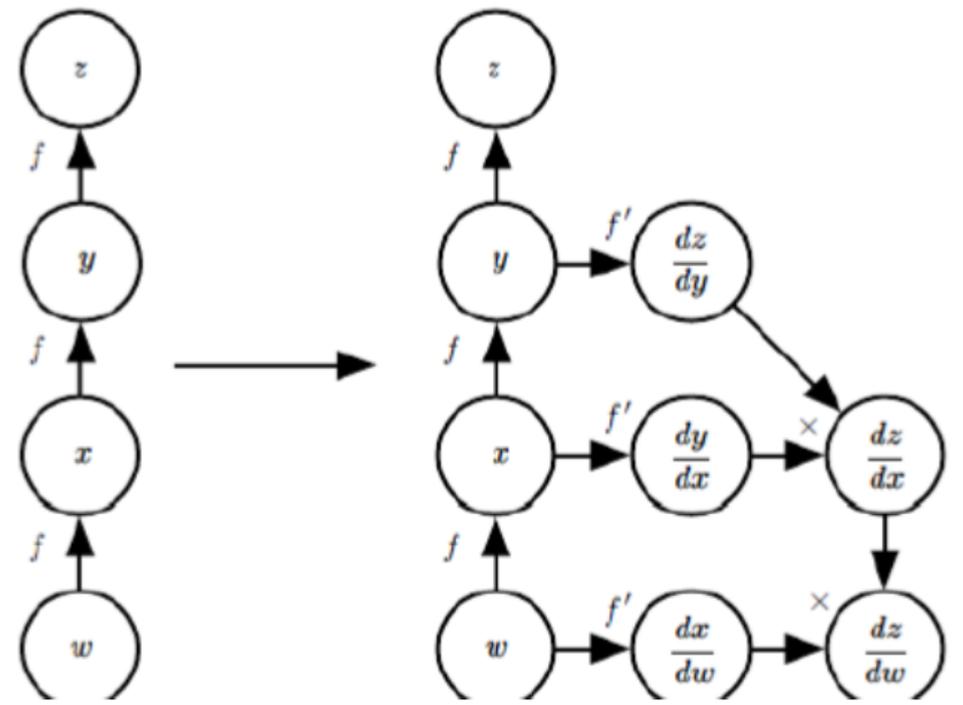
$$C(a^L(z^L))$$

$$\frac{\partial C}{\partial w^L} = \begin{array}{|c|} \hline \text{Costo} \\ \hline \frac{\partial C}{\partial a^L} \\ \hline \end{array} \cdot \begin{array}{|c|} \hline \text{Activación} \\ \hline \frac{\partial a^L}{\partial z^L} \\ \hline \end{array} \cdot \begin{array}{|c|} \hline \text{Sumatoria} \\ \hline \frac{\partial z^L}{\partial w^L} \\ \hline \end{array}$$
$$\frac{\partial C}{\partial b^L} = \begin{array}{|c|} \hline \text{Costo} \\ \hline \frac{\partial C}{\partial a^L} \\ \hline \end{array} \cdot \begin{array}{|c|} \hline \text{Activación} \\ \hline \frac{\partial a^L}{\partial z^L} \\ \hline \end{array} \cdot \begin{array}{|c|} \hline \text{Sumatoria} \\ \hline \frac{\partial z^L}{\partial b^L} \\ \hline \end{array}$$

$$z^L = W^L a^{L-1} + b^L$$

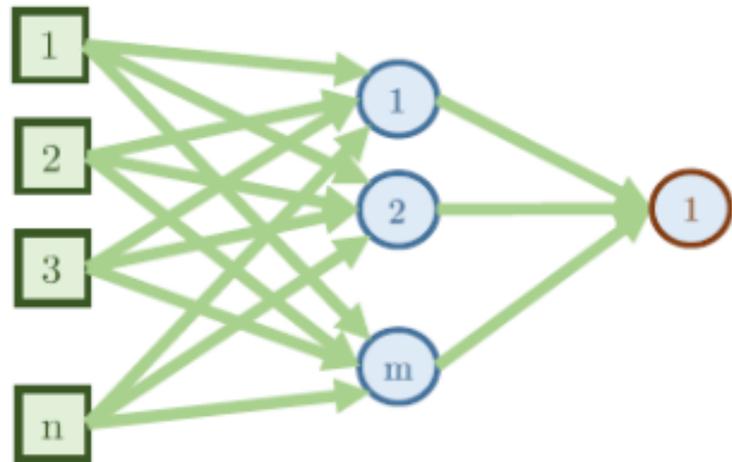


Regla de la cadena.

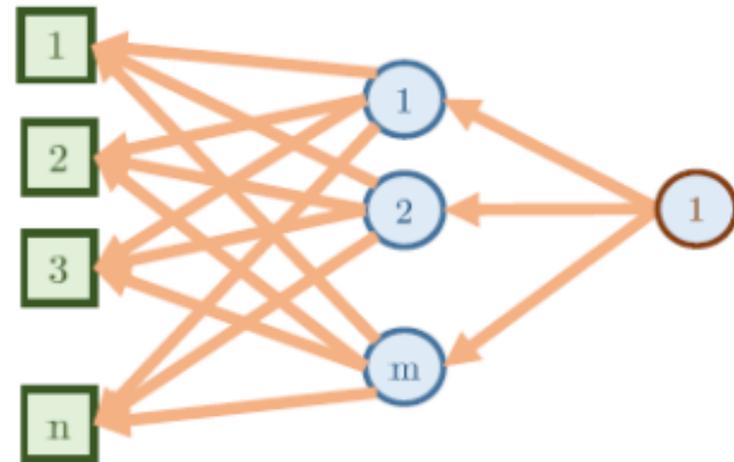




Propagación hacia adelante

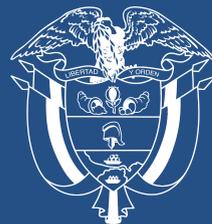


Propagación hacia atrás



Consideraciones importantes

- **Función de Pérdida:** la elección de la función de pérdida depende de la tarea (clasificación, regresión, etc.).
- **Tasa de Aprendizaje:** la tasa de aprendizaje influye en la rapidez con la que los pesos se ajustan. Se debe ajustar cuidadosamente para evitar convergencia demasiado rápida o demasiado lenta.
- **Overfitting:** puede ocurrir si la red se ajusta demasiado a los datos de entrenamiento y no generaliza bien a datos nuevos. Se aborda con técnicas como regularización.
- La **retropropagación** es esencial para el entrenamiento eficiente de redes neuronales y ha sido fundamental para el éxito del aprendizaje profundo en diversas tareas.



TIC

▶ TALENTO
TECH

AZ | PROYECTOS
EDUCATIVOS

