Tema 2 – Optimización y Regularización (Parte 3)

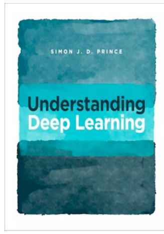
Aprendizaje Automático II - Grado en Inteligencia Artificial
Universidad Rey Juan Carlos

Iván Ramírez Díaz
ivan.ramirez@urjc.es

José Miguel Buenaposada Biencinto
josemiguel.buenaposada@urjc.es

Bibliografía

- Understanding Deep Learning. Capítulo 7.



- Deep Learning for Computer Vision. Lecture 4.
Stanford University. 2017
Curso en youtube.
- Deep Learning: CS 182 2021. Lecture 5.
Sergey Levine. UC Berkeley.
Curso en youtube.

Algoritmo de descenso de gradiente

- Paso 0. Inicializar los parámetros $\boldsymbol{\Phi}_0$.
- Repetir:
  - Paso 1. Calcular derivadas de **la función de coste** con respecto a los parámetros $\boldsymbol{\Phi}$

$$\left.\frac{\partial J(\boldsymbol{\Phi})}{\partial \boldsymbol{\Phi}}\right|_{\boldsymbol{\Phi}=\boldsymbol{\Phi}_t} = \begin{bmatrix} \frac{\partial J(\boldsymbol{\Phi})}{\partial \Phi_0} \\ \frac{\partial J(\boldsymbol{\Phi})}{\partial \Phi_1} \\ \vdots \\ \frac{\partial J(\boldsymbol{\Phi})}{\partial \Phi_k} \end{bmatrix}$$

  - Paso 2. Actualizar los parámetros de acuerdo con:

$$\boldsymbol{\Phi}_{t+1} \leftarrow \boldsymbol{\Phi}_t - \alpha \left.\frac{\partial J(\boldsymbol{\Phi})}{\partial \boldsymbol{\Phi}}\right|_{\boldsymbol{\Phi}=\boldsymbol{\Phi}_t}$$

donde el escalar positivo α determina la magnitud del cambio.

Implementación del descenso de gradiente

- Problema 1. Calcular derivadas de **la función de coste** con respecto a los parámetros

$$\left.\frac{\partial J(\boldsymbol{\Phi})}{\partial \boldsymbol{\Phi}}\right|_{\boldsymbol{\Phi}=\boldsymbol{\Phi}_t} = \begin{bmatrix} \frac{\partial J(\boldsymbol{\Phi})}{\partial \Phi_0} \\ \frac{\partial J(\boldsymbol{\Phi})}{\partial \Phi_1} \\ \vdots \\ \frac{\partial J(\boldsymbol{\Phi})}{\partial \Phi_k} \end{bmatrix}$$

¿Por qué es un problema?

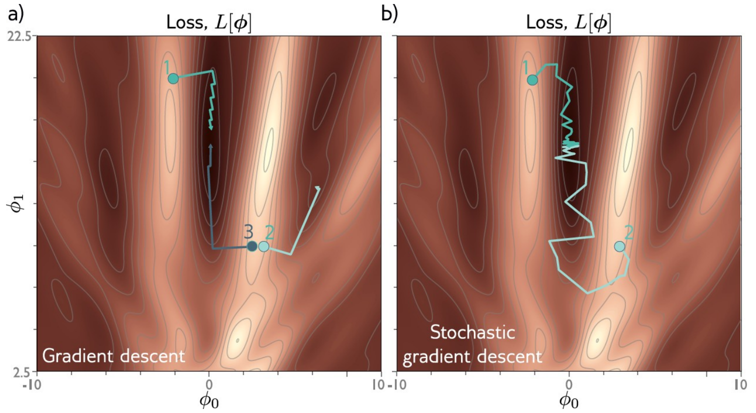
- Muchos modelos son muy complicados con multitud de parámetros:

$$
\begin{aligned}
y' = \phi'_0 + \phi'_1 \mathrm{a}\left[\psi_{10} + \psi_{11}\mathrm{a}[\theta_{10} + \theta_{11}x] + \psi_{12}\mathrm{a}[\theta_{20} + \theta_{21}x] + \psi_{13}\mathrm{a}[\theta_{30} + \theta_{31}x]\right] \\
+ \phi'_2 \mathrm{a}[\psi_{20} + \psi_{21}\mathrm{a}[\theta_{10} + \theta_{11}x] + \psi_{22}\mathrm{a}[\theta_{20} + \theta_{21}x] + \psi_{23}\mathrm{a}[\theta_{30} + \theta_{31}x]] \\
+ \phi'_3 \mathrm{a}[\psi_{30} + \psi_{31}\mathrm{a}[\theta_{10} + \theta_{11}x] + \psi_{32}\mathrm{a}[\theta_{20} + \theta_{21}x] + \psi_{33}\mathrm{a}[\theta_{30} + \theta_{31}x]]
\end{aligned}
$$

- Hay que calcular la derivada:
  - Para cada parámetro
  - Para cada muestra en el mini-batch
  - Para cada iteración del SGD

Implementación del descenso de gradiente

- Problema 2. Inicialización



¿Dónde deberíamos inicializar los parámetros antes de arrancar el SGD?

2.4 Cálculo eficiente del gradiente

- Conceptos matemáticos
- Retropropagación del gradiente (Backpropagation)
- Diferenciación algorítmica

2.4 Cálculo eficiente del gradiente

- Conceptos matemáticos
- Retropropagación del gradiente (Backpropagation)
- Diferenciación algorítmica

Recordatorio: regla de la cadena

- Tenemos una composición de funciones sobre escalares:

$$x \xrightarrow{g} y \xrightarrow{f} z$$

- Por la regla de la cadena de Cálculo:

$$\frac{d}{dx}(f(g(x))) = \frac{dz}{dx} = \frac{dy}{dx} \cdot \frac{dz}{dy}$$

Recordatorio: regla de la cadena

- Tenemos una composición de funciones sobre escalares:

$$x \xrightarrow{g} y \xrightarrow{f} z$$

- Por la regla de la cadena de Cálculo:

$$\frac{d}{dx}(f(g(x))) = \frac{dz}{dx} = \frac{dy}{dx} \cdot \frac{dz}{dy}$$

Ejemplo: *$g(x) = 1 + 4x$; $f(y) = y^2$*

$$\frac{d}{dx}(f(g(x))) = \frac{dy}{dx} \cdot \frac{dz}{dy}$$

Recordatorio: regla de la cadena

- Tenemos una composición de funciones sobre escalares:

$$x \xrightarrow{g} y \xrightarrow{f} z$$

- Por la regla de la cadena de Cálculo:

$$\frac{d}{dx}(f(g(x))) = \frac{dz}{dx} = \frac{dy}{dx} \cdot \frac{dz}{dy}$$

Ejemplo: *$g(x) = 1 + 4x$; $f(y) = y^2$*

$$\frac{d}{dx}(f(g(x))) = \frac{dg(x)}{dx} \cdot \left.\frac{df(y)}{dy}\right|_{y=g(x)}$$

Recordatorio: regla de la cadena

- Tenemos una composición de funciones sobre escalares:

$$x \xrightarrow{g} y \xrightarrow{f} z$$

- Por la regla de la cadena de Cálculo:

$$\frac{d}{dx}(f(g(x))) = \frac{dz}{dx} = \frac{dy}{dx} \cdot \frac{dz}{dy}$$

Ejemplo: *$g(x) = 1 + 4x$; $f(y) = y^2$*

$$\frac{d}{dx}(f(g(x))) = 4 \cdot \left.\frac{df(y)}{dy}\right|_{y=g(x)}$$

Recordatorio: regla de la cadena

- Tenemos una composición de funciones sobre escalares:

$$x \xrightarrow{g} y \xrightarrow{f} z$$

- Por la regla de la cadena de Cálculo:

$$\frac{d}{dx}(f(g(x))) = \frac{dz}{dx} = \frac{dy}{dx} \cdot \frac{dz}{dy}$$

Ejemplo: *g(x) = 1 + 4x; f(y) = y²*

$$\frac{d}{dx}(f(g(x))) = 4 \cdot (2 \cdot y)\big|_{y=g(x)}$$

Recordatorio: regla de la cadena

- Tenemos una composición de funciones sobre escalares:

$$x \xrightarrow{g} y \xrightarrow{f} z$$

- Por la regla de la cadena de Cálculo:

$$\frac{d}{dx}(f(g(x))) = \frac{dz}{dx} = \frac{dy}{dx} \cdot \frac{dz}{dy}$$

Ejemplo: *g(x) = 1 + 4x; f(y) = y²*

$$\frac{d}{dx}(f(g(x))) = 4 \cdot (2 \cdot y)\big|_{y=1+4x}$$

Recordatorio: regla de la cadena

- Tenemos una composición de funciones sobre escalares:

$$x \xrightarrow{g} y \xrightarrow{f} z$$

- Por la regla de la cadena de Cálculo:

$$\frac{d}{dx}(f(g(x))) = \frac{dz}{dx} = \frac{dy}{dx} \cdot \frac{dz}{dy}$$

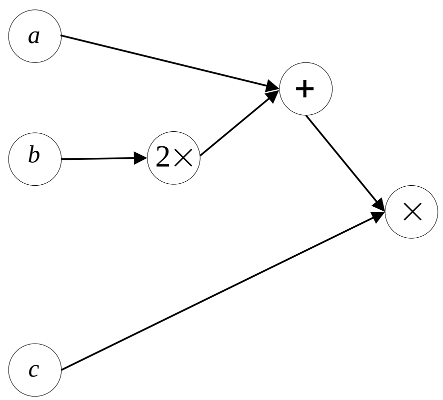
Ejemplo: *g(x) = 1 + 4x; f(y) = y²*

$$\frac{d}{dx}(f(g(x))) = 4 \cdot 2 \cdot (1+4x) = 8 + 32x$$

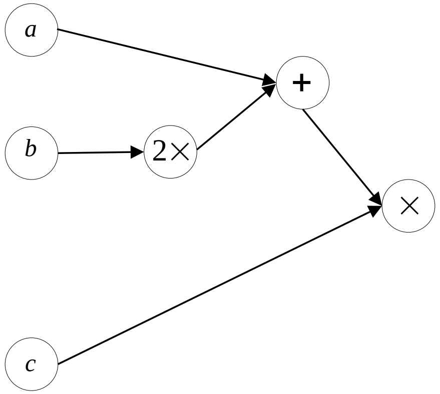
Grafos de cómputo

¿qué **expresión** calcula el grafo?

$$J(a,b,c) = (a + 2 \cdot b) \cdot c$$



Grafos de cómputo



¿qué **expresión** calcula el grafo?

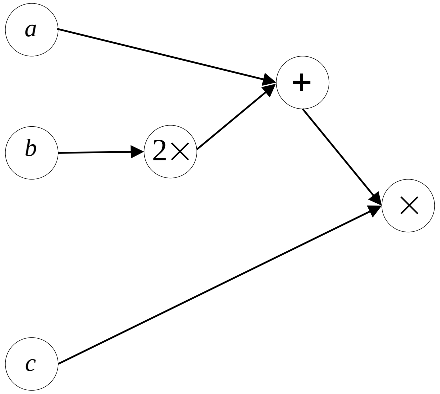
$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

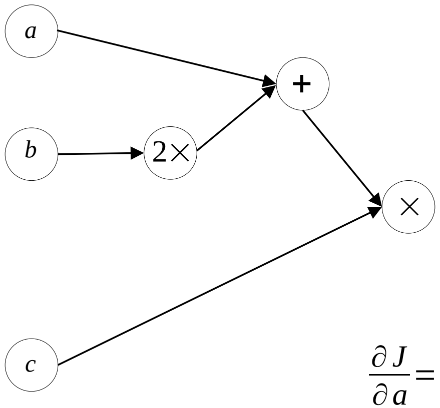
Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial a}$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

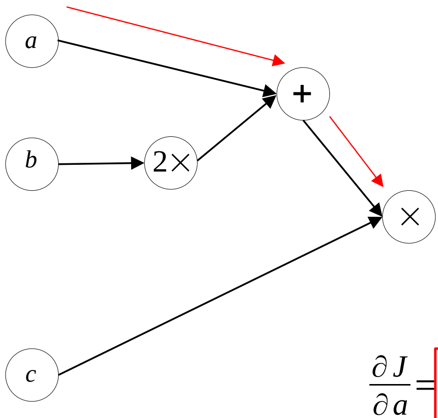
$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial a}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial a}=$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

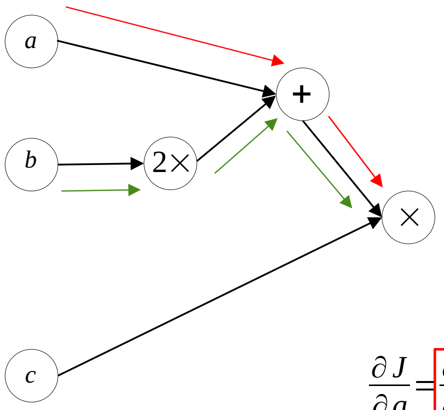
$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial a}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial a}=\boxed{\frac{\partial g}{\partial a}\cdot\frac{\partial f}{\partial x}}+\ldots$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

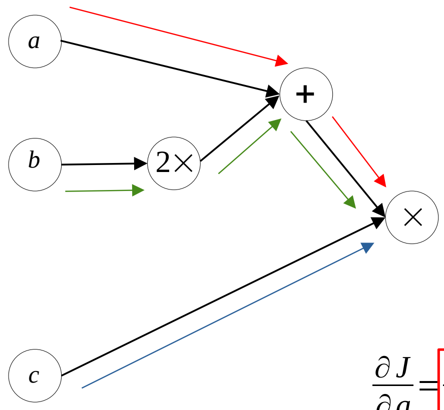
$$f(x,c)=x\cdot c \qquad g(a,s)=a+s \qquad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial a}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial a}=\boxed{\frac{\partial g}{\partial a}\cdot\frac{\partial f}{\partial x}}+\boxed{\frac{\partial b}{\partial a}\cdot\frac{\partial h}{\partial b}\cdot\frac{\partial g}{\partial s}\cdot\frac{\partial f}{\partial x}}+\dots$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

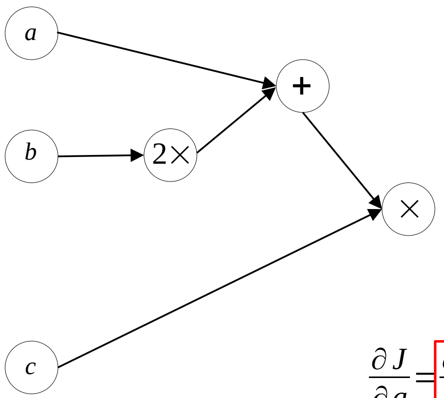
$$f(x,c)=x\cdot c \qquad g(a,s)=a+s \qquad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial a}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial a}=\boxed{\frac{\partial g}{\partial a}\cdot\frac{\partial f}{\partial x}}+\boxed{\frac{\partial b}{\partial a}\cdot\frac{\partial h}{\partial b}\cdot\frac{\partial g}{\partial s}\cdot\frac{\partial f}{\partial x}}+\boxed{\frac{\partial c}{\partial a}\cdot\frac{\partial f}{\partial c}}$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

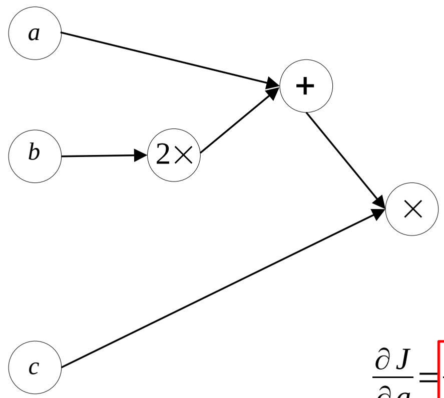
$$f(x,c)=x\cdot c \qquad g(a,s)=a+s \qquad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial a}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial a}=\boxed{\frac{\partial g}{\partial a}\cdot\frac{\partial f}{\partial x}}+\boxed{\cancelto{0}{\frac{\partial b}{\partial a}}\cdot\frac{\partial h}{\partial b}\cdot\frac{\partial g}{\partial s}\cdot\frac{\partial f}{\partial x}}+\boxed{\cancelto{0}{\frac{\partial c}{\partial a}}\cdot\frac{\partial f}{\partial c}}$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

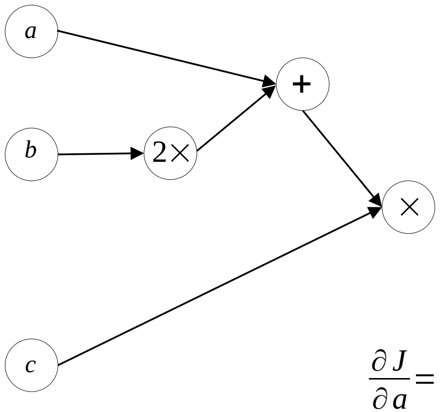
$$f(x,c)=x\cdot c \qquad g(a,s)=a+s \qquad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial a}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial a}=\boxed{\frac{\partial g}{\partial a}\cdot\frac{\partial f}{\partial x}}$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

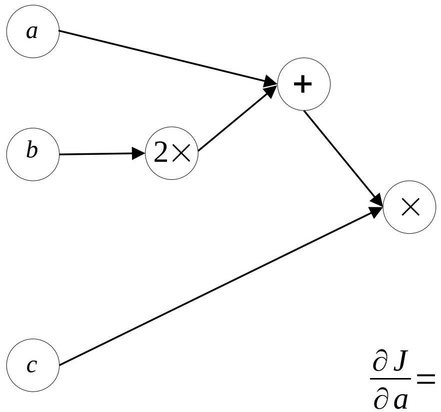
$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial a}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial a}=\frac{\partial g}{\partial a}\bigg|_{s=h(b)}\cdot\frac{\partial f}{\partial x}\bigg|_{x=g(a,h(b))}$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

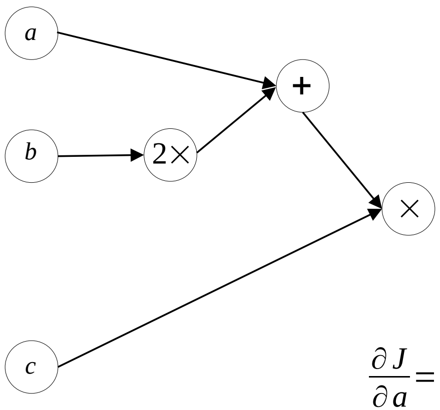
$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial a}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial a}=1\cdot\frac{\partial f}{\partial x}\bigg|_{x=g(a,h(b))}$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

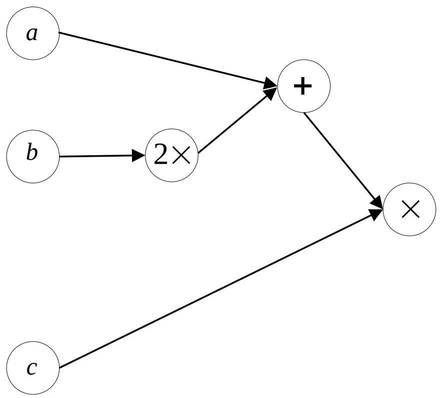
$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial a}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial a}=1\cdot c$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

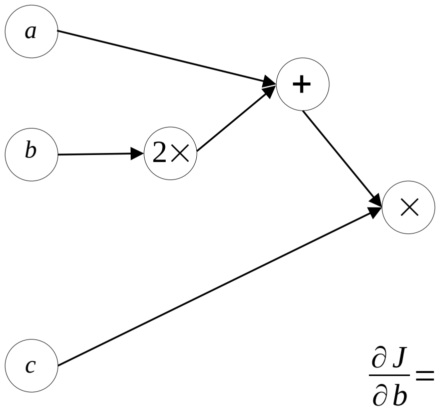
Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial b}$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

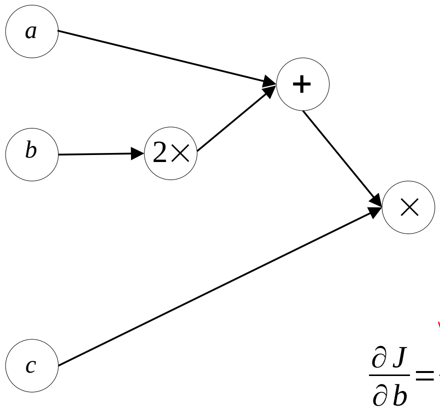
$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial b}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial b}=$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

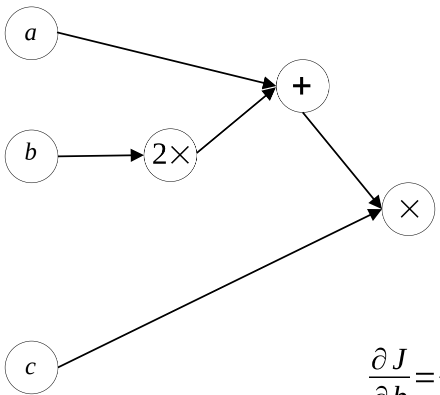
$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial b}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial b}=\cancelto{0}{\frac{\partial a}{\partial b}}\cdot\frac{\partial g}{\partial a}\cdot\frac{\partial f}{\partial x}+\frac{\partial h}{\partial b}\cdot\frac{\partial g}{\partial s}\cdot\frac{\partial f}{\partial x}+\cancelto{0}{\frac{\partial c}{\partial b}}\cdot\frac{\partial f}{\partial c}$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

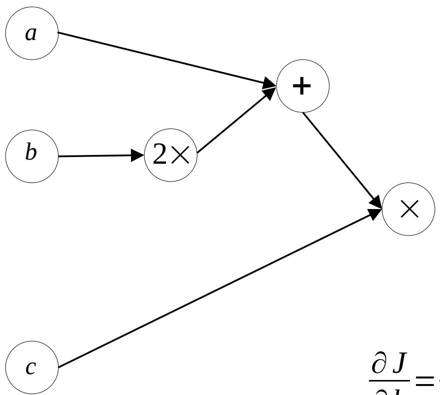
$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial b}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial b}=\frac{\partial h}{\partial b}\cdot\frac{\partial g}{\partial s}\cdot\frac{\partial f}{\partial x}$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

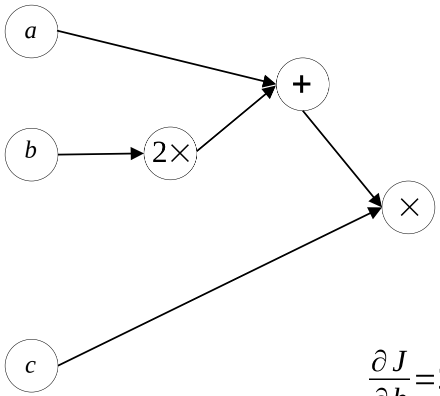
$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial b}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial b}=\frac{\partial h}{\partial b}\cdot\left.\frac{\partial g}{\partial s}\right|_{s=h(b)}\cdot\left.\frac{\partial f}{\partial x}\right|_{x=g(a,h(b))}$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

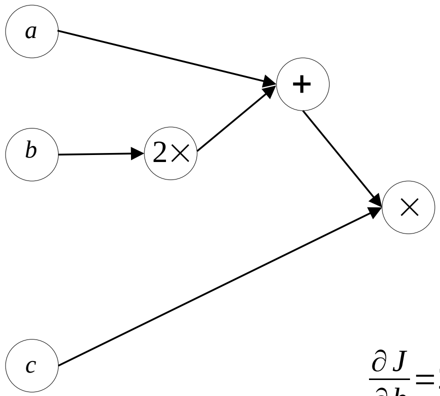
$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Calculemos: $\dfrac{\partial J}{\partial b}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial b}=2\cdot\left.\frac{\partial g}{\partial s}\right|_{s=h(b)}\cdot\left.\frac{\partial f}{\partial x}\right|_{x=g(a,h(b))}$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

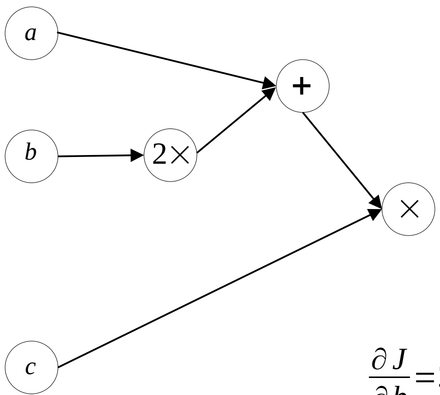
$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Calculemos: $\dfrac{\partial J}{\partial b}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial b}=2\cdot 1\cdot\left.\frac{\partial f}{\partial x}\right|_{x=g(a,h(b))}$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

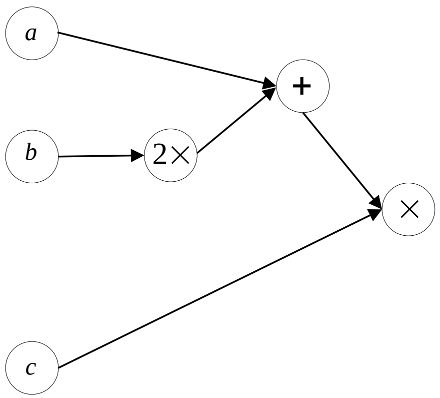
$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Calculemos: $\dfrac{\partial J}{\partial b}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial b}=2\cdot 1\cdot c=2\cdot c$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

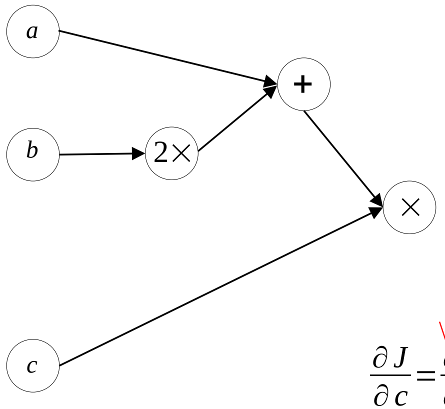
Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

Calculemos: $\dfrac{\partial J}{\partial c}$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

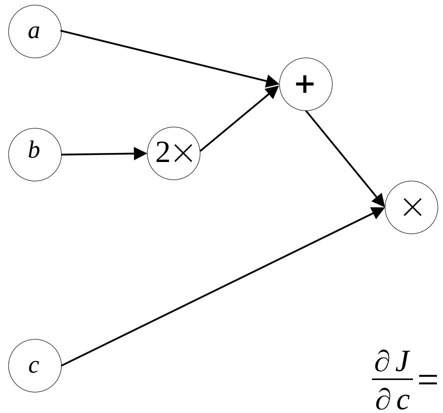
$$f(x,c)=x\cdot c \qquad g(a,s)=a+s \qquad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial c}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial c}=\underbrace{\frac{\partial a}{\partial c}}_{0}\cdot\frac{\partial g}{\partial a}\cdot\frac{\partial f}{\partial x}+\underbrace{\frac{\partial b}{\partial c}}_{0}\cdot\frac{\partial h}{\partial b}\cdot\frac{\partial g}{\partial s}\cdot\frac{\partial f}{\partial x}+\frac{\partial f}{\partial c}$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

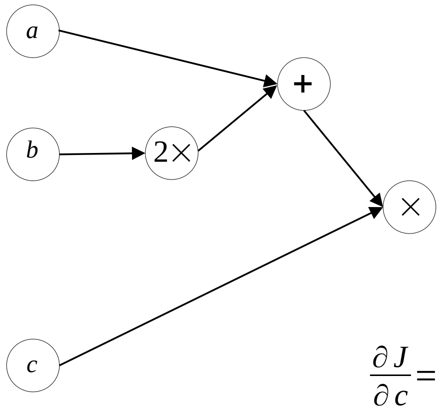
$$f(x,c)=x\cdot c \qquad g(a,s)=a+s \qquad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial c}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial c}=\left.\frac{\partial f}{\partial c}\right|_{x=g(a,h(b))}$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

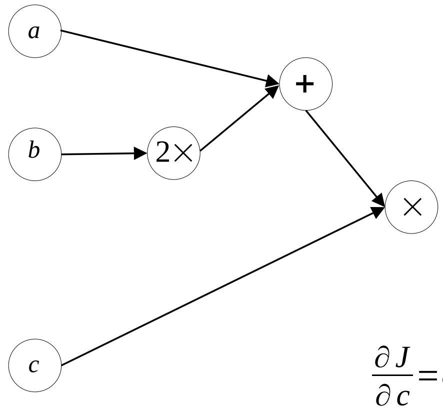
$$f(x,c)=x\cdot c \qquad g(a,s)=a+s \qquad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial c}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial c}=\left.x\right|_{x=g(a,h(b))}$$

Grafos de cómputo



¿qué **expresión** calcula el grafo?

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

Composición de funciones:

$$J(a,b,c)=f(g(a,h(b)),c)$$

$$f(x,c)=x\cdot c \qquad g(a,s)=a+s \qquad h(b)=2\cdot b$$

Calculemos: $\frac{\partial J}{\partial c}$

La derivada usa todos los caminos:

$$\frac{\partial J}{\partial c}=a+2b$$

2.4 Cálculo eficiente del gradiente

- Conceptos matemáticos
- Retropropagación del gradiente (Backpropagation)
- Diferenciación algorítmica

Retropropagación en grafos de cómputo

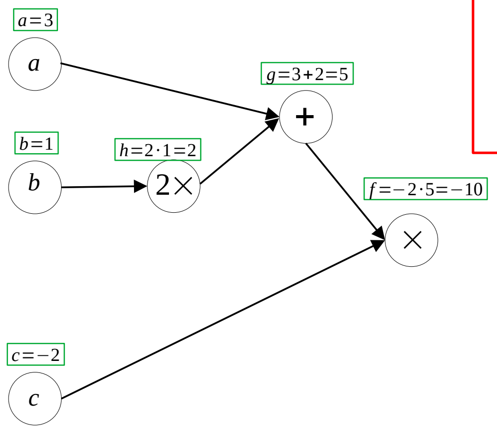
Fase 1: ejecución del grafo (forward)

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

$$\frac{\partial f}{\partial x}=c \quad \frac{\partial g}{\partial a}=1 \quad \frac{\partial h}{\partial b}=2$$

$$\frac{\partial f}{\partial c}=x \quad \frac{\partial g}{\partial s}=1$$



Retropropagación en grafos de cómputo

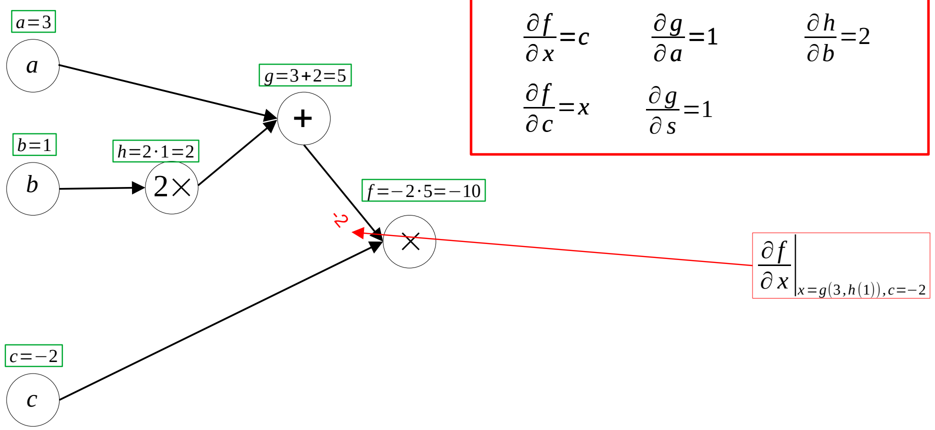
Fase 2: retropropagación (backward)

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

$$\frac{\partial f}{\partial x}=c \quad \frac{\partial g}{\partial a}=1 \quad \frac{\partial h}{\partial b}=2$$

$$\frac{\partial f}{\partial c}=x \quad \frac{\partial g}{\partial s}=1$$



Retropropagación en grafos de cómputo

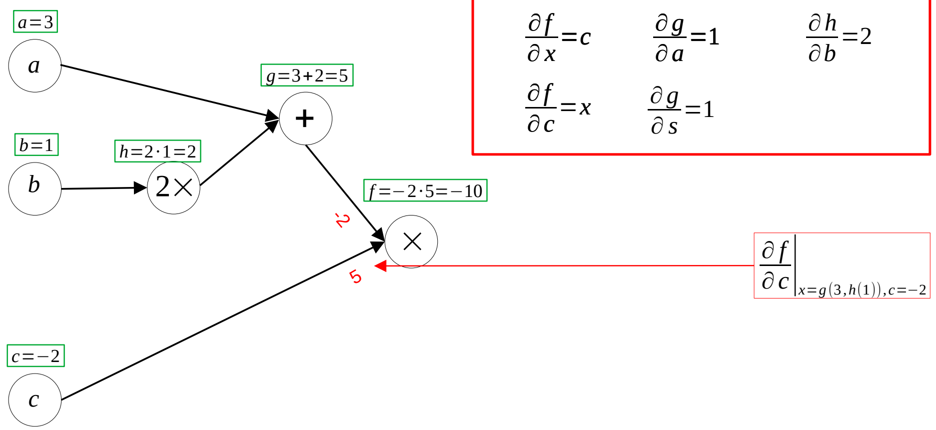
Fase 2: retropropagación (backward)

$$J(a,b,c)=(a+2\cdot b)\cdot c$$

$$f(x,c)=x\cdot c \quad g(a,s)=a+s \quad h(b)=2\cdot b$$

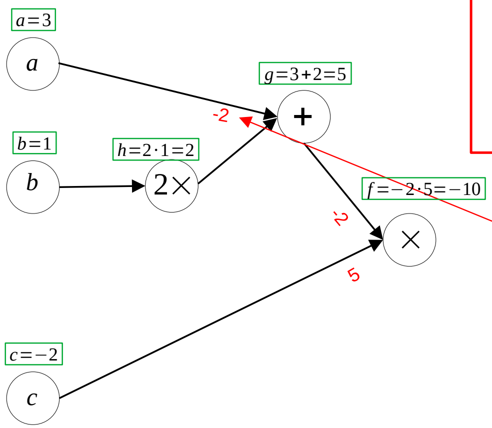
$$\frac{\partial f}{\partial x}=c \quad \frac{\partial g}{\partial a}=1 \quad \frac{\partial h}{\partial b}=2$$

$$\frac{\partial f}{\partial c}=x \quad \frac{\partial g}{\partial s}=1$$



Retropropagación en grafos de cómputo

Fase 2: retropropagación (backward)



$$J(a,b,c)=(a+2\cdot b)\cdot c$$

$$f(x,c)=x\cdot c \qquad g(a,s)=a+s \qquad h(b)=2\cdot b$$

$$\frac{\partial f}{\partial x}=c \qquad \frac{\partial g}{\partial a}=1 \qquad \frac{\partial h}{\partial b}=2$$

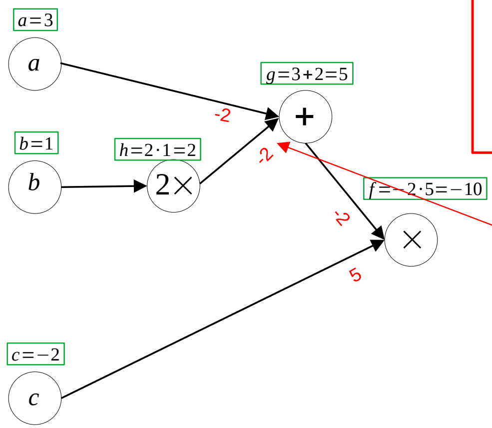
$$\frac{\partial f}{\partial c}=x \qquad \frac{\partial g}{\partial s}=1$$

$$\frac{\partial f}{\partial a}=\left.\frac{\partial g}{\partial a}\right|_{s=h(1)}\cdot\left.\frac{\partial f}{\partial x}\right|_{x=g(3,h(1)),c=-2}$$

No se recalcula lo que ya conocemos (la derivada de f con respecto a x)

Retropropagación en grafos de cómputo

Fase 2: retropropagación (backward)



$$J(a,b,c)=(a+2\cdot b)\cdot c$$

$$f(x,c)=x\cdot c \qquad g(a,s)=a+s \qquad h(b)=2\cdot b$$

$$\frac{\partial f}{\partial x}=c \qquad \frac{\partial g}{\partial a}=1 \qquad \frac{\partial h}{\partial b}=2$$

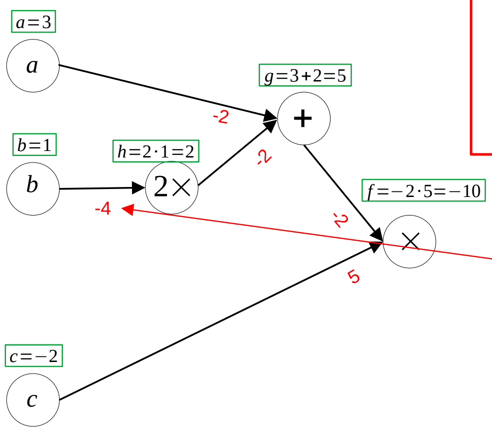
$$\frac{\partial f}{\partial c}=x \qquad \frac{\partial g}{\partial s}=1$$

$$\frac{\partial f}{\partial s}=\left.\frac{\partial g}{\partial s}\right|_{s=h(1)}\cdot\left.\frac{\partial f}{\partial x}\right|_{x=g(3,h(1)),c=-2}$$

No se recalcula lo que ya conocemos (la derivada de f con respecto a x)

Retropropagación en grafos de cómputo

Fase 2: retropropagación (backward)



$$J(a,b,c)=(a+2\cdot b)\cdot c$$

$$f(x,c)=x\cdot c \qquad g(a,s)=a+s \qquad h(b)=2\cdot b$$

$$\frac{\partial f}{\partial x}=c \qquad \frac{\partial g}{\partial a}=1 \qquad \frac{\partial h}{\partial b}=2$$

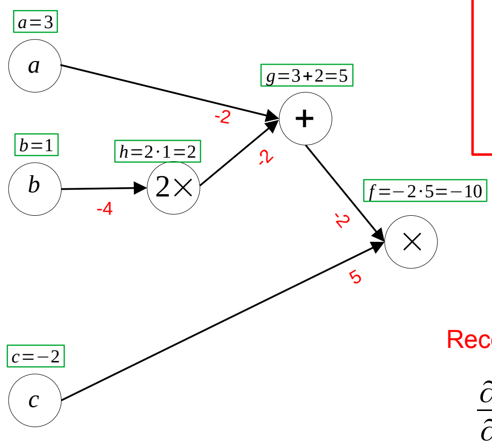
$$\frac{\partial f}{\partial c}=x \qquad \frac{\partial g}{\partial s}=1$$

$$\frac{\partial f}{\partial b}=\left.\frac{\partial h}{\partial b}\right|_{b=1}\cdot\left.\frac{\partial f}{\partial s}\right|_{x=g(3,h(1)),c=-2}$$

No se recalcula lo que ya conocemos (la derivada de f con respecto a s)

Retropropagación en grafos de cómputo

Fase 2: retropropagación (backward)



$$J(a,b,c)=(a+2\cdot b)\cdot c$$

$$f(x,c)=x\cdot c \qquad g(a,s)=a+s \qquad h(b)=2\cdot b$$

$$\frac{\partial f}{\partial x}=c \qquad \frac{\partial g}{\partial a}=1 \qquad \frac{\partial h}{\partial b}=2$$

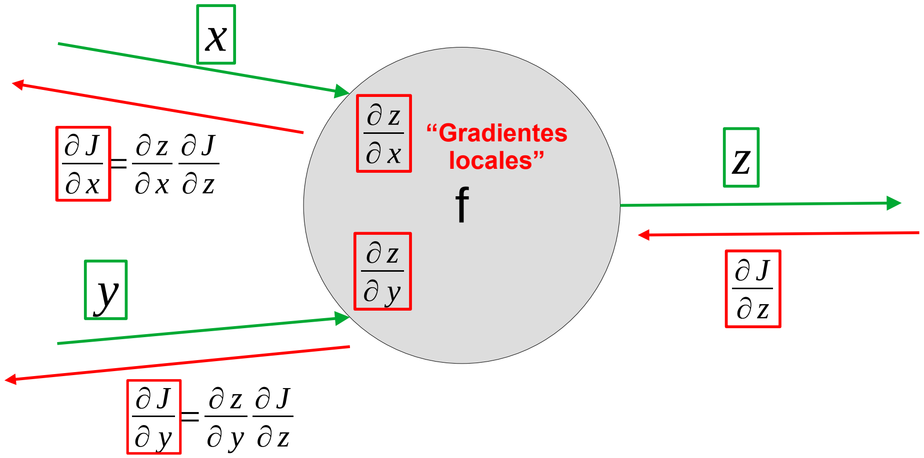
$$\frac{\partial f}{\partial c}=x \qquad \frac{\partial g}{\partial s}=1$$

Recordatorio:

$$\frac{\partial J}{\partial a}=c \qquad \frac{\partial J}{\partial b}=2\cdot c \qquad \frac{\partial J}{\partial c}=a+2b$$

Retropropagación en grafos de cómputo

- En cada nodo únicamente necesitamos calcular las derivadas locales con respecto a sus entradas:

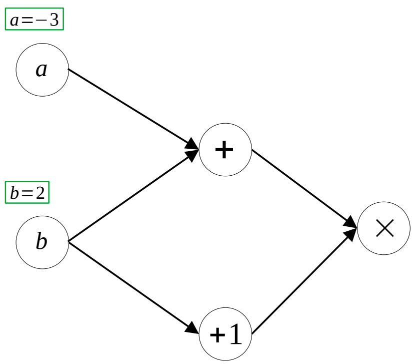


Algoritmo de retropropagación

- Forward: Se ejecuta el grafo de cómputo dadas unas entradas
- Backward:
  - Inicializar $\delta = \frac{dJ}{d\boxed{z^{(n)}}}$ — Salida de la última operación del grafo de cómputo
  - Para cada ***f*** con entrada x_f desde el final al principio:

$$\delta \leftarrow \frac{df}{dx_f} \cdot \delta$$

Grafos de cómputo: Ejemplo 2



¿qué **expresión** calcula el grafo?

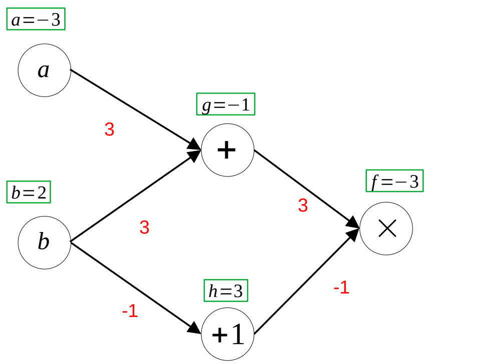
$$J(a,b) = (a+b)\cdot(b+1)$$

$$g = a+b$$
$$h = b+1$$
$$f = g\cdot h$$

$$\frac{\partial f}{\partial g} = h \qquad \frac{\partial f}{\partial h} = g$$

$$\frac{\partial h}{\partial b} = 1 \qquad \frac{\partial g}{\partial a} = \frac{\partial g}{\partial b} = 1$$

Grafos de cómputo: Ejemplo 2



$$\frac{df}{db} = ¿?$$

$$\frac{df}{db} = a + 2\cdot b + 1$$

¿qué **expresión** calcula el grafo?

$$J(a,b) = (a+b)\cdot(b+1)$$

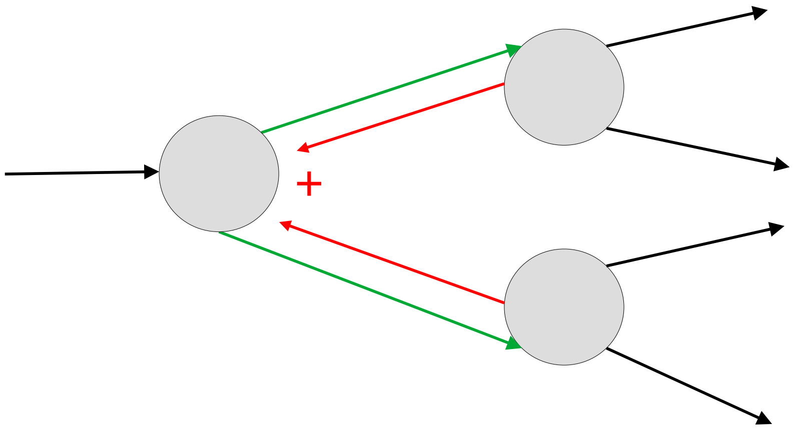
$$g = a+b$$
$$h = b+1$$
$$f = g\cdot h$$

$$\frac{\partial f}{\partial g} = h \qquad \frac{\partial f}{\partial h} = g$$

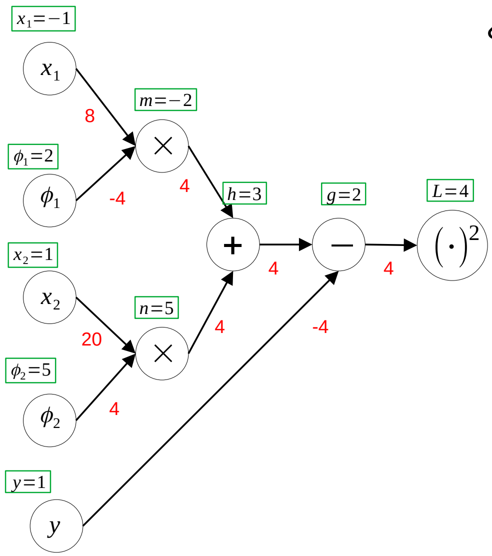
$$\frac{\partial h}{\partial b} = 1 \qquad \frac{\partial g}{\partial a} = \frac{\partial g}{\partial b} = 1$$

Retropropagación en grafos de cómputo

- Los gradientes se suman en las ramificaciones:



Grafos de cómputo: Ejemplo 3



¿qué **expresión** calcula el grafo?

$$L(\boldsymbol{x}, y; \boldsymbol{\phi}) = ((x_1\phi_1 + x_2\phi_2) - y)^2$$

$$m = x_1 \cdot \phi_1 \qquad n = x_2 \cdot \phi_2$$

$$h = m + n$$

$$g = h - y$$

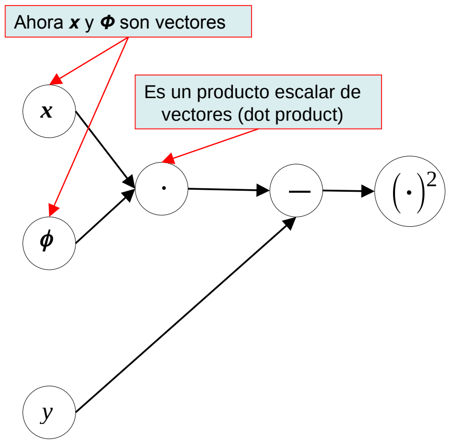
$$L = g^2$$

$$\frac{dL}{dg} = 2 \cdot g \qquad \frac{dg}{dh} = 1 \qquad \frac{dh}{dm} = \frac{dh}{dn} = 1$$

$$\frac{dm}{dx_1} = \phi_1 \qquad \frac{dm}{d\phi_1} = x_1 \qquad \frac{dg}{dy} = -1$$

$$\frac{dn}{dx_2} = \phi_2 \qquad \frac{dn}{d\phi_2} = x_2$$

Grafos de cómputo: vectores y matrices

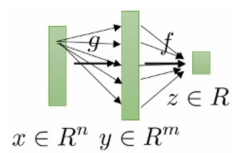


¿qué **expresión** calcula el grafo?

$$L(\boldsymbol{x}, y; \boldsymbol{\phi}) = (\boldsymbol{\phi}^T \boldsymbol{x} - y)^2$$

Recordatorio: regla de la cadena

- Si tenemos una composición de funciones sobre vectores:

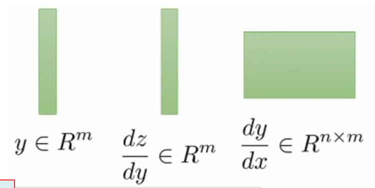


- Por la regla de la cadena de Cálculo:

$$\frac{d}{d\boldsymbol{x}}(f(g(\boldsymbol{x}))) = \frac{d\boldsymbol{z}}{d\boldsymbol{x}} = \frac{d\boldsymbol{y}}{d\boldsymbol{x}} \cdot \frac{d\boldsymbol{z}}{d\boldsymbol{y}}$$

Jacobiano de *g*

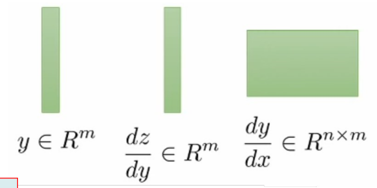
Jacobiano de *f*



Recordatorio: regla de la cadena

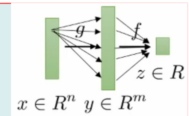
- Por la regla de la cadena de Cálculo:

$$\frac{d}{d\mathbf{x}}(f(g(\mathbf{x})))=\frac{d\mathbf{z}}{d\mathbf{x}}=\underbrace{\frac{d\mathbf{y}}{d\mathbf{x}}}_{\text{Jacobiano de } g}\cdot\underbrace{\frac{d\mathbf{z}}{d\mathbf{y}}}_{\text{Jacobiano de } f}$$



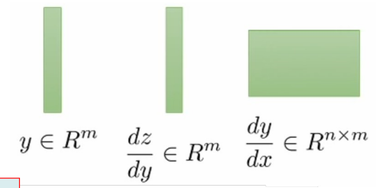
Recordatorio: regla de la cadena

Ejemplo: $g(\mathbf{x})=\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ $\quad f(\mathbf{y})=[4 \quad 3]\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$



- Por la regla de la cadena de Cálculo:

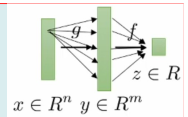
$$\frac{d}{d\mathbf{x}}(f(g(\mathbf{x})))=\frac{d\mathbf{z}}{d\mathbf{x}}=\underbrace{\frac{d\mathbf{y}}{d\mathbf{x}}}_{\text{Jacobiano de } g}\cdot\underbrace{\frac{d\mathbf{z}}{d\mathbf{y}}}_{\text{Jacobiano de } f}$$



Recordatorio: regla de la cadena

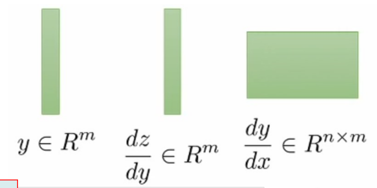
Ejemplo: $g(\mathbf{x})=\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ $\quad f(\mathbf{y})=[4 \quad 3]\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$

$$f(g(\mathbf{x}))=[4 \quad 3]\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$



- Por la regla de la cadena de Cálculo:

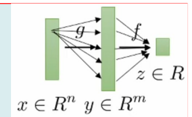
$$\frac{d}{d\mathbf{x}}(f(g(\mathbf{x})))=\frac{d\mathbf{z}}{d\mathbf{x}}=\underbrace{\frac{d\mathbf{y}}{d\mathbf{x}}}_{\text{Jacobiano de } g}\cdot\underbrace{\frac{d\mathbf{z}}{d\mathbf{y}}}_{\text{Jacobiano de } f}$$



Recordatorio: regla de la cadena

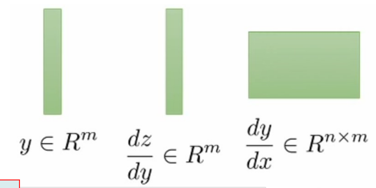
Ejemplo: $g(\mathbf{x})=\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ $\quad f(\mathbf{y})=[4 \quad 3]\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$

$$f(g(\mathbf{x}))=[4 \quad 3]\begin{bmatrix} w_{11}\cdot x_1+w_{12}\cdot x_2+w_{13}\cdot x_3 \\ w_{21}\cdot x_1+w_{22}\cdot x_2+w_{23}\cdot x_3 \end{bmatrix}$$



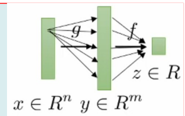
- Por la regla de la cadena de Cálculo:

$$\frac{d}{d\mathbf{x}}(f(g(\mathbf{x})))=\frac{d\mathbf{z}}{d\mathbf{x}}=\underbrace{\frac{d\mathbf{y}}{d\mathbf{x}}}_{\text{Jacobiano de } g}\cdot\underbrace{\frac{d\mathbf{z}}{d\mathbf{y}}}_{\text{Jacobiano de } f}$$



Recordatorio: regla de la cadena

Ejemplo: $g(\mathbf{x})=\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ $f(\mathbf{y})=\begin{bmatrix} 4 & 3 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$



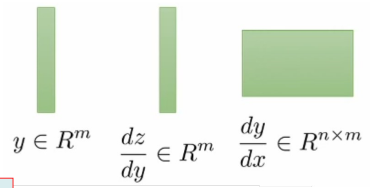
$f(g(\mathbf{x}))=4\cdot(w_{11}\cdot x_1+w_{12}\cdot x_2+w_{13}\cdot x_3)+3\cdot(w_{21}\cdot x_1+w_{22}\cdot x_2+w_{23}\cdot x_3)$

- Por la regla de la cadena de Cálculo:

$$\frac{d}{d\,\mathbf{x}}(f(g(\mathbf{x})))=\frac{d\,\mathbf{z}}{d\,\mathbf{x}}=\frac{d\,\mathbf{y}}{d\,\mathbf{x}}\cdot\frac{d\,\mathbf{z}}{d\,\mathbf{y}}$$

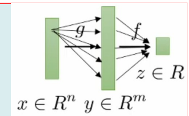
Jacobiano de *g*

Jacobiano de *f*



Recordatorio: regla de la cadena

Ejemplo: $g(\mathbf{x})=\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ $f(\mathbf{y})=\begin{bmatrix} 4 & 3 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$



$f(g(\mathbf{x}))=4\cdot(w_{11}\cdot x_1+w_{12}\cdot x_2+w_{13}\cdot x_3)+3\cdot(w_{21}\cdot x_1+w_{22}\cdot x_2+w_{23}\cdot x_3)$

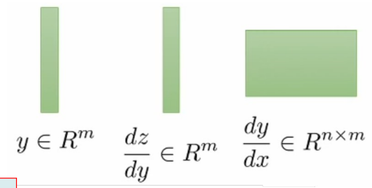
$$\frac{d}{d\,\mathbf{x}}(f(g(\mathbf{x})))=\begin{bmatrix} \frac{\partial}{\partial x_1}(f(g(\mathbf{x}))) \\ \frac{\partial}{\partial x_2}(f(g(\mathbf{x}))) \\ \frac{\partial}{\partial x_3}(f(g(\mathbf{x}))) \end{bmatrix}=\begin{bmatrix} 4\cdot w_{11}+3\cdot w_{21} \\ 4\cdot w_{12}+3\cdot w_{22} \\ 4\cdot w_{13}+3\cdot w_{23} \end{bmatrix}$$

- Por la regla de la cadena de Cálculo:

$$\frac{d}{d\,\mathbf{x}}(f(g(\mathbf{x})))=\frac{d\,\mathbf{z}}{d\,\mathbf{x}}=\frac{d\,\mathbf{y}}{d\,\mathbf{x}}\cdot\frac{d\,\mathbf{z}}{d\,\mathbf{y}}$$

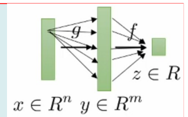
Jacobiano de *g*

Jacobiano de *f*



Recordatorio: regla de la cadena

Ejemplo: $g(\mathbf{x})=\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ $f(\mathbf{y})=\begin{bmatrix} 4 & 3 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$



$f(g(\mathbf{x}))=4\cdot(w_{11}\cdot x_1+w_{12}\cdot x_2+w_{13}\cdot x_3)+3\cdot(w_{21}\cdot x_1+w_{22}\cdot x_2+w_{23}\cdot x_3)$

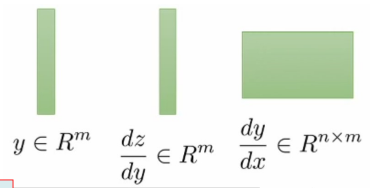
$$\boxed{\frac{d}{d\,\mathbf{x}}(f(g(\mathbf{x})))=\begin{bmatrix} 4\cdot w_{11}+3\cdot w_{21} \\ 4\cdot w_{12}+3\cdot w_{22} \\ 4\cdot w_{13}+3\cdot w_{23} \end{bmatrix}}$$

- Por la regla de la cadena de Cálculo:

$$\frac{d}{d\,\mathbf{x}}(f(g(\mathbf{x})))=\frac{d\,\mathbf{z}}{d\,\mathbf{x}}=\frac{d\,\mathbf{y}}{d\,\mathbf{x}}\cdot\frac{d\,\mathbf{z}}{d\,\mathbf{y}}$$

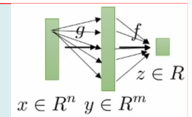
Jacobiano de *g*

Jacobiano de *f*



Recordatorio: regla de la cadena

Ejemplo: $g(\mathbf{x})=\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ $f(\mathbf{y})=\begin{bmatrix} 4 & 3 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$



$f(g(\mathbf{x}))=4\cdot(w_{11}\cdot x_1+w_{12}\cdot x_2+w_{13}\cdot x_3)+3\cdot(w_{21}\cdot x_1+w_{22}\cdot x_2+w_{23}\cdot x_3)$

$$\boxed{\frac{d}{d\,\mathbf{x}}(f(g(\mathbf{x})))=\begin{bmatrix} 4\cdot w_{11}+3\cdot w_{21} \\ 4\cdot w_{12}+3\cdot w_{22} \\ 4\cdot w_{13}+3\cdot w_{23} \end{bmatrix}}$$

Regla de la cadena:

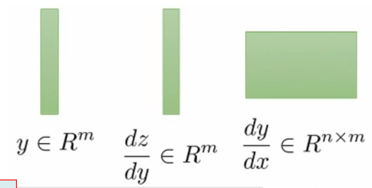
$$\frac{d}{d\,\mathbf{x}}(f(g(\mathbf{x})))=\frac{g(\mathbf{x})}{d\,\mathbf{x}}\cdot\frac{d\,f(\mathbf{y})}{d\,\mathbf{y}}\bigg|_{\mathbf{y}=g(\mathbf{x})}$$

- Por la regla de la cadena de Cálculo:

$$\frac{d}{d\,\mathbf{x}}(f(g(\mathbf{x})))=\frac{d\,\mathbf{z}}{d\,\mathbf{x}}=\frac{d\,\mathbf{y}}{d\,\mathbf{x}}\cdot\frac{d\,\mathbf{z}}{d\,\mathbf{y}}$$

Jacobiano de *g*

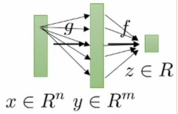
Jacobiano de *f*



Recordatorio: regla de la cadena

Ejemplo:

$$g(\mathbf{x})=\begin{bmatrix} w_{11} & w_{12} & w_{13}\\ w_{21} & w_{22} & w_{22}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} \qquad f(\mathbf{y})=\begin{bmatrix}4 & 3\end{bmatrix}\begin{bmatrix}y_1\\y_2\end{bmatrix}$$



$$f(g(\mathbf{x}))=4\cdot(w_{11}\cdot x_1+w_{12}\cdot x_2+w_{13}\cdot x_3)+3\cdot(w_{21}\cdot x_1+w_{22}\cdot x_2+w_{23}\cdot x_3)$$

$$\boxed{\frac{d}{d\mathbf{x}}(f(g(\mathbf{x})))=\begin{bmatrix}4\cdot w_{11}+3\cdot w_{21}\\4\cdot w_{12}+3\cdot w_{22}\\4\cdot w_{13}+3\cdot w_{23}\end{bmatrix}}$$

Regla de la cadena:

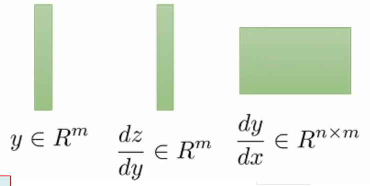
$$\frac{d}{d\mathbf{x}}(f(g(\mathbf{x})))=\begin{bmatrix} w_{11} & w_{12} & w_{13}\\ w_{21} & w_{22} & w_{22}\end{bmatrix}^T\cdot\left.\frac{d\,f(\mathbf{y})}{d\,\mathbf{y}}\right|_{y=g(x)}$$

- Por la regla de la cadena de Cálculo:

$$\frac{d}{d\,\mathbf{x}}(f(g(\mathbf{x})))=\frac{d\,\mathbf{z}}{d\,\mathbf{x}}=\frac{d\,\mathbf{y}}{d\,\mathbf{x}}\cdot\frac{d\,\mathbf{z}}{d\,\mathbf{y}}$$

Jacobiano de *g*

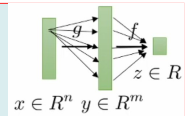
Jacobiano de *f*



Recordatorio: regla de la cadena

Ejemplo:

$$g(\mathbf{x})=\begin{bmatrix} w_{11} & w_{12} & w_{13}\\ w_{21} & w_{22} & w_{22}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} \qquad f(\mathbf{y})=\begin{bmatrix}4 & 3\end{bmatrix}\begin{bmatrix}y_1\\y_2\end{bmatrix}$$



$$f(g(\mathbf{x}))=4\cdot(w_{11}\cdot x_1+w_{12}\cdot x_2+w_{13}\cdot x_3)+3\cdot(w_{21}\cdot x_1+w_{22}\cdot x_2+w_{23}\cdot x_3)$$

$$\boxed{\frac{d}{d\mathbf{x}}(f(g(\mathbf{x})))=\begin{bmatrix}4\cdot w_{11}+3\cdot w_{21}\\4\cdot w_{12}+3\cdot w_{22}\\4\cdot w_{13}+3\cdot w_{23}\end{bmatrix}}$$

Regla de la cadena:

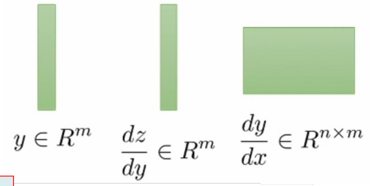
$$\frac{d}{d\mathbf{x}}(f(g(\mathbf{x})))=\begin{bmatrix} w_{11} & w_{12} & w_{13}\\ w_{21} & w_{22} & w_{22}\end{bmatrix}^T\cdot\left.\frac{d\,f(\mathbf{y})}{d\,\mathbf{y}}\right|_{y=g(x)}$$

- Por la regla de la cadena de Cálculo:

$$\frac{d}{d\,\mathbf{x}}(f(g(\mathbf{x})))=\frac{d\,\mathbf{z}}{d\,\mathbf{x}}=\frac{d\,\mathbf{y}}{d\,\mathbf{x}}\cdot\frac{d\,\mathbf{z}}{d\,\mathbf{y}}$$

Jacobiano de *g*

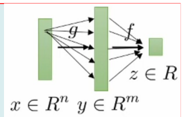
Jacobiano de *f*



Recordatorio: regla de la cadena

Ejemplo:

$$g(\mathbf{x})=\begin{bmatrix} w_{11} & w_{12} & w_{13}\\ w_{21} & w_{22} & w_{22}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} \qquad f(\mathbf{y})=\begin{bmatrix}4 & 3\end{bmatrix}\begin{bmatrix}y_1\\y_2\end{bmatrix}$$



$$f(g(\mathbf{x}))=4\cdot(w_{11}\cdot x_1+w_{12}\cdot x_2+w_{13}\cdot x_3)+3\cdot(w_{21}\cdot x_1+w_{22}\cdot x_2+w_{23}\cdot x_3)$$

$$\boxed{\frac{d}{d\mathbf{x}}(f(g(\mathbf{x})))=\begin{bmatrix}4\cdot w_{11}+3\cdot w_{21}\\4\cdot w_{12}+3\cdot w_{22}\\4\cdot w_{13}+3\cdot w_{23}\end{bmatrix}}$$

Regla de la cadena:

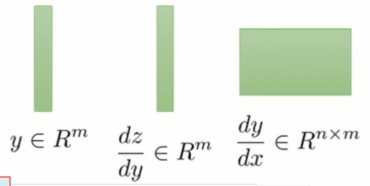
$$\frac{d}{d\mathbf{x}}(f(g(\mathbf{x})))=\begin{bmatrix} w_{11} & w_{12} & w_{13}\\ w_{21} & w_{22} & w_{22}\end{bmatrix}^T\cdot\begin{bmatrix}4\\3\end{bmatrix}$$

- Por la regla de la cadena de Cálculo:

$$\frac{d}{d\,\mathbf{x}}(f(g(\mathbf{x})))=\frac{d\,\mathbf{z}}{d\,\mathbf{x}}=\frac{d\,\mathbf{y}}{d\,\mathbf{x}}\cdot\frac{d\,\mathbf{z}}{d\,\mathbf{y}}$$

Jacobiano de *g*

Jacobiano de *f*



Derivadas con matrices

Function escalar (1D) *f[]* de un vector **a**

$$\mathbf{a}=\begin{bmatrix}a_1\\a_2\\a_3\\a_4\end{bmatrix} \qquad \frac{\partial f}{\partial \mathbf{a}}=\begin{bmatrix}\frac{\partial f}{\partial a_1}\\[4pt]\frac{\partial f}{\partial a_2}\\[4pt]\frac{\partial f}{\partial a_3}\\[4pt]\frac{\partial f}{\partial a_4}\end{bmatrix}$$

Derivadas con matrices

Function escalar (1D) *f[]* de un a matriz **A**

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix}$$

$$\frac{\partial f}{\partial \mathbf{A}} = \begin{bmatrix} \frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} & \frac{\partial f}{\partial a_{13}} \\ \frac{\partial f}{\partial a_{21}} & \frac{\partial f}{\partial a_{22}} & \frac{\partial f}{\partial a_{23}} \\ \frac{\partial f}{\partial a_{31}} & \frac{\partial f}{\partial a_{32}} & \frac{\partial f}{\partial a_{33}} \\ \frac{\partial f}{\partial a_{41}} & \frac{\partial f}{\partial a_{42}} & \frac{\partial f}{\partial a_{43}} \end{bmatrix}$$

Derivadas con matrices

Function vectorial *f[]* de un vector **a**

$$\mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix}$$

$$\frac{\partial \mathbf{f}}{\partial \mathbf{a}} = \begin{bmatrix} \frac{\partial f_1}{\partial a_1} & \frac{\partial f_2}{\partial a_1} & \frac{\partial f_3}{\partial a_1} \\ \frac{\partial f_1}{\partial a_2} & \frac{\partial f_2}{\partial a_2} & \frac{\partial f_3}{\partial a_2} \\ \frac{\partial f_1}{\partial a_3} & \frac{\partial f_2}{\partial a_3} & \frac{\partial f_3}{\partial a_3} \\ \frac{\partial f_1}{\partial a_4} & \frac{\partial f_2}{\partial a_4} & \frac{\partial f_4}{\partial a_4} \end{bmatrix}$$

Vectores vs Matrices

Derivadas escalares:

$$f_3 = \beta_3 + \omega_3 h_3 \qquad \frac{\partial f_3}{\partial h_3} = \frac{\partial}{\partial h_3}(\beta_3 + \omega_3 h_3) = \omega_3$$

Derivadas de matrices

$$\mathbf{f}_3 = \boldsymbol{\beta}_3 + \boldsymbol{\Omega}_3 \mathbf{h}_3 \qquad \frac{\partial \mathbf{f}_3}{\partial \mathbf{h}_3} = \frac{\partial}{\partial \mathbf{h}_3}(\boldsymbol{\beta}_3 + \boldsymbol{\Omega}_3 \mathbf{h}_3) = \boldsymbol{\Omega}_3^T$$

Vectores vs Matrices

Derivadas escalares:

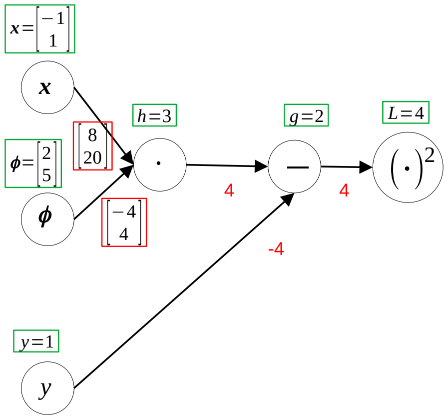
$$f_3 = \beta_3 + \omega_3 h_3 \qquad \frac{\partial f_3}{\partial \beta_3} = \frac{\partial}{\partial \omega_3}\beta_3 + \omega_3 h_3 = 1$$

Derivadas de matrices

$$\mathbf{f}_3 = \boldsymbol{\beta}_3 + \boldsymbol{\Omega}_3 \mathbf{h}_3 \qquad \frac{\partial \mathbf{f}_3}{\partial \boldsymbol{\beta}_3} = \frac{\partial}{\partial \boldsymbol{\beta}_3}(\boldsymbol{\beta}_3 + \boldsymbol{\Omega}_3 \mathbf{h}_3) = \mathbf{I}$$

Grafos de cómputo

¿qué **expresión** calcula el grafo?



$$L(\boldsymbol{x},y;\boldsymbol{\phi})=(\boldsymbol{\phi}^T\boldsymbol{x}-y)^2$$

$$h=\boldsymbol{\phi}^T\boldsymbol{x}$$
$$g=h-y$$
$$L=g^2$$

$$\frac{dL}{dg}=2\cdot g \qquad \frac{dg}{dh}=1 \qquad \frac{dg}{dy}=-1$$

$$\frac{dh}{d\boldsymbol{x}}=\boldsymbol{\phi} \qquad \frac{dh}{d\boldsymbol{\phi}}=\boldsymbol{x}$$

2.4 Cálculo eficiente del gradiente

- Conceptos matemáticos
- Retropropagación del gradiente (Backpropagation)
- Diferenciación algorítmica

Diferenciación algorítmica

- Los frameworks de deep learning calculan derivadas automáticamente
- Únicamente se especifica el modelo y la función de pérdida
- ¿Cómo? Diferenciación Algorítmica
  - Cada componente sabe como calcular su propia derivada
    - Una función lineal "sabe" cómo calcular su derivada de la salida con respecto a su entrada
    - Una función lineal "sabe" cómo calcular su derivada de la salida con respecto a sus parámetros
  - Se especifica el orden de los componentes (grafo de cómputo)
  - Puede calcular la cadena de derivadas

Algoritmo de retropropagación

- Forward: Se ejecuta el grafo de cómputo dadas unas entradas
- Backward:
  - Inicializar $\delta=\frac{dJ}{d\boxed{z^{(n)}}}$

    Salida de la última operación del grafo de cómputo

  - Para cada $\boldsymbol{f}$ con entrada $\boldsymbol{x_f}$ y parámetros $\boldsymbol{\Phi_f}$ desde el final al principio:

Derivada de J con respecto a los parámetros de este nodo

$$\boxed{\frac{df}{d\boldsymbol{\phi}_f}}=\frac{df}{d\boldsymbol{\phi}_f}\cdot\delta$$

$$\delta\leftarrow\frac{df}{d\boldsymbol{x}_f}\cdot\delta$$