



UNIVERSIDAD AUTÓNOMA DE ZACATECAS
"Francisco García Salinas"

Grafos computacionales

Redes Neuronales

Gamaliel Moreno Chávez

`gamalielmch@uaz.edu.mx`

Enero-julio 2021



Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

1 Grafos computacionales

2 Caso escalar

- Propagación hacia adelante
- Propagación hacia atrás
- Operación de un nodo

3 Caso vectorial





Grafos computacionales

Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

- Cálculo de $J(\theta)$ y $\nabla_{\theta}J(\theta)$ es complejo: Concatenación de productos matriz-vector, extensiones para el sesgo y funciones de activación no lineales contribuyen a complejidad.
- En redes modernas, capas varían (función de activación, tipo, tamaño), lo que aumenta complejidad.
- Redes clásicas: $J(\theta)$ y $\nabla_{\theta}J(\theta)$ calculados de previo y plasmados en código fijo para el caso concreto de J .
- Marcos de trabajo (TensorFlow, Torch, ...) son flexibles: tipos de capas y funciones de activación se cambian fácilmente.
- ¿cómo se pueden hacer los cálculos eficientemente?
- ¿cómo puedo agregar mis propios tipos de capas?





Grafos computacionales

Grafos computacionales

Caso escalar

Propagación hacia adelante
Propagación hacia atrás
Operación de un nodo

Caso vectorial

- Cálculo analítico y código correspondiente es complejo, tarda tiempo, y es poco reutilizable.
- Cálculo numérico del gradiente:
 - es excesivamente caro
 - diferencias divididas requieren la evaluación de J 1 o 2 veces por cada parámetro escalar θ
 - es muy fácil y rápido de implementar
 - se usa para verificar cálculo analítico
- Alternativa: grafos computacionales:
 - Similar a la diferenciación automática
 - Aplican regla de la cadena para descomponer cálculo complejo en pasos sencillos
 - Compromiso entre los métodos numéricos y los métodos analíticos
 - Altamente flexibles



Grafos computacionales

Caso escalar

Propagación hacia adelante

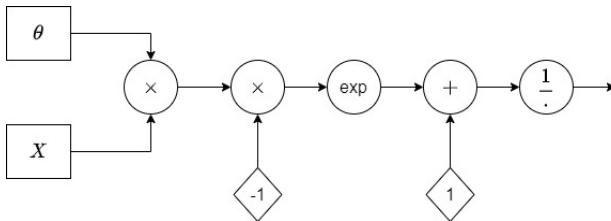
Propagación hacia atrás

Operación de un nodo

Caso vectorial

- idea:
 - 1 Partir evaluación de una función en pasos elementales
 - 2 Cada paso es un nodo del grafo
 - 3 Relación entre pasos se expresa con aristas del grafo
- Explicaremos conceptos con ejemplo:

$$f(x, \theta) = \frac{1}{1 + e^{-\theta x}}$$



Grafos computacionales

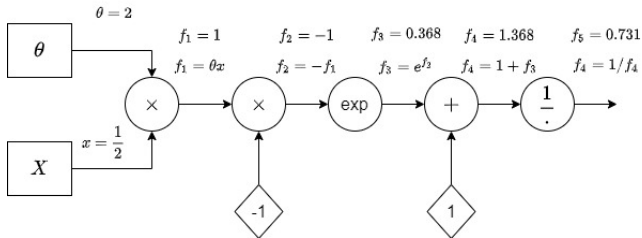
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial





Propagación hacia atrás

Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

- Buscamos ahora $dy/d\theta$ y dy/dx
- Esto se interpreta como el cambio en la salida debido a un cambio en la entrada
- La regla de la cadena subyace el cálculo nodo por nodo:

$$\frac{\partial y}{\partial \theta} = \underbrace{\frac{\partial y}{\partial f_4}}_{b1} \cdot \underbrace{\frac{\partial f_4}{\partial f_3}}_{b2} \cdot \underbrace{\frac{\partial f_3}{\partial f_2}}_{b3} \cdot \underbrace{\frac{\partial f_2}{\partial f_1}}_{b4} \cdot \frac{\partial f_1}{\partial \theta}$$

- Lo calculamos de atrás hacia adelante



Propagación hacia atrás

Grafos computacionales

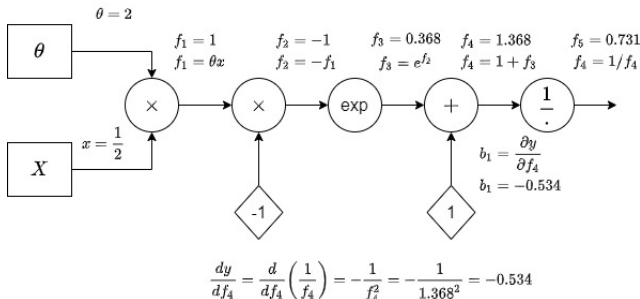
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial



Grafos computacionales

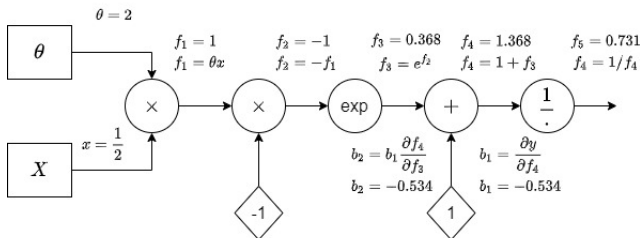
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial



$$\frac{df_4}{df_3} = \frac{d}{df_3} (1 + f_3) = 1$$

Grafos computacionales

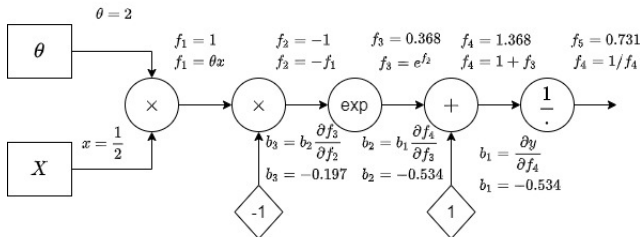
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial



$$\frac{df_3}{df_2} = \frac{d}{df_2}(e^{f_2}) = e^{f_2} = e^{-1} = 0.368$$

Propagación hacia atrás

Grafos computacionales

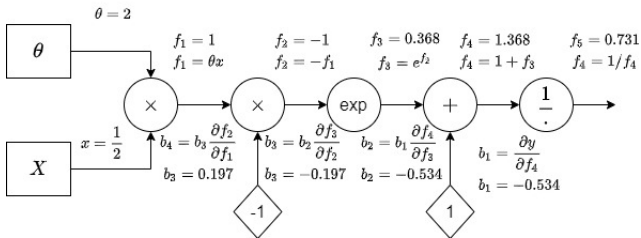
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial



$$\frac{df_2}{df_1} = \frac{d}{df_1}(-f_1) = -1$$



Propagación hacia atrás

Grafos computacionales

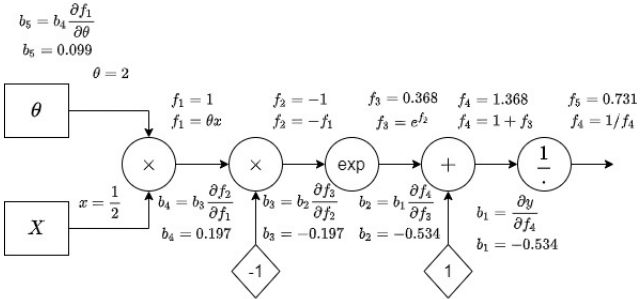
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial



$$\frac{df_1}{d\theta} = \frac{d}{d\theta} \theta x = x$$





Propagación hacia atrás

Grafos computacionales

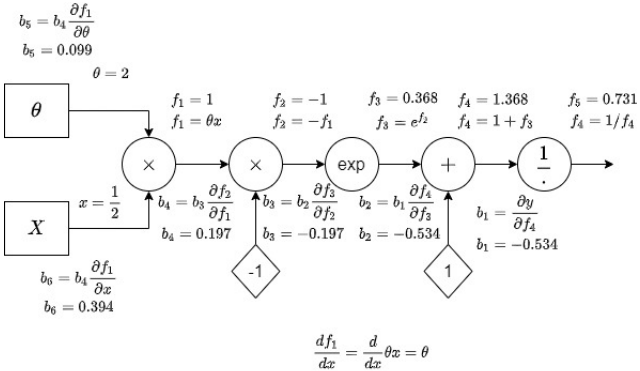
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial





Propagación hacia atrás

Grafos computacionales

Caso escalar

Propagación hacia adelante

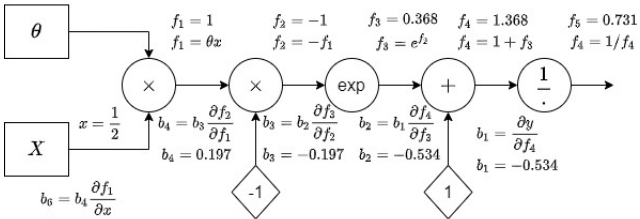
Propagación hacia atrás

Operación de un nodo

Caso vectorial

$$b_5 = b_4 \frac{\partial f_1}{\partial \theta}$$
$$b_5 = 0.099$$

$$\theta = 2$$



$$\frac{dy}{d\theta} = b_5 = b_4 x$$

$$\frac{dy}{dx} = b_6 = b_4 \theta$$





Propagación hacia atrás

Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

- Obsérvese que calculamos $dy/d\theta$ y dy/dx
- con eso, podemos adaptar las entradas en un proceso de descenso de gradiente:

$$\theta' \leftarrow \theta - \lambda \frac{\partial y}{\partial \theta}$$

$$x' \leftarrow x - \lambda \frac{\partial y}{\partial x}$$

- Estos cambios en θ y x asegurarían que $y = f(x', \theta') \leq f(x, \theta)$
- La iteración de este proceso llevará hasta el mínimo (si λ está bien elegido)





Operación de un nodo

Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

- Comportamiento de nodos en escalares
 - Múltiples entradas y una salida
 - Una entrada y múltiples salidas
 - Múltiples entradas y múltiples salidas
- Generalización a vectores y matrices



Operación de un nodo

Grafos computacionales

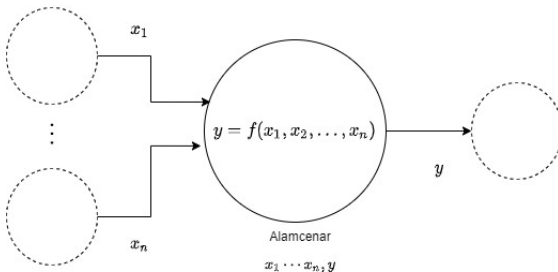
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial



Operación de un nodo

Grafos computacionales

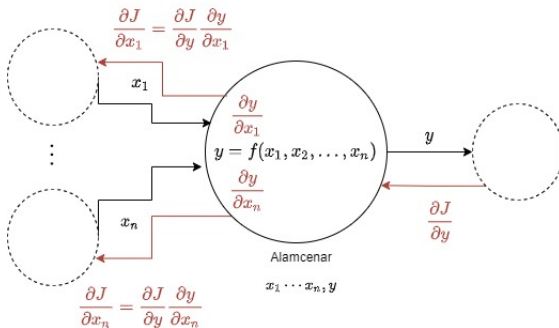
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial



Operación de un nodo

Grafos computacionales

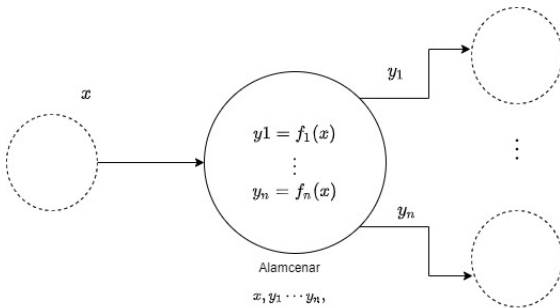
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial



Operación de un nodo

Grafos computacionales

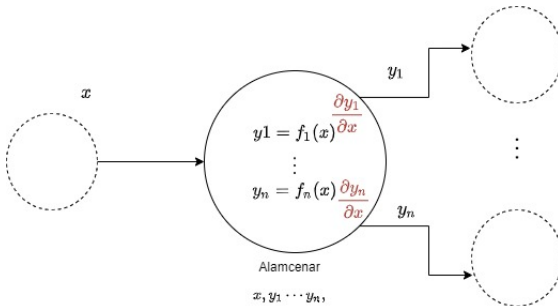
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial



Operación de un nodo

Grafos computacionales

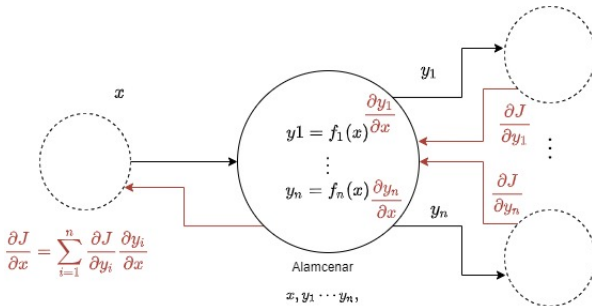
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial



Grafos computacionales

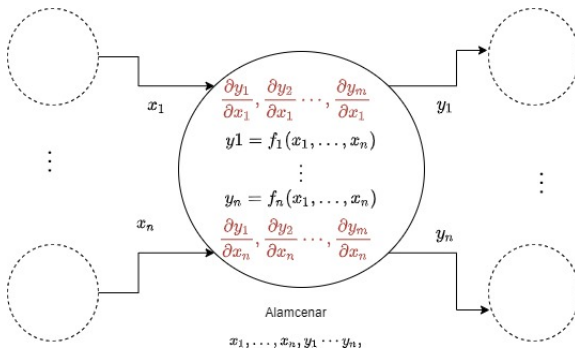
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial





Operación de un nodo

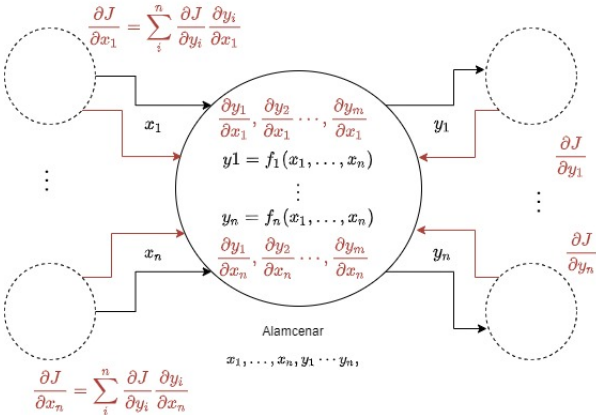
Grafos computacionales

Caso escalar

Propagación hacia adelante
Propagación hacia atrás

Operación de un nodo

Caso vectorial





Usando notación vectorial

Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

- El nodo anterior se simplifica usando notación vectorial

$$\mathbf{x} = [x_1 \quad x_2 \cdots x_n]^T$$

$$\mathbf{y} = [y_1 \quad y_2 \cdots y_m]^T = [f_1(\mathbf{x}) \quad f_2(\mathbf{x}) \cdots f_m(\mathbf{x})]$$

$$\nabla_{\mathbf{x}} J = \left[\frac{\partial J}{\partial x_1} \quad \frac{\partial J}{\partial x_2} \cdots \frac{\partial J}{\partial x_n} \right]^T \quad \nabla_{\mathbf{y}} J = \left[\frac{\partial J}{\partial y_1} \quad \frac{\partial J}{\partial y_2} \cdots \frac{\partial J}{\partial y_m} \right]^T$$

$$\mathbf{D}_{\mathbf{x}} \mathbf{f} = \begin{bmatrix} \nabla_{\mathbf{x}}^T f_1(\mathbf{x}) \\ \nabla_{\mathbf{x}}^T f_2(\mathbf{x}) \\ \vdots \\ \nabla_{\mathbf{x}}^T f_m(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

$$\nabla_{\mathbf{x}} J = \mathbf{D}_{\mathbf{x}}^T \mathbf{f} \nabla_{\mathbf{y}} J$$





Operación de un nodo, vectorial

Grafos computacionales

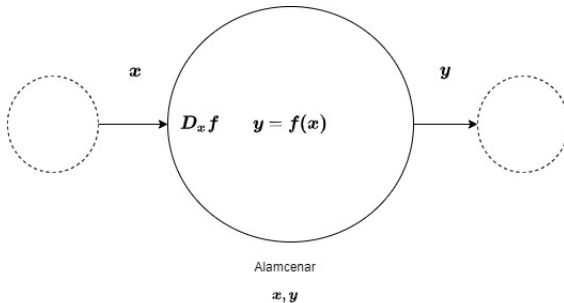
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial





Operación de un nodo, vectorial

Grafos computacionales

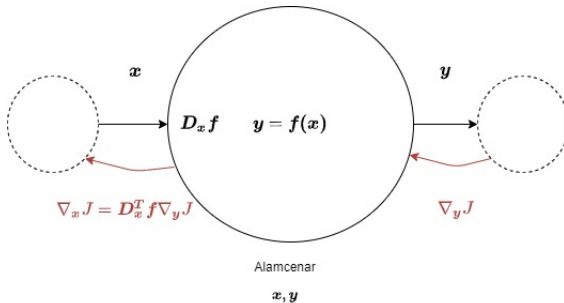
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial





Consideraciones prácticas

Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

- $\mathbf{D}_x^T \mathbf{f}$ es el jacobiano de \mathbf{f} respecto \mathbf{x}
- La ecuación $\nabla_x J = \mathbf{D}_x^T \mathbf{f} \nabla_y J$ integra todas las posibles influencias de cada salida y cada entrada, y aplica la regla de la cadena para ello.
- con frecuencia sólo existen algunas dependencias y el Jacobiano $\mathbf{D}_x^T \mathbf{f}$ es disperso (tiene muchos ceros).
 - Si y_i solo depende de x_i entonces el Jacobiano es diagonal.
- En la práctica se calcula $\nabla_x J$, que es lo que se retropropaga a los nodos precedentes.
- Por eso el nombre: retro-propagación (backpropagation)





Generalización

Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

- Observe que $f(\mathbf{x})$ puede ser cualquier función
- $f(\mathbf{x})$ agrupa varios cálculos en un nodo
- Granularidad según complejidad deseada para cada nodo
- Resumiendo: cada nodo tiene tres tareas:
 - propagar entrada a salida
 - almacenar valores temporales para calcular gradientes
 - calcula gradiente de J respecto a entradas del nodo, combinando información del gradiente J respecto a las salidas y derivadas/gradientes de salidas respecto a entradas
- Esto último es conceptual. En la práctica se resume el cálculo
- Si \mathbf{x} o \mathbf{y} fuese matrices, se analiza cada elemento de la matriz por separado, $\nabla_{\mathbf{x}}J$ o $\nabla_{\mathbf{y}}J$ serían matrices.



Ejemplo: Sigmoide

Grafos computacionales

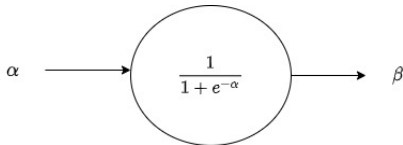
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial



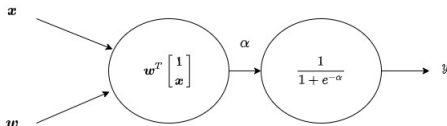
- Entrada: α
- Salida: β
- Cálculo en propagación hacia adelante: $\frac{1}{1+e^{-\alpha}}$
- Cálculo en propagación hacia atrás:

$$\begin{aligned}
 \frac{\partial \beta}{\partial \alpha} &= \frac{e^{-\alpha}}{(1 + e^{-\alpha})^2} = \frac{e^{-\alpha}}{(1 + e^{-\alpha})} + \frac{1}{(1 + e^{-\alpha})} \\
 &= \frac{1 + e^{-\alpha} - 1}{(1 + e^{-\alpha})} \beta = \left(\frac{1 + e^{-\alpha}}{1 + e^{-\alpha}} - \frac{1}{1 + e^{-\alpha}} \right) = (1 - \beta) \beta
 \end{aligned}$$

- Basta con almacenar β para calcular $d\beta/d\alpha$



Ejemplo: Neurona totalmente conectada con sesgo



- primer nodo extiende vector \mathbf{x} con 1 para poder aplicar el sesgo w_0 (primer componente de \mathbf{w})

$$\alpha = \begin{pmatrix} w_0 \\ + w_1 x_1 \\ + w_2 x_2 \\ \vdots \\ + w_n x_n \end{pmatrix} \Rightarrow \mathbf{D}_w^T \alpha = \nabla_w \alpha = \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix}$$

$$\mathbf{D}_x^T \alpha = \nabla_x \alpha = \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$$

Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial





Ejemplo: Neurona totalmente conectada con sesgo

Grafos computacionales

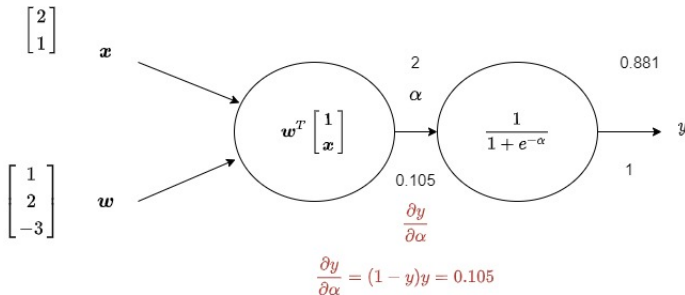
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial





Ejemplo: Neurona totalmente conectada con sesgo

Grafos computacionales

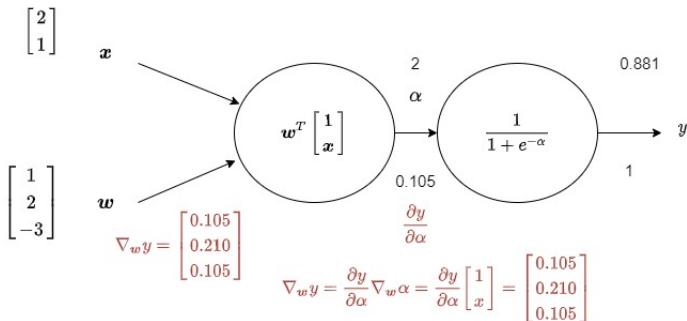
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial





Ejemplo: Neurona totalmente conectada con sesgo

Grafos computacionales

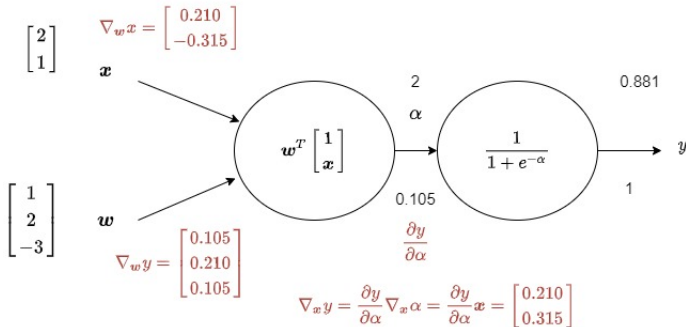
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial





Compuertas

Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

- Los nodos de un grafo computacional se comportan como compuertas
- A continuación se analizarán algunos casos comunes





Compuertas

Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

- Los nodos de un grafo computacional se comportan como compuertas
- A continuación se analizarán algunos casos comunes



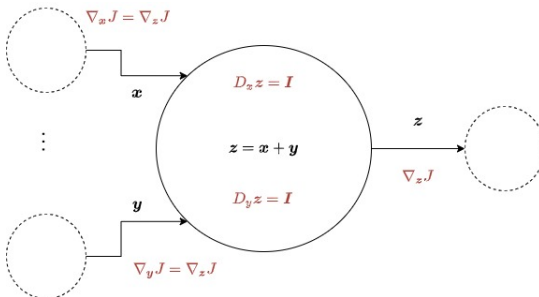
Grafos computacionales

Caso escalar

Propagación hacia adelante
Propagación hacia atrás
Operación de un nodo

Caso vectorial

- La suma distribuye el gradiente a todas las entradas





Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

- El nodo $z = \max(x, y)$ produce el máximo de variables de entrada
- la pregunta es ¿qué hace la derivada?
- la interpretación es la siguiente
 - si $x > y$, entonces $z = x$ y por tanto $\frac{\partial z}{\partial x} = 1$ y $\frac{\partial z}{\partial y} = 0$
 - si $x < y$, entonces $z = y$ y por tanto $\frac{\partial z}{\partial x} = 0$ y $\frac{\partial z}{\partial y} = 1$
- La compuerta \max enruta entonces el gradiente por aquella entrada con el máximo valor
- Esto se generaliza fácilmente a vectores componente por componente.





Grafos computacionales

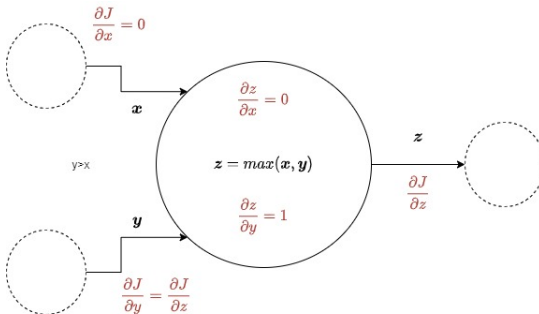
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial





Conmutadores

Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

- Los distintos tipos de producto intercambian de algún modo u otro las entradas en la propagación hacia atrás.
- Solo en algunos casos es posible calcular el jacobiano
- A continuación analizaremos cuatro casos:
 - Producto de dos escalares $z = xy$
 - Producto interno de dos vectores $z = \mathbf{x}^T \mathbf{y}$
 - Producto externo de dos vectores $\mathbf{Z} = \mathbf{x} \mathbf{y}^T$
 - Producto de dos matrices $\mathbf{Z} = \mathbf{X} \mathbf{Y}$





Producto escalar

Grafos computacionales

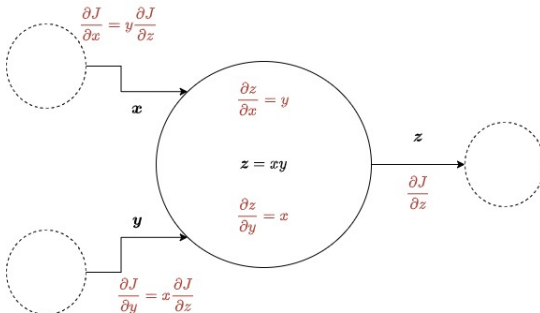
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial



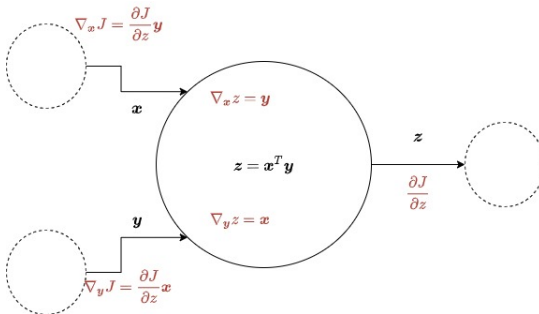
Producto interno de dos vectores

Grafos computacionales

Caso escalar

Propagación hacia adelante
Propagación hacia atrás
Operación de un nodo

Caso vectorial





Producto externo de vectores

Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

- El producto externo produce una matriz a la salida
- Supongamos $\mathbf{x} \in \mathbb{R}^n$ e $\mathbf{y} \in \mathbb{R}^m$

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1m} \\ z_{21} & z_{22} & \cdots & z_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nm} \end{bmatrix} = \mathbf{xy}^T = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_m \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_m \\ \vdots & \vdots & \ddots & \vdots \\ x_n y_1 & x_n y_2 & \cdots & x_n y_m \end{bmatrix}$$

- siguiendo el principio de múltiples salidas (elementos de \mathbf{Z})

$$\frac{\partial J}{\partial x_k} = \sum_i \sum_j \frac{\partial J}{\partial z_{ij}} \frac{\partial z_{ij}}{\partial x_k} = \sum_i \sum_j \frac{\partial J}{\partial z_{ij}} \frac{\partial (x_i y_j)}{\partial x_k}$$

donde solo los términos con $i = k$ sobreviven

$$\frac{\partial J}{\partial x_k} = \sum_j \frac{\partial J}{\partial z_{kj}} y_j \Rightarrow \nabla_{\mathbf{x}} J = \nabla_{\mathbf{z}} J \mathbf{y}$$





Producto externo de vectores

Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

- De forma equivalente para \mathbf{y}

$$\frac{\partial J}{\partial y_k} = \sum_i \sum_j \frac{\partial J}{\partial z_{ij}} \frac{\partial z_{ij}}{\partial y_k} = \sum_i \sum_j \frac{\partial J}{\partial z_{ij}} \frac{\partial (x_i y_j)}{\partial y_k}$$

donde solo los términos con $j = k$ sobreviven

$$\frac{\partial J}{\partial y_k} = \sum_j \frac{\partial J}{\partial z_{kj}} x_j \Rightarrow \nabla_{\mathbf{y}} J = (\nabla_{\mathbf{z}} J)^T \mathbf{x}$$





Producto externo de vectores

Grafos computacionales

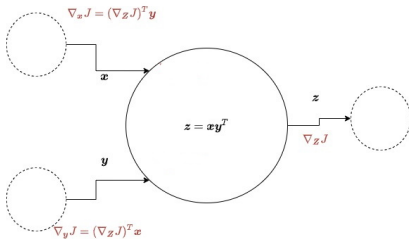
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial





Producto de matrices

Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

■ Sean $\mathbf{X} \in \mathbb{R}^{r \times c}$, $\mathbf{Y} \in \mathbb{R}^{c \times d}$, $\mathbf{Z} = \mathbf{XY} \in \mathbb{R}^{r \times d}$

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1m} \\ z_{21} & z_{22} & \cdots & z_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nm} \end{bmatrix} = \mathbf{XY}$$

$$\begin{bmatrix} \mathbf{x}_{1,:}^T \mathbf{y}_{:,1} & \mathbf{x}_{1,:}^T \mathbf{y}_{:,2} & \cdots & \mathbf{x}_{1,:}^T \mathbf{y}_{:,m} \\ \mathbf{x}_{2,:}^T \mathbf{y}_{:,1} & \mathbf{x}_{2,:}^T \mathbf{y}_{:,2} & \cdots & \mathbf{x}_{2,:}^T \mathbf{y}_{:,m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{n,:}^T \mathbf{y}_{:,1} & \mathbf{x}_{n,:}^T \mathbf{y}_{:,2} & \cdots & \mathbf{x}_{n,:}^T \mathbf{y}_{:,m} \end{bmatrix}$$

$$z_{ij} = \mathbf{x}_{i,:}^T \mathbf{y}_{:,j} = \sum_{k=1}^c x_{ik} y_{kj}$$





Producto de matrices

Grafos computacionales

Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial

- Siguiendo el principio de múltiples salidas (elementos de \mathbf{Z})

$$\frac{\partial J}{\partial x_{lm}} = \sum_{i=1}^r \sum_{j=1}^d \frac{\partial J}{\partial z_{ij}} \frac{\partial z_{ij}}{\partial x_{lm}}$$

donde solo los términos con $i = l$ y $k = m$ sobreviven

$$\frac{\partial J}{\partial x_{lm}} = \sum_{j=1}^d \frac{\partial J}{\partial z_{lj}} y_{mj} \Rightarrow \nabla_{\mathbf{x}} J = \nabla_{\mathbf{z}} J \mathbf{Y}^T$$



Grafos computacionales

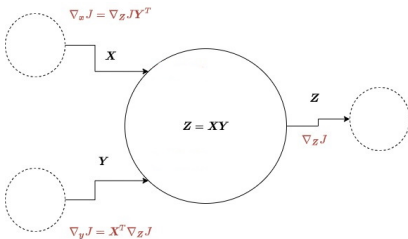
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial





Producto matriz vector

Grafos computacionales

Caso escalar

Propagación hacia adelante
Propagación hacia atrás
Operación de un nodo

Caso vectorial

- Un caso particular del producto de matrices es el producto matriz vector

$$\mathbf{z} = \mathbf{X}\mathbf{y}$$

- Simplemente tómese $\mathbf{Y} = \mathbf{y}$
- En ese caso, $\nabla_{\mathbf{z}}J$ es un vector
 - $\nabla_{\mathbf{y}}J = \mathbf{X}^T \nabla_{\mathbf{z}}J$ es un vector
 - $\nabla_{\mathbf{x}}J = \nabla_{\mathbf{z}}J \mathbf{y}^T$ es una matriz (producto externo)





Producto de matrices

Grafos computacionales

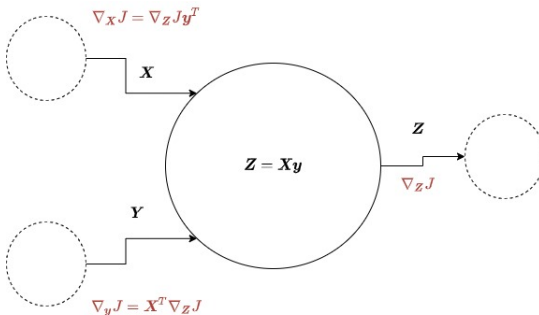
Caso escalar

Propagación hacia adelante

Propagación hacia atrás

Operación de un nodo

Caso vectorial



Gracias



Gamaliel Moreno Chávez