Día 02 · Tema 1 — Ejercicios: neurona, MLP y backpropagation

Cómo se usa esta hoja. Hay dos tipos de ejercicio, y sólo dos.

El ejercicio del día. Uno por hoja. Es el obligatorio y es el que entra en el examen, con el cálculo incluido. Basta con hacer ese para seguir el curso con solvencia.

Los ejercicios de ampliación. El resto. Están para enseñar de qué es capaz el método, no para exigir: **no se piden calculados en el examen**. De ellos sólo puede preguntarse la idea, a nivel conceptual.

EL EJERCICIO DEL DÍA *obligatorio · entra en el examen*

1. Pase forward completo por una red $2 \to 3 \to 1$ y error final

Un operador de telecomunicaciones estima la probabilidad de fuga (*churn*) de cada cliente con una red de dos entradas, una capa oculta de tres neuronas con activación ReLU y una neurona de salida con activación sigmoide. Las dos entradas están estandarizadas sobre la cartera de clientes (media 0, desviación típica 1): x_1 es el gasto mensual y x_2 el número de llamadas al servicio de soporte en el último trimestre.

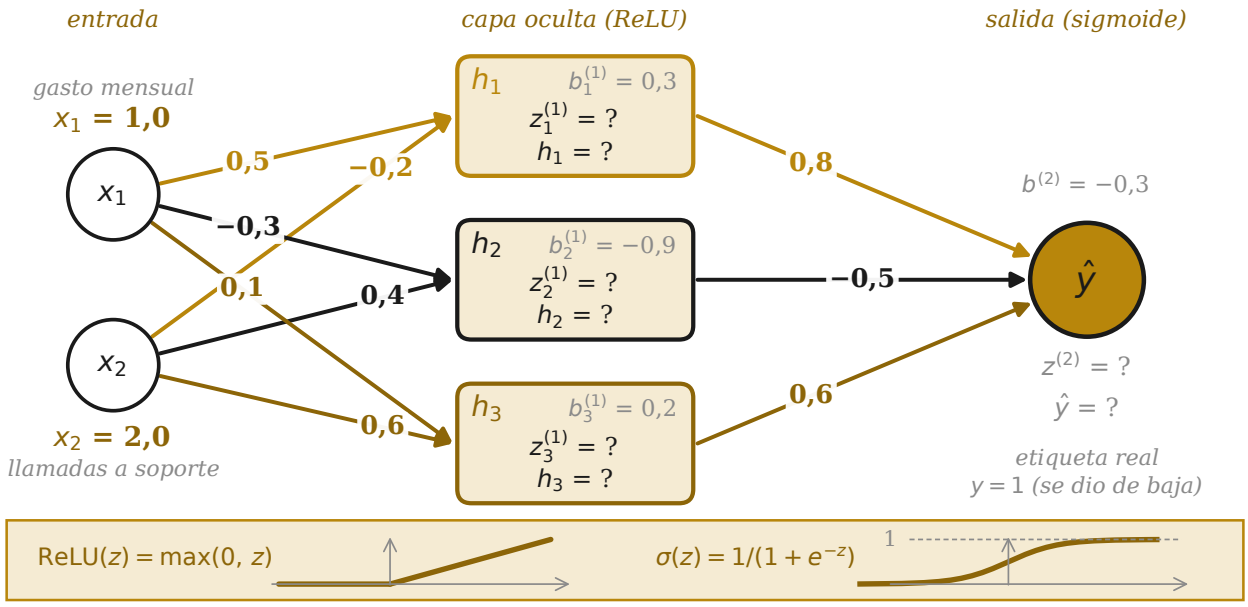
Del cliente 4713 se observa

$$\mathbf{x} = (1{,}0,\ 2{,}0)^\top,$$

es decir, gasta una desviación típica por encima de la media y llama al soporte dos desviaciones típicas por encima de la media. El cliente pertenece al histórico de entrenamiento y efectivamente se dio de baja el trimestre siguiente, de modo que la etiqueta real es $y = 1$. Los parámetros de la red son

$$W^{(1)} = \begin{pmatrix} 0{,}5 & -0{,}2 \\ -0{,}3 & 0{,}4 \\ 0{,}1 & 0{,}6 \end{pmatrix}, \quad \mathbf{b}^{(1)} = \begin{pmatrix} 0{,}3 \\ -0{,}9 \\ 0{,}2 \end{pmatrix}, \quad \mathbf{w}^{(2)} = \begin{pmatrix} 0{,}8 \\ -0{,}5 \\ 0{,}6 \end{pmatrix}, \quad b^{(2)} = -0{,}3,$$

donde la fila j de $W^{(1)}$ contiene los pesos de la neurona oculta j, de manera que $W^{(1)}_{jk}$ es el peso con el que la entrada x_k entra en la neurona oculta j.



Las mismas matrices, dibujadas. Cada arista lleva su peso y va del color de la neurona a la que llega (igual que las filas de $W^{(1)}$ arriba): la fila j son las dos aristas que entran en h_j, y la columna k, las tres que salen de x_k. Los sesgos entran en la suma sin multiplicar por nada. Rellena lo que aparece con "?".

a) **Paso 1.** Calcula las tres pre-activaciones de la capa oculta, $\mathbf{z}^{(1)} = W^{(1)}\mathbf{x} + \mathbf{b}^{(1)}$.
b) **Paso 2.** Aplica la activación: $\mathbf{h} = \mathrm{ReLU}(\mathbf{z}^{(1)})$, con $\mathrm{ReLU}(z) = \max(0, z)$ componente a componente.
c) **Paso 3.** Calcula la pre-activación de la salida, $z^{(2)} = \mathbf{w}^{(2)\top}\mathbf{h} + b^{(2)}$.
d) **Paso 4.** Aplica la sigmoide $\sigma(z) = 1/(1 + e^{-z})$ para obtener la predicción $\hat{y}$.
e) **Paso 5.** Calcula el error final con la entropía cruzada binaria, $\mathcal{L} = -\big[y \ln \hat{y} + (1 - y)\ln(1 - \hat{y})\big]$.

© 2026 Eduardo C. Garrido-Merchán · CC BY-NC-SA 4.0

f) **Lectura de negocio.** ¿Se marca al cliente como *churn* con umbral 0,5? ¿Acertó la red? ¿Cuánto habría valido el error si el cliente *no* se hubiera dado de baja ($y = 0$)? Identifica qué neurona oculta no aporta nada al veredicto de este cliente y qué variable domina la decisión.

AMPLIACIÓN *opcional · no se pide calculado en el examen*

2. Pase forward por una neurona

Un operador de telecomunicaciones quiere predecir la fuga de clientes. Se dispone de una única neurona con tres entradas: $x_1 = 24$ (meses de antigüedad), $x_2 = 80$ (gasto mensual en euros) y $x_3 = 5$ (llamadas a soporte). Los pesos y el sesgo son

$$\mathbf{w} = (-0{,}04,\ 0{,}02,\ 0{,}55), \qquad b = -0{,}50.$$

a) Calcula la pre-activación $z = \mathbf{w}^\top \mathbf{x} + b$.
b) Calcula la salida si la activación es ReLU: $\mathrm{ReLU}(z) = \max(0, z)$.
c) Calcula la salida si la activación es sigmoide: $\sigma(z) = \dfrac{1}{1 + e^{-z}}$.
d) Interpreta el resultado de la sigmoide como probabilidad de fuga. ¿Clasificarías a este cliente como *churn* con un umbral de 0,5?

AMPLIACIÓN *opcional · no se pide calculado en el examen*

3. Pase forward por un MLP de dos capas

Considera un MLP con 2 entradas, 2 neuronas ocultas (activación ReLU) y 1 neurona de salida (activación sigmoide). La entrada es $\mathbf{x} = (1{,}0,\ 2{,}0)^\top$. Los parámetros son:

$$W^{(1)} = \begin{pmatrix} 0{,}3 & -0{,}5 \\ 0{,}8 & 0{,}1 \end{pmatrix}, \quad \mathbf{b}^{(1)} = \begin{pmatrix} -0{,}1 \\ 0{,}2 \end{pmatrix}, \qquad \mathbf{w}^{(2)} = \begin{pmatrix} 0{,}6 \\ -0{,}4 \end{pmatrix}, \quad b^{(2)} = 0{,}1.$$

a) Calcula las pre-activaciones de la capa oculta $\mathbf{z}^{(1)}$.
b) Aplica ReLU para obtener $\mathbf{h} = \mathrm{ReLU}(\mathbf{z}^{(1)})$.
c) Calcula la pre-activación de la capa de salida $z^{(2)}$.
d) Aplica sigmoide para obtener $\hat{y}$.
e) Si la etiqueta real es $y = 0$, ¿es esta una buena predicción?

AMPLIACIÓN *opcional · no se pide calculado en el examen*

4. Entropía cruzada binaria

La entropía cruzada binaria para una observación es

$$\mathcal{L}(\hat{y}, y) = -\big[y \ln \hat{y} + (1 - y) \ln(1 - \hat{y})\big].$$

a) Calcula $\mathcal{L}$ para una observación con $y = 1$, $\hat{y} = 0{,}82$.
b) Calcula $\mathcal{L}$ para una observación con $y = 0$, $\hat{y} = 0{,}35$.
c) Calcula la pérdida media del mini-batch formado por estas dos observaciones.
d) Si en vez de $\hat{y} = 0{,}82$ el modelo hubiera predicho $\hat{y} = 0{,}99$ para la primera observación, ¿cómo cambia la pérdida? Comenta.

AMPLIACIÓN *opcional · no se pide calculado en el examen*

5. Un paso de SGD

En un instante del entrenamiento, los pesos de una neurona son

$$\mathbf{w} = (1{,}50,\ -0{,}80,\ 0{,}30), \qquad b = 0{,}20.$$

Los gradientes calculados para el mini-batch actual son

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = (0{,}12,\ -0{,}35,\ 0{,}08), \qquad \frac{\partial \mathcal{L}}{\partial b} = 0{,}15.$$

La tasa de aprendizaje es $\eta = 0{,}01$.

a) Escribe la regla de actualización de SGD.
b) Calcula los nuevos valores de $\mathbf{w}$ y b.
c) ¿En qué dirección se mueve cada peso? ¿Tiene sentido?

AMPLIACIÓN *opcional · no se pide calculado en el examen*

© 2026 Eduardo C. Garrido-Merchán · CC BY-NC-SA 4.0

6. Backpropagation completo en una red 2-2-1

Considera la red del siguiente diagrama: 2 entradas, 2 neuronas ocultas (ReLU) y 1 neurona de salida (sigmoide) con pérdida de entropía cruzada binaria. La entrada es $\mathbf{x} = (1,\ -1)^\top$ y la etiqueta $y = 1$. Los parámetros son:

$$W^{(1)} = \begin{pmatrix} 0{,}5 & -0{,}3 \\ 0{,}2 & 0{,}4 \end{pmatrix}, \quad \mathbf{b}^{(1)} = \begin{pmatrix} 0{,}1 \\ -0{,}1 \end{pmatrix}, \qquad \mathbf{w}^{(2)} = \begin{pmatrix} 0{,}7 \\ -0{,}5 \end{pmatrix}, \quad b^{(2)} = 0{,}2.$$

a) Realiza el pase forward completo: calcula $\mathbf{z}^{(1)}$, $\mathbf{h}$, $z^{(2)}$ y $\hat{y}$.
b) Calcula la pérdida $\mathcal{L}$.
c) Calcula la señal de error de salida $\delta^{(2)} = \hat{y} - y$.
d) Calcula los gradientes de la capa de salida: $\partial\mathcal{L}/\partial w_j^{(2)}$ y $\partial\mathcal{L}/\partial b^{(2)}$.
e) Retropropaga el error a la capa oculta: calcula $\delta_j^{(1)} = \delta^{(2)}\, w_j^{(2)}\ \mathrm{ReLU}'(z_j^{(1)})$ para $j = 1, 2$.
f) Calcula todos los gradientes de la capa oculta: $\partial\mathcal{L}/\partial W_{jk}^{(1)}$ y $\partial\mathcal{L}/\partial b_j^{(1)}$.
g) Comenta qué ocurre con los gradientes de la segunda neurona oculta y por qué.

AMPLIACIÓN *opcional · no se pide calculado en el examen*

7. Un paso de Adam

Se optimiza un parámetro escalar θ con Adam. Los hiperparámetros son $\alpha = 0{,}01$, $\beta_1 = 0{,}9$, $\beta_2 = 0{,}999$, $\varepsilon = 10^{-8}$. El estado inicial es $\theta_0 = 2{,}0$, $m_0 = 0$, $v_0 = 0$.

a) En $t = 1$ el gradiente es $g_1 = 0{,}5$. Calcula m_1, v_1, sus versiones corregidas $\hat{m}_1$ y $\hat{v}_1$, y el nuevo parámetro θ_1.
b) En $t = 2$ el gradiente es $g_2 = -0{,}3$. Repite el cálculo para obtener θ_2.
c) Observa que en $t = 1$, partiendo de $m_0 = v_0 = 0$, el tamaño efectivo de paso es α independientemente de la magnitud del gradiente. Demuéstralo.
d) Compara θ_1 y θ_2 con los que obtendría SGD simple con la misma tasa α. ¿Qué efecto tiene el cambio de signo del gradiente en Adam frente a SGD?

AMPLIACIÓN *opcional · no se pide calculado en el examen*

8. Precision, recall y F1 a partir de una matriz de confusión

Un modelo de detección de fraude bancario se evalúa sobre 100 transacciones y produce la siguiente matriz de confusión:

	Predicho **fraude**	Predicho **legítimo**
Real **fraude**	42	8
Real **legítimo**	12	38

a) Identifica TP, FP, FN y TN.
b) Calcula precision, recall y F1.
c) Calcula la accuracy. ¿Es un buen resumen del rendimiento en este problema? ¿Por qué?
d) Si el coste de un falso negativo (fraude no detectado) fuera 10 veces mayor que el de un falso positivo (transacción legítima bloqueada), ¿qué métrica priorizarías?

AMPLIACIÓN *opcional · no se pide calculado en el examen*

9. Selección del umbral que maximiza F1

Un modelo de riesgo crediticio produce las siguientes puntuaciones para 8 clientes, ordenados de mayor a menor probabilidad de impago:

Cliente	A	B	C	D	E	F	G	H
$\hat{y}$	0,91	0,73	0,62	0,55	0,44	0,38	0,31	0,15
y	1	1	0	1	0	0	1	0

a) Para cada umbral $\tau \in \{0{,}3,\ 0{,}4,\ 0{,}5,\ 0{,}6,\ 0{,}7\}$, clasifica como positivo si $\hat{y} \geq \tau$ y calcula precision, recall y F1.
b) ¿Qué umbral maximiza F1?
c) ¿Qué umbral maximiza precision? ¿Y recall?
d) Describe cualitativamente cómo cambiaría la curva ROC al mover el umbral de 0,3 a 0,7.

© 2026 Eduardo C. Garrido-Merchán · CC BY-NC-SA 4.0