Ejercicio 1: Backpropagation

Jordi Blasco Lozano

Planteamiento

Se considera una red neuronal feedforward con una neurona por capa. La arquitectura es:

$$a^{(0)} = x \;\longrightarrow\; (L-2) \;\longrightarrow\; (L-1) \;\longrightarrow\; (L) = \hat{y}.$$

En cada capa se calcula una preactivación $z^{(l)}$ y una activación $a^{(l)}$.

Datos iniciales

- Entrada: $x = 1$.
- Pesos iniciales: $W^{(L-2)} = 0{,}32$, $W^{(L-1)} = 0{,}18$, $W^{(L)} = 0{,}23$.
- Sesgos iniciales: $b^{(L-2)} = 0$, $b^{(L-1)} = 0$, $b^{(L)} = 0$.
- Tasa de aprendizaje: $\alpha = 0{,}1$.
- Etiqueta (target): $y = 1$.

Notación y fórmulas

Para cada capa l:

$$\begin{aligned} z^{(l)} &= W^{(l)}\, a^{(l-1)} + b^{(l)}, \\ a^{(l)} &= \sigma\big(z^{(l)}\big). \end{aligned}$$

Se usa la sigmoide:

$$\sigma(t) = \frac{1}{1+e^{-t}}, \qquad \sigma'(t) = \sigma(t)\,\big(1-\sigma(t)\big).$$

Función de pérdida:

$$L \;=\; (\hat{y} - y)^2.$$

Al trabajar con un solo ejemplo, “loss” y “cost” coinciden numéricamente.

1. Apartado 1: Forward pass (cálculo de z, a y $\hat{y}$)

En el forward pass se calculan las activaciones desde la entrada hasta la salida.

Capa $L-2$

$$\begin{aligned} z^{(L-2)} &= W^{(L-2)}\, x + b^{(L-2)} = 0{,}32 \cdot 1 + 0 = 0{,}32, \\ a^{(L-2)} &= \sigma\big(z^{(L-2)}\big) = \frac{1}{1+e^{-0{,}32}} \approx \frac{1}{1+0{,}7261} \approx 0{,}5793. \end{aligned}$$

Capa $L-1$

$$z^{(L-1)} = W^{(L-1)}\,a^{(L-2)} + b^{(L-1)} = 0{,}18 \cdot 0{,}5793 + 0 \approx 0{,}1043,$$
$$a^{(L-1)} = \sigma\big(z^{(L-1)}\big) = \frac{1}{1+e^{-0{,}1043}} \approx \frac{1}{1+0{,}9009} \approx 0{,}5260.$$

Capa de salida L

$$z^{(L)} = W^{(L)}\,a^{(L-1)} + b^{(L)} = 0{,}23 \cdot 0{,}5260 + 0 \approx 0{,}1210,$$
$$\hat{y} = a^{(L)} = \sigma\big(z^{(L)}\big) = \frac{1}{1+e^{-0{,}1210}} \approx \frac{1}{1+0{,}8860} \approx 0{,}5302.$$

Pérdida (loss)

$$L = (\hat{y}-y)^2 = (0{,}5302-1)^2 = (-0{,}4698)^2 \approx 0{,}2207.$$

Resumen del forward pass

Capa	$z^{(l)}$	$a^{(l)}$	Comentario
$L-2$	0,32	0,5793	$a^{(0)} = x = 1$
$L-1$	0,1043	0,5260	$\sigma(z)$
L	0,1210	$\hat{y} = 0{,}5302$	salida

2. Apartado 2: Backward pass (gradientes y actualización)

El objetivo es actualizar $W^{(L)}$ y $b^{(L-1)}$ mediante descenso por gradiente.

Mapa de dependencias (para justificar la regla de la cadena)

$$W^{(L)} \to z^{(L)} \to \hat{y} \to L, \qquad b^{(L-1)} \to z^{(L-1)} \to a^{(L-1)} \to z^{(L)} \to \hat{y} \to L.$$

Para simplificar, definimos el "error" (delta) por capa:

$$\delta^{(l)} := \frac{\partial L}{\partial z^{(l)}}.$$

Los gradientes se calculan con los parámetros del forward actual (antes de aplicar la actualización).

Capa de salida L

1) Derivada de la pérdida respecto a la salida

$$\frac{\partial L}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}}(\hat{y}-y)^2 = 2(\hat{y}-y) = 2(0{,}5302-1) \approx -0{,}9396.$$

2) Delta de la capa de salida

$$\delta^{(L)} = \frac{\partial L}{\partial z^{(L)}} = \frac{\partial L}{\partial \hat{y}}\,\frac{\partial \hat{y}}{\partial z^{(L)}} = \frac{\partial L}{\partial \hat{y}}\,\sigma'\big(z^{(L)}\big)$$
$$= (-0{,}9396)\,\big(\hat{y}(1-\hat{y})\big) = (-0{,}9396)\,(0{,}5302 \cdot 0{,}4698) \approx (-0{,}9396) \cdot 0{,}2491 \approx -0{,}2340.$$

3) Gradiente respecto a $W^{(L)}$

Como $z^{(L)} = W^{(L)} a^{(L-1)} + b^{(L)}$, se tiene $\frac{\partial z^{(L)}}{\partial W^{(L)}} = a^{(L-1)}$. Por tanto:

$$\frac{\partial L}{\partial W^{(L)}} = \frac{\partial L}{\partial z^{(L)}} \frac{\partial z^{(L)}}{\partial W^{(L)}} = \delta^{(L)}\, a^{(L-1)} \approx (-0{,}2340) \cdot 0{,}5260 \approx -0{,}1231.$$

4) Actualización de $W^{(L)}$

$$W^{(L)}_{\text{nuevo}} = W^{(L)} - \alpha \frac{\partial L}{\partial W^{(L)}} = 0{,}23 - 0{,}1 \cdot (-0{,}1231) \approx 0{,}2423.$$

Capa $L-1$ (para actualizar $b^{(L-1)}$)

5) Propagación del delta a la capa anterior

$$\delta^{(L-1)} = \frac{\partial L}{\partial z^{(L-1)}} = \underbrace{\frac{\partial L}{\partial z^{(L)}}}_{\delta^{(L)}} \underbrace{\frac{\partial z^{(L)}}{\partial a^{(L-1)}}}_{W^{(L)}} \underbrace{\frac{\partial a^{(L-1)}}{\partial z^{(L-1)}}}_{\sigma'(z^{(L-1)})}.$$

Sustituyendo valores del forward (con $W^{(L)} = 0{,}23$ antes de actualizar):

$$\begin{aligned}
\delta^{(L-1)} &= \delta^{(L)}\, W^{(L)}\, \sigma'(z^{(L-1)}) \\
&\approx (-0{,}2340) \cdot 0{,}23 \cdot \left(a^{(L-1)}(1 - a^{(L-1)})\right) \\
&\approx (-0{,}2340) \cdot 0{,}23 \cdot (0{,}5260 \cdot 0{,}4740) \\
&\approx (-0{,}2340) \cdot 0{,}23 \cdot 0{,}2493 \approx -0{,}0134.
\end{aligned}$$

6) Gradiente y actualización de $b^{(L-1)}$

Como $z^{(L-1)} = W^{(L-1)} a^{(L-2)} + b^{(L-1)}$, se cumple $\frac{\partial z^{(L-1)}}{\partial b^{(L-1)}} = 1$, y por tanto:

$$\frac{\partial L}{\partial b^{(L-1)}} = \delta^{(L-1)} \approx -0{,}0134.$$

La actualización queda:

$$b^{(L-1)}_{\text{nuevo}} = b^{(L-1)} - \alpha \frac{\partial L}{\partial b^{(L-1)}} = 0 - 0{,}1 \cdot (-0{,}0134) \approx 0{,}00134.$$

Resumen final

Magnitud	Valor
$\hat{y}$	0,5302
$L = (\hat{y} - y)^2$	0,2207
$\frac{\partial L}{\partial W^{(L)}}$	$-0{,}1231$
$W^{(L)}_{\text{nuevo}}$	0,2423
$\frac{\partial L}{\partial b^{(L-1)}}$	$-0{,}0134$
$b^{(L-1)}_{\text{nuevo}}$	0,00134