Análisis Matemático para Inteligencia Artificial

Martín Errázquin (merrazquin@fi.uba.ar)

Especialización en Inteligencia Artificial

Diferenciación automática

Regla de la Cadena en forma matricial

$f, g : \mathbb{R} \to \mathbb{R}$

$z = g(y)$
$y = f(x)$

$$\frac{dz}{dx} = \underbrace{\frac{dz}{dy}(y)}_{\substack{\mathbb{R} \\ 1\times 1}} \cdot \underbrace{\frac{dy}{dx}(x)}_{\substack{\mathbb{R} \\ 1\times 1}}$$

Sea $f(x_1(s,t), x_2(s,t))$

$$\begin{pmatrix} s \\ t \end{pmatrix} \xrightarrow{\vec{x}} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \xrightarrow{f} f$$

$J_{\vec{x}}(s,t)$ 2×2

$\nabla_f(x_1,x_2)$ 1×2

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial s} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial s}$$

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t}$$

Y luego

$\nabla_f(x_1,x_2) \cdot$ $J_{\vec{x}}(s,t)$

$$\nabla_f(s,t) = \frac{df}{d(s,t)} = \frac{df}{dx}\frac{dx}{d(s,t)} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix} \begin{bmatrix} \frac{\partial x_1}{\partial s} & \frac{\partial x_1}{\partial t} \\ \frac{\partial x_2}{\partial s} & \frac{\partial x_2}{\partial t} \end{bmatrix}$$

Nota: Observar que si bien $\frac{df}{d(s,t)} \in \mathbb{R}^2$, hay que pasar por $\frac{dx}{d(s,t)} \in \mathbb{R}^{2\times 2}$. Podría interesarnos *compilar* el cálculo analítico para ser más eficientes en cómputo.

Derivadas y código

Situación: parte de mi código involucra, para un cierto valor de entrada x, calcular una función `f(x)` y su derivada `df(x)` en ese punto.

```
def f(x):
    return x^2
```

- Cálculo 100 % analítico
  - (+) máxima eficiencia de cómputo
  - (−) máxima inflexibilidad de código

```
def df(x):
    return 2x
```

- Cálculo 100 % numérico
  - (−) máxima ineficiencia de cómputo
  - (+) máxima flexibilidad de código

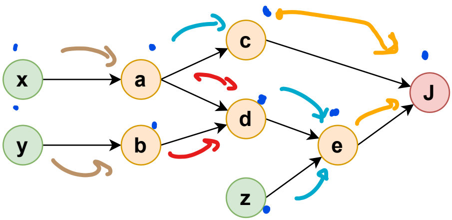
```
def df(x, h=1e-3):
    return (f(x+h) - f(x-h)) / 2h
```

$$\texttt{return}\ \frac{f(x+h)-f(x-h)}{2h}$$

Grafo de cómputo

- Uso específico: composición de funciones de diversas entradas y me importa la derivada de la salida "final" respecto de cada entrada.
- Punto medio entre eficiencia y flexibilidad a través de *bloques diferenciables*.
- Cada bloque debe poder calcular la derivada de su salida respecto de todas sus entradas (en gral. analítica).
- Un "orquestador" se ocupa de ir aplicando regla de la cadena.
- ¡Cada bloque podría ser (adentro) una función compuesta!

$$\frac{\partial J}{\partial x} = \left(\frac{\partial J}{\partial c} \cdot \frac{\partial c}{\partial a} + \frac{\partial J}{\partial e} \cdot \frac{\partial e}{\partial d} \cdot \frac{\partial d}{\partial a} \right) \cdot \frac{\partial a}{\partial x}$$



Ejemplo (simplificado) de código

```
class Product(BaseOperation):
```

$\mathbb{R}^2 \to \mathbb{R}$

```
  def f(self, x, y):
      return x * y
```

$f(x,y) = xy$

```
  def df(self, x, y):
      return (y, x)
```

$\nabla_f(x,y) = (y, x)$

```
class Cosine(BaseOperation):
```

$\mathbb{R} \to \mathbb{R}$

```
  def f(self, x):
      return np.cos(x)
```

$f(x) = \cos(x)$

```
  def df(self, x):
      return -np.sin(x)
```

$\nabla_f(x) = -\sin(x)$

Sea $z(x, y) = \underbrace{x^2}_{a} \cdot e^{\underbrace{x+\overbrace{\cos(y)}^{b}}_{c}}$, quiero $\nabla_z(2, \frac{\pi}{2})$.

x, y

$a = x^2 \qquad \frac{\partial a}{\partial x} = 2x$

$b = \cos(y) \qquad \frac{\partial b}{\partial y} = -\mathrm{sen}(y)$

$c = x + b \qquad \nabla_c(x, b) = (1, 1)$

$d = \exp(c) \qquad \frac{\partial d}{\partial c} = \exp(c)$

$z = a \cdot d \qquad \nabla_z(a, d) = (d, a)$

$x = 2, \; y = \pi/2$

$a = 4$

$b = 0$

$c = 2 + 0 = 2$

$d = e^2 \approx 7{,}4$

$z = 4 \cdot 7{,}4 = 29{,}6$

forward

backwards

$dx = da \cdot \frac{\partial a}{\partial x} + dc \cdot \frac{\partial c}{\partial x} = 7{,}4 \cdot 2 \cdot 2 + 29{,}6 \cdot 1 = 59{,}2$

$dy = db \cdot \frac{\partial b}{\partial y} = 29{,}6 \cdot -1 = -29{,}6$

$da = dz \cdot \frac{\partial z}{\partial a} = 1 \cdot 7{,}4 = 7{,}4$

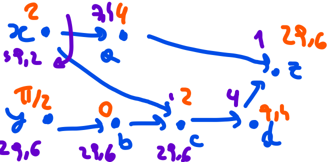
$db = dc \cdot \frac{\partial c}{\partial b} = 29{,}6 \cdot 1 = 29{,}6$

$dc = dd \cdot \frac{\partial d}{\partial c} = 4 \cdot 7{,}4 = 29{,}6$

$dd = 1 \cdot 4 = 1 \cdot 4 = 4$

$dz = \frac{\partial z}{\partial z} = 1$

$z(2, \frac{\pi}{2}) = 29{,}6$



$di = \frac{\partial z}{\partial i}$

$\nabla_z(2, \frac{\pi}{2}) = (59{,}2, \; -29{,}6)$