Modelos de Deep Learning

Derivadas

Universidad ORT Uruguay

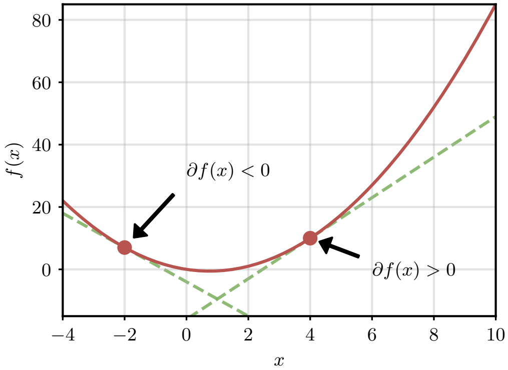
18 de Agosto, 2025

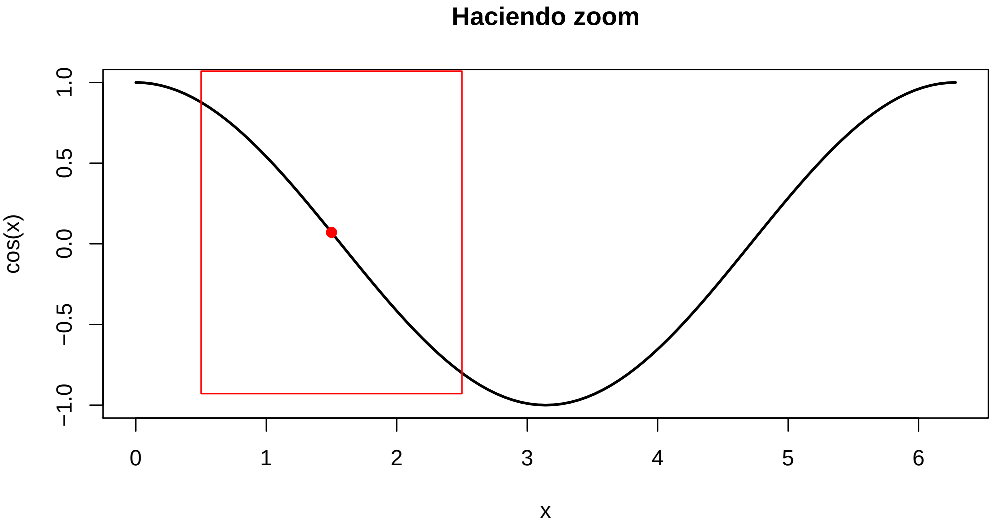
Repaso: definición de derivada

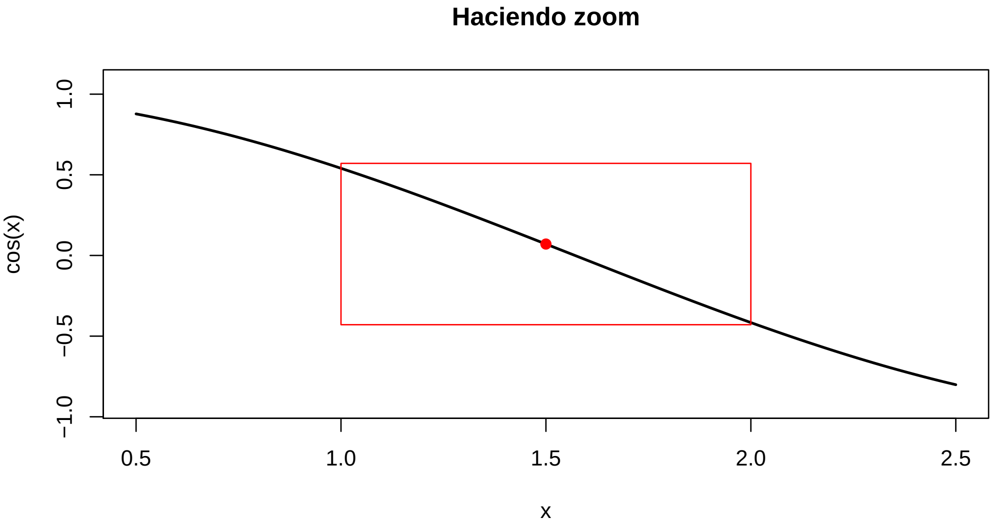
Para una función $y = f(x)$ con entrada escalar y salida escalar, su derivada es

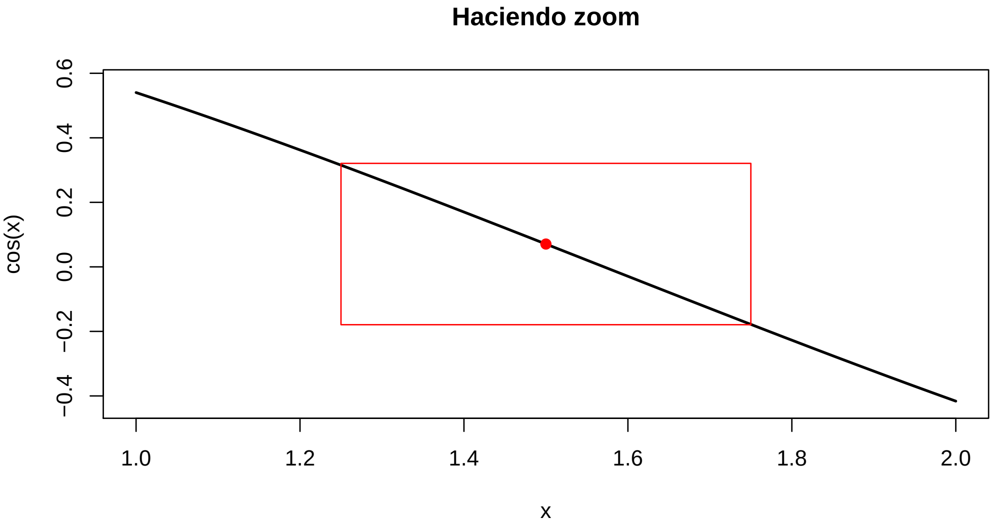
$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}$$

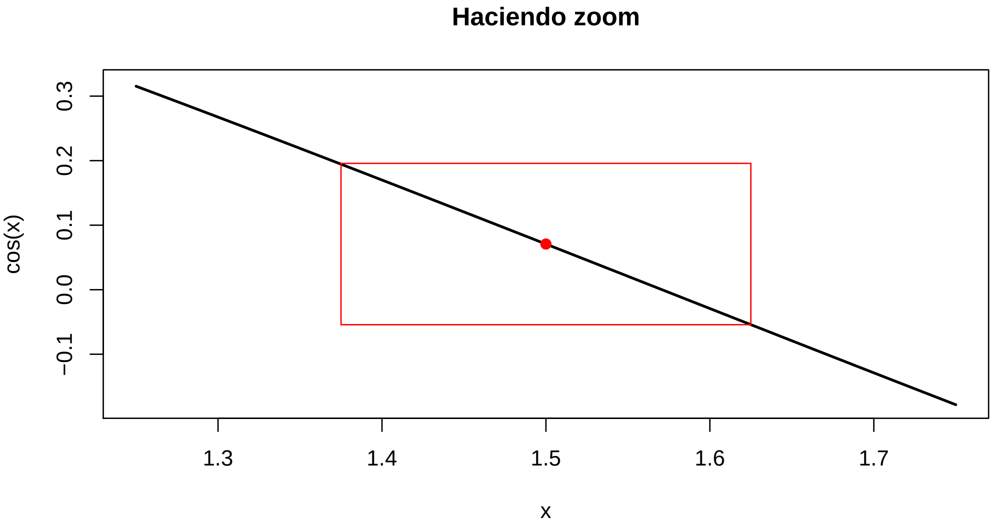
Otra notación: $\partial f(x)$

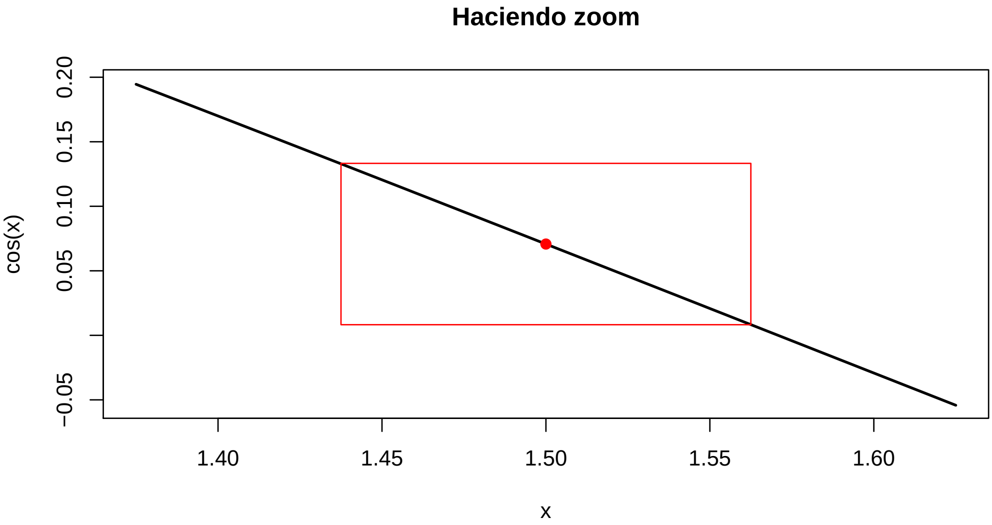


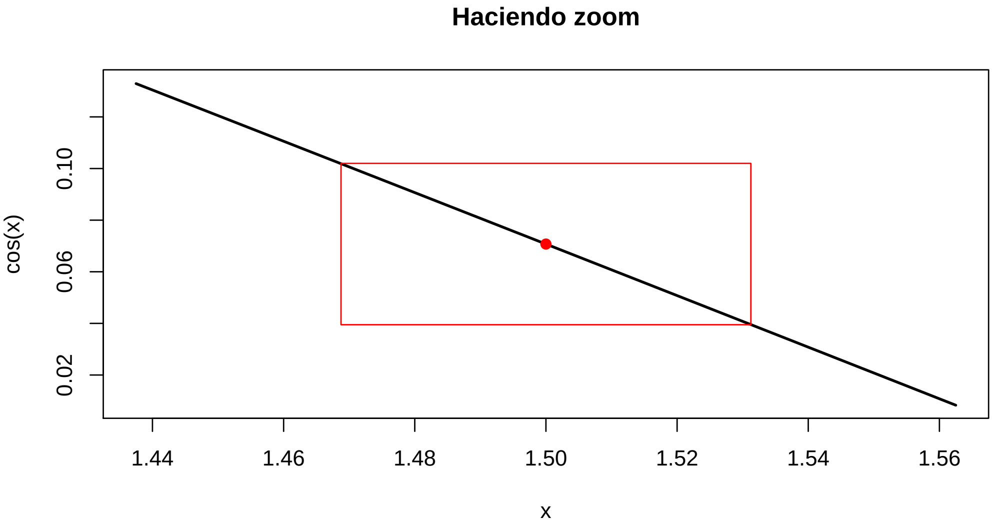

Haciendo zoom
1.0
0.5
0.0
−0.5
−1.0
cos(x)
0
1
2
3
4
5
6
x


Haciendo zoom
cos(x)
1.0
0.5
0.0
−0.5
−1.0
0.5
1.0
1.5
2.0
2.5
x


Haciendo zoom
cos(x)
x
0.6
0.4
0.2
0.0
−0.2
−0.4
1.0
1.2
1.4
1.6
1.8
2.0


Haciendo zoom
cos(x)
0.3
0.2
0.1
0.0
−0.1
1.3
1.4
1.5
1.6
1.7
x

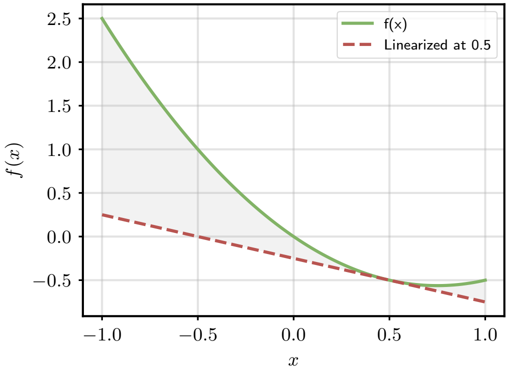

Haciendo zoom
cos(x)
0.20
0.15
0.10
0.05
−0.05
1.40
1.45
1.50
1.55
1.60
x


Haciendo zoom
cos(x)
0.02
0.06
0.10
1.44
1.46
1.48
1.50
1.52
1.54
1.56
x

Aproximación lineal

Para $x \approx x_0$:

$$f(x) \approx f(x_0) + \partial f(x_0) \cdot (x - x_0)$$



Propiedades básicas

- **Linealidad**: si a y b son constantes, entonces

$$\partial\Big[a\, f(x) + b\, g(x)\Big] = a\,\partial f(x) + b\,\partial g(x)$$

- **Regla del producto**:

$$\partial\Big[f(x)g(x)\Big] = \partial f(x)\cdot g(x) + f(x)\cdot \partial g(x)$$

- **Regla de la cadena**:

$$\partial\Big[f(g(x))\Big] = \partial f(g(x))\cdot \partial g(x)$$

Derivadas parciales y gradiente

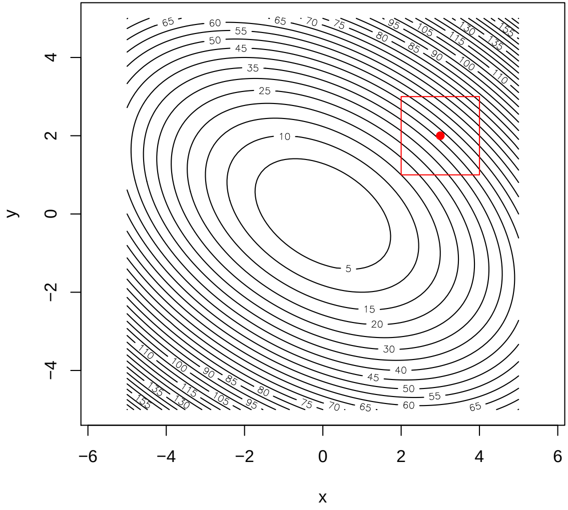
- Sea $y = f(\mathbf{x})$ una función con entrada $\mathbf{x} \sim (d)$ y salida escalar.
- Su derivada parcial respecto a x_i es

$$\partial_i f(\mathbf{x}) = \lim_{\Delta x_i \to 0} \frac{\Delta f}{\Delta x_i} \qquad \text{(dejando las } x_j,\ j \neq i \text{ constantes)}$$

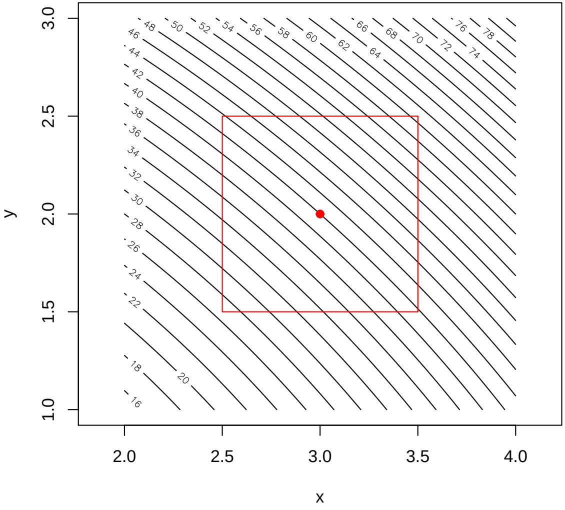
- El gradiente de $y = f(\mathbf{x})$ es

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \partial_1 f(\mathbf{x}) \\ \vdots \\ \partial_d f(\mathbf{x}) \end{bmatrix}$$

Curvas de nivel de f=2x^2+2xy+y^2

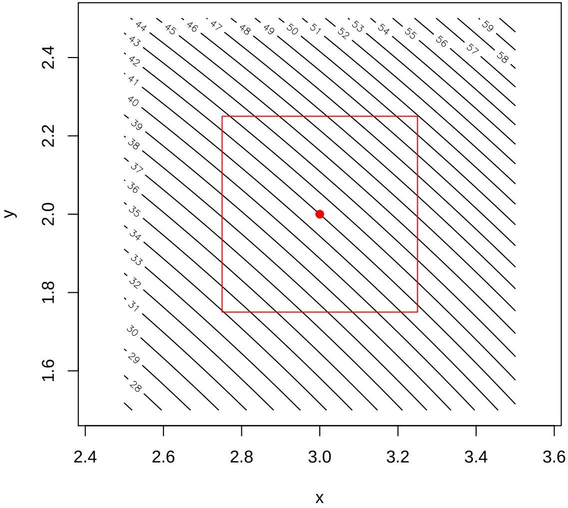


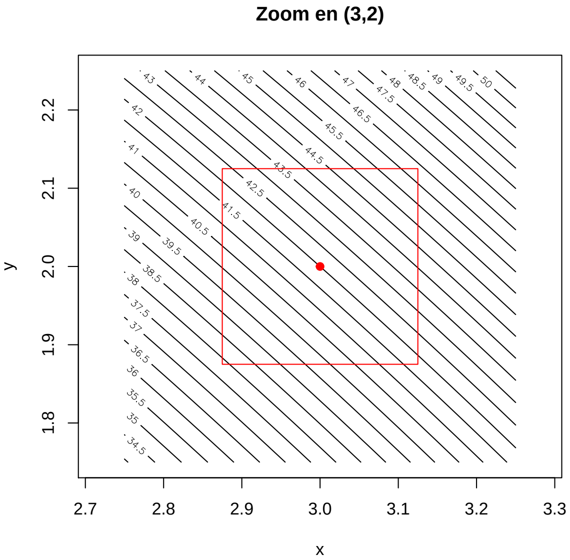
Zoom en (3,2)



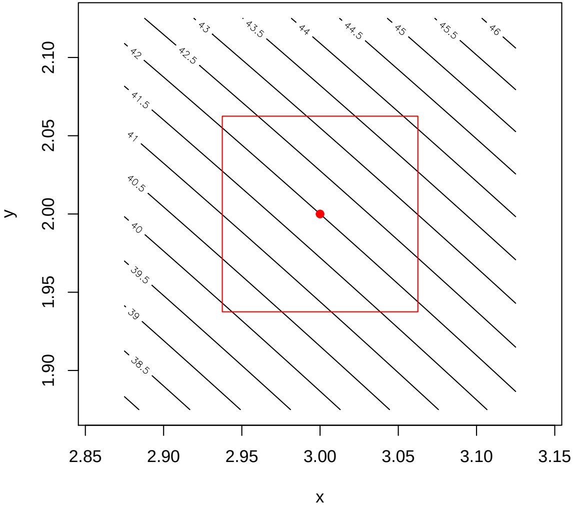
x
y
2.0
2.5
3.0
3.5
4.0
1.0
1.5
2.0
2.5
3.0
16
18
20
22
24
26
28
30
32
34
36
38
40
42
44
46
48
50
52
54
56
58
60
62
64
66
68
70
72
74
76
78

Zoom en (3,2)




Zoom en (3,2)
y
x
2.7
2.8
2.9
3.0
3.1
3.2
3.3
1.8
1.9
2.0
2.1
2.2

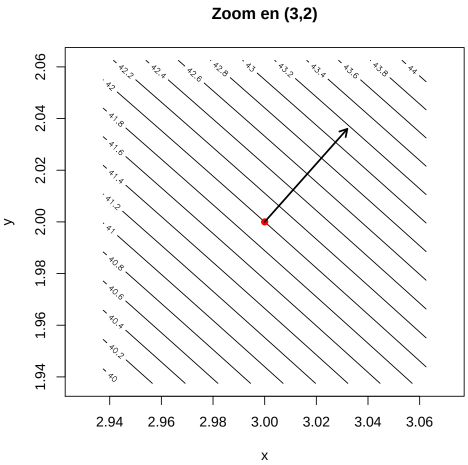
Zoom en (3,2)


38.5
39
39.5
40
40.5
41
41.5
42
42.5
43
43.5
44
44.5
45
45.5
46
2.85
2.90
2.95
3.00
3.05
3.10
3.15
x
1.90
1.95
2.00
2.05
2.10
y

Aproximación lineal

Para $\mathbf{x} \approx \mathbf{x}_0$:

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + \langle \nabla f(\mathbf{x}_0), \mathbf{x} - \mathbf{x}_0 \rangle$$



El Jacobiano

- Sea $\mathbf{y} = f(\mathbf{x})$ una función con entrada $\mathbf{x} \sim (d)$ y salida $\mathbf{y} \sim (k)$.
- Su Jacobiano se define como

$$\partial f(\mathbf{x}) = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_k}{\partial x_1} & \cdots & \frac{\partial y_k}{\partial x_d} \end{pmatrix} = \begin{pmatrix} \nabla f_1^\top \\ \vdots \\ \nabla f_k^\top \end{pmatrix} \sim (k, d)$$

  Siendo $f = (f_1, \cdots, f_k)$.

- **Regla de la cadena:**

$$\partial\left[f(g(\mathbf{x}))\right] = \partial f(g(\mathbf{x}))\,\partial g(\mathbf{x}) \qquad \text{(multiplicación de matrices)}$$

Ejemplo

- Consideremos la función lineal $\mathbf{y} = \mathbf{W}\mathbf{x}$
- Vista como función de $\mathbf{x}$, su Jacobiano es la matriz (k, d):

$$\partial_{\mathbf{x}} \left[\mathbf{W}\mathbf{x}\right] = \mathbf{W}$$

- Vista como función de $\mathbf{W}$, su Jacobiano es la matriz (k, kd):

$$y_i = \sum_j W_{ij} x_j \Rightarrow \frac{\partial y_i}{\partial W_{r\ell}} = \begin{cases} x_\ell & \text{si } r = i \\ 0 & \text{si } r \neq i \end{cases} \Rightarrow \partial_{\mathbf{W}} \mathbf{W}\mathbf{x} = \begin{pmatrix} \mathbf{x}^\top & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}^\top & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \vdots & \mathbf{0} & \mathbf{x}^\top \end{pmatrix}$$