

# Diferenciación Automática

Daniel Corzo García

# Métodos

- ▶ Para realizar la diferenciación, existen cuatro posibilidades:
  - ▶ A mano
  - ▶ Derivación numérica
  - ▶ Derivación simbólica
  - ▶ Diferenciación automática

# Métodos

- ▶ La derivación simbólica o la derivación a mano hacen uso de las reglas de derivación para encontrar la solución de una forma totalmente mecánica.
- ▶ Un ejemplo de las reglas sería:

$$\frac{d}{dx} (f(x) + g(x)) \rightsquigarrow \frac{d}{dx} f(x) + \frac{d}{dx} g(x)$$
$$\frac{d}{dx} (f(x) g(x)) \rightsquigarrow \left( \frac{d}{dx} f(x) \right) g(x) + f(x) \left( \frac{d}{dx} g(x) \right) .$$

# Métodos

- ▶ La derivación numérica busca el valor de la derivada en un punto mediante una aproximación.
- ▶ Un ejemplo de un método básico sería:

$$\frac{\partial f(\mathbf{x})}{\partial x_i} \approx \frac{f(\mathbf{x} + h\mathbf{e}_i) - f(\mathbf{x})}{h}$$

# Métodos

- ▶ Calcular las derivadas a mano tiene el inconveniente de que requiere bastante tiempo y es propenso a errores.
- ▶ La derivación numérica produce errores por redondeo/truncado y no escala bien a la hora de calcular los gradientes. (al menos  $O(n)$  para calcular un gradiente de  $n$  componentes)
- ▶ La derivación simbólica tiene el problema de producir expresiones con un tamaño excesivo, que dificulta su uso posterior. Además, impone condiciones estrictas sobre el formato de entrada que limitan la expresividad.

# Diferenciación Automática

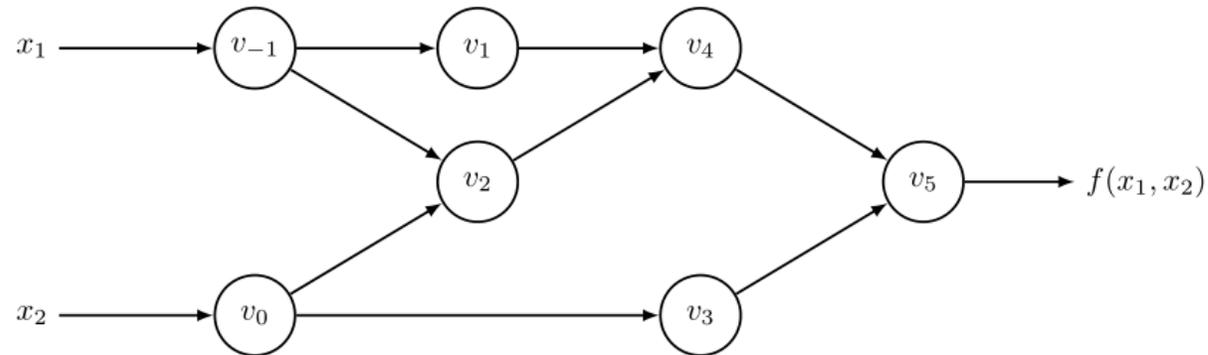
- ▶ Es una mezcla entre la derivación simbólica y la numérica, que busca aprovechar sus ventajas y evitar sus limitaciones.
- ▶ A diferencia de la derivación numérica, produce un resultado con la precisión máquina donde se utilice y es eficiente de forma asintótica.
- ▶ A diferencia de la derivación simbólica, permite elementos externos como condiciones, saltos y bucles, y no produce expresiones de salida enormes.

# Conceptos previos

- ▶ Una función puede verse como una composición de operaciones elementales para las cuales su derivada es conocida.
- ▶ Las operaciones de dicha función generan una traza que representa un grafo computacional.
- ▶ Para el grafo vamos a usar una notación con tres elementos: variables de entrada, variables intermedias y variables de salida.

# Ejemplo de grafo

Para  $f(x_1, x_2) = \ln(x_1) + x_1x_2 - \sin(x_2)$



## Forward Evaluation Trace

$v_{-1}$	$= x_1$	$= 2$
$v_0$	$= x_2$	$= 5$
<hr/>		
$v_1$	$= \ln v_{-1}$	$= \ln 2$
$v_2$	$= v_{-1} \times v_0$	$= 2 \times 5$
$v_3$	$= \sin v_0$	$= \sin 5$
$v_4$	$= v_1 + v_2$	$= 0.693 + 10$
$v_5$	$= v_4 - v_3$	$= 10.693 + 0.959$
<hr/>		
$y$	$= v_5$	$= 11.652$

# DA hacia delante con acumulación

- ▶ Es la forma más sencilla de implementar la diferenciación automática.
- ▶ Consiste en derivar cada nodo del grafo, de forma que al recorrerlo con distintas configuraciones de entrada nos den las derivadas parciales. Ejemplo:

Forward Evaluation Trace	Forward Derivative Trace
$v_{-1} = x_1 = 2$	$\dot{v}_{-1} = \dot{x}_1 = 1$
$v_0 = x_2 = 5$	$\dot{v}_0 = \dot{x}_2 = 0$
<hr/>	<hr/>
$v_1 = \ln v_{-1} = \ln 2$	$\dot{v}_1 = \dot{v}_{-1}/v_{-1} = 1/2$
$v_2 = v_{-1} \times v_0 = 2 \times 5$	$\dot{v}_2 = \dot{v}_{-1} \times v_0 + \dot{v}_0 \times v_{-1} = 1 \times 5 + 0 \times 2$
$v_3 = \sin v_0 = \sin 5$	$\dot{v}_3 = \dot{v}_0 \times \cos v_0 = 0 \times \cos 5$
$v_4 = v_1 + v_2 = 0.693 + 10$	$\dot{v}_4 = \dot{v}_1 + \dot{v}_2 = 0.5 + 5$
$v_5 = v_4 - v_3 = 10.693 + 0.959$	$\dot{v}_5 = \dot{v}_4 - \dot{v}_3 = 5.5 - 0$
<hr/>	<hr/>
$y = v_5 = 11.652$	$\dot{y} = \dot{v}_5 = 5.5$

# DA hacia delante con acumulación

- ▶ Cada recorrido del grafo nos proporciona la derivada para una de las variables de entrada:

$$\dot{y}_j = \left. \frac{\partial y_j}{\partial x_i} \right|_{\mathbf{x}=\mathbf{a}}, \quad j = 1, \dots, m$$

- ▶ El resultado es una de las columnas de la matriz jacobiana:

$$\mathbf{J}_f \mathbf{r} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$$

# DA hacia delante con acumulación

- ▶ Se puede aprovechar cada recorrido para realizar al mismo tiempo una multiplicación vectorial, simplemente inicializando  $\dot{x} = r$  tal que:

$$\mathbf{J}_f \mathbf{r} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$$

- ▶ DA hacia delante es preferible cuando  $m > n$ , pero cuando  $n \gg m$  preferiremos el siguiente método.

# DA hacia atrás con acumulación

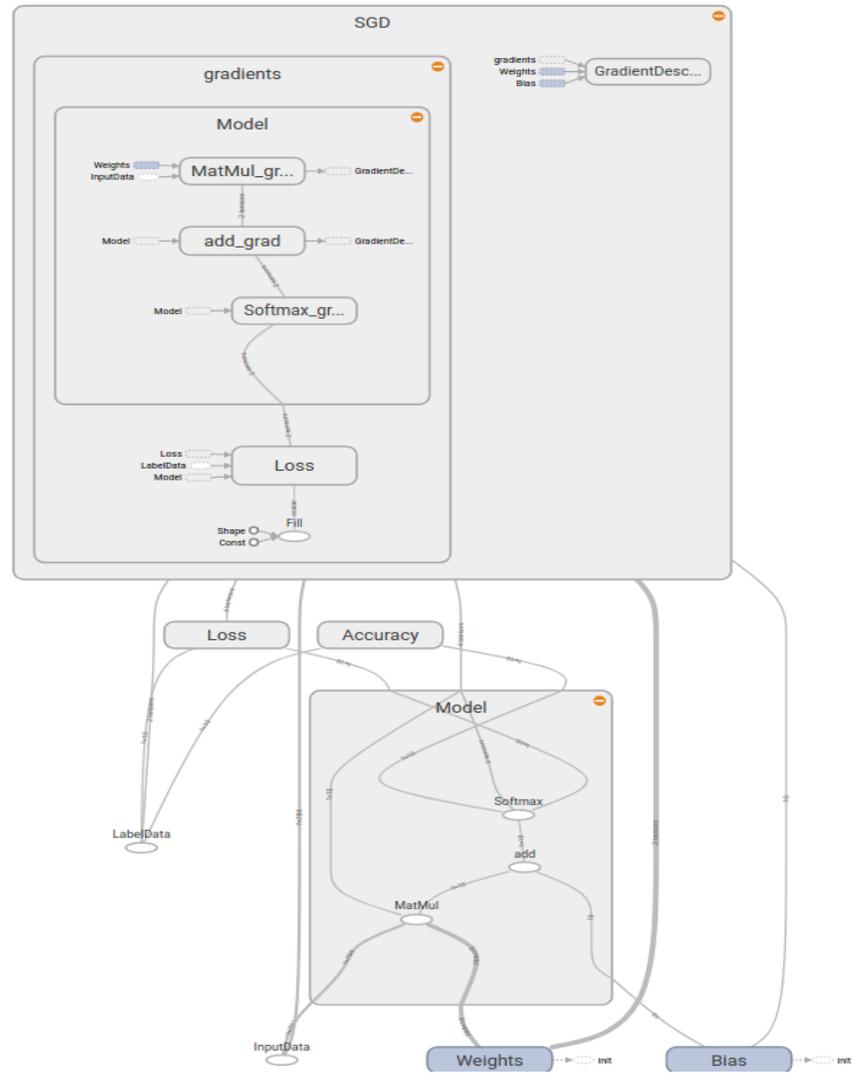
- ▶ Es una forma general del algoritmo de propagación hacia atrás.
- ▶ Primero se recorre el grafo de la función y se guardan los valores de cada variable y las dependencias entre nodos.
- ▶ Después el grafo se recorre empezando por el final y las derivadas de cada variable se calculan respecto a la variable que le antecede. (en orden inverso)
- ▶ De esa forma, el valor de la derivada de un nodo puede verse como la contribución de dicho nodo al cambio del siguiente.

# DA hacia atrás con acumulación

- Un ejemplo usando también  $f(x_1, x_2) = \ln(x_1) + x_1x_2 - \sin(x_2)$

Forward Evaluation Trace	Reverse Adjoint Trace
$v_{-1} = x_1 = 2$	$\bar{x}_1 = \bar{v}_{-1} = 5.5$
$v_0 = x_2 = 5$	$\bar{x}_2 = \bar{v}_0 = 1.716$
$v_1 = \ln v_{-1} = \ln 2$	$\bar{v}_{-1} = \bar{v}_{-1} + \bar{v}_1 \frac{\partial v_1}{\partial v_{-1}} = \bar{v}_{-1} + \bar{v}_1 / v_{-1} = 5.5$
$v_2 = v_{-1} \times v_0 = 2 \times 5$	$\bar{v}_0 = \bar{v}_0 + \bar{v}_2 \frac{\partial v_2}{\partial v_0} = \bar{v}_0 + \bar{v}_2 \times v_{-1} = 1.716$
$v_3 = \sin v_0 = \sin 5$	$\bar{v}_{-1} = \bar{v}_2 \frac{\partial v_2}{\partial v_{-1}} = \bar{v}_2 \times v_0 = 5$
$v_4 = v_1 + v_2 = 0.693 + 10$	$\bar{v}_0 = \bar{v}_3 \frac{\partial v_3}{\partial v_0} = \bar{v}_3 \times \cos v_0 = -0.284$
$v_5 = v_4 - v_3 = 10.693 + 0.959$	$\bar{v}_2 = \bar{v}_4 \frac{\partial v_4}{\partial v_2} = \bar{v}_4 \times 1 = 1$
$y = v_5 = 11.652$	$\bar{v}_1 = \bar{v}_4 \frac{\partial v_4}{\partial v_1} = \bar{v}_4 \times 1 = 1$
	$\bar{v}_3 = \bar{v}_5 \frac{\partial v_5}{\partial v_3} = \bar{v}_5 \times (-1) = -1$
	$\bar{v}_4 = \bar{v}_5 \frac{\partial v_5}{\partial v_4} = \bar{v}_5 \times 1 = 1$
	$\bar{v}_5 = \bar{y} = 1$

# Ejemplo con Tensorflow



# Características

- ▶ La complejidad del modo hacia atrás solo depende de la cantidad de variables de salida, mientras que el modo hacia delante solo depende de las de entrada.
- ▶ Igual que hemos visto antes, se le pueden dar valores iniciales para calcular a la misma vez un producto vectorial.
- ▶ Ambos métodos permiten que el grafo computacional sea dinámico, permitiendo operaciones de control y otros elementos típicos del código.

# Aplicaciones

- ▶ Redes neuronales y backpropagation: Son modelos que dependen de numerosas entradas, que requieren por si mismos de recorrer el grafo hacia delante y que suelen definirse por partes en el código. Todo esto hace la DA hacia atrás idónea para esta clase de problemas.
- ▶ Aquellas tareas de visión por computador que requieren de matrices jacobianas o matrices hessianas: reducción de ruido, segmentación, registración, etc. pueden beneficiarse del uso de DA y suelen mostrar mejor rendimiento.
- ▶ Los modelos estadísticos dentro del procesado del lenguaje natural que se entrenan usando descenso por gradiente son también candidatos a emplear la DA para mejorar el coste del entrenamiento y simplificar su implementación.

# Implementación

- ▶ Tres modos principales:
  - ▶ Usando una librería que proporcione métodos-operadores con los que construir a mano las funciones.
  - ▶ Extendiendo la toolchain de algún lenguaje con el uso de un preprocesador que se encargue de descomponer las funciones en las operaciones elementales y sus derivadas.
  - ▶ Usando sobrecarga de operadores para realizar los pasos necesarios de la DA cuando se empleen tipos específicos.

The background features abstract, overlapping geometric shapes in various shades of green, ranging from light lime to dark forest green. These shapes are primarily located on the right side of the frame, creating a modern, layered effect. The rest of the background is plain white.

**FIN**

Preguntas