

Día 16 · Tema 6 — Ejercicios: LoRA, RLHF y DPO

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.

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

1. Recuento de parámetros LoRA en un Transformer

Un Transformer tiene dimensión oculta $d = 4096$. En cada capa de atención hay cuatro matrices de proyección W^Q, W^K, W^V, W^O , cada una de tamaño 4096×4096 . Se aplica LoRA con rango $r = 8$ a las cuatro matrices.

Recuerda que LoRA reemplaza el ajuste completo de $W \in \mathbb{R}^{m \times n}$ por

$$W' = W + \frac{\alpha}{r} B A, \quad A \in \mathbb{R}^{r \times n}, \quad B \in \mathbb{R}^{m \times r}.$$

Solo se entrenan A y B ; la matriz original W permanece congelada.

- ¿Cuántos parámetros tiene la capa de atención original (las cuatro matrices, sin contar sesgos)?
- ¿Cuántos parámetros entrenables introduce LoRA en esa misma capa?
- Calcula la razón de compresión (parámetros originales / parámetros LoRA).
- Si el modelo tiene 32 capas y se aplica LoRA a todas, ¿cuántos parámetros LoRA hay en total? ¿Qué porcentaje representan del total original?

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

2. Forward pass de LoRA (ejemplo numérico)

Sean la matriz original $W \in \mathbb{R}^{2 \times 2}$, los adaptadores LoRA $A \in \mathbb{R}^{1 \times 2}$ y $B \in \mathbb{R}^{2 \times 1}$ (rango $r = 1$), el factor de escala $\alpha = 1$, y el vector de entrada $\mathbf{x}$:

$$W = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad A = (0,1 \quad 0,2), \quad B = \begin{pmatrix} 0,5 \\ 0,3 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- Calcula $W\mathbf{x}$.
- Calcula el producto $B A \mathbf{x}$ paso a paso.
- Calcula la salida completa $W'\mathbf{x} = W\mathbf{x} + \frac{\alpha}{r} B A \mathbf{x}$.
- Comenta: ¿cuántos parámetros entrenables tiene este adaptador frente a la matriz W completa? ¿Cuál es la naturaleza (rango) de la corrección $\Delta W = B A$?

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

3. Ahorro de memoria con LoRA

Un LLM de 7 B (7 mil millones) de parámetros en FP16 ocupa 14 GB en memoria.

Supón que las matrices de atención representan aproximadamente $\frac{1}{3}$ de los parámetros totales. Se aplica LoRA con rango $r = 16$ únicamente a esas matrices. Cada matriz original es cuadrada de dimensión $d = 4096$, y LoRA añade $r(d + d) = 2rd$ parámetros por cada una, con una razón de compresión $\frac{d^2}{2rd} = \frac{d}{2r}$.

- ¿Cuántos parámetros de atención tiene el modelo original?
- ¿Cuántos parámetros LoRA se introducen? (Divide los parámetros de atención por la razón de compresión $\frac{d}{2r}$.)
- ¿Cuánta memoria ocupan los adaptadores LoRA en FP16 (2 bytes por parámetro)?
- ¿Qué porcentaje de la memoria del modelo base representan?

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

4. Memoria con QLoRA

Partiendo del mismo LLM de 7 B:

- Si se cuantiza el modelo base a 4 bits (0,5 bytes por parámetro), ¿cuántos GB ocupa?
- Añade los adaptadores LoRA del ejercicio anterior (en FP16). ¿Cuál es la memoria total de QLoRA?

- (c) El fine-tuning completo en FP16 necesita almacenar parámetros (14 GB) más estados del optimizador (momentos de Adam $\approx 2 \times$ los parámetros) ≈ 42 GB. ¿Cuántas veces más memoria necesita el fine-tuning completo respecto a QLoRA?

AMPLIACIÓN opcional · no se pide calculado en el examen

5. Pérdida del modelo de recompensa (Bradley-Terry)

En RLHF, el modelo de recompensa se entrena con la pérdida de Bradley-Terry:

$$\mathcal{L}_{\text{RM}} = -\mathbb{E} \left[\log \sigma(r(x, y_w) - r(x, y_l)) \right],$$

donde y_w es la respuesta preferida, y_l la rechazada, y $\sigma(z) = \frac{1}{1+e^{-z}}$.

- (a) **Caso 1 (preferencia clara):** $r(x, y_w) = 2,5$ y $r(x, y_l) = 1,0$. Calcula la diferencia, aplica σ (usa $\sigma(1,5) \approx 0,818$) y obtén la pérdida.
- (b) **Caso 2 (preferencia ambigua):** $r(x, y_w) = 1,2$ y $r(x, y_l) = 1,0$. Calcula la diferencia, aplica σ (usa $\sigma(0,2) \approx 0,550$) y obtén la pérdida.
- (c) Compara ambas pérdidas. ¿Cuándo le resulta más difícil al modelo aprender las preferencias? ¿Qué implicación tiene para la calidad de las anotaciones humanas?

AMPLIACIÓN opcional · no se pide calculado en el examen

6. Pérdida DPO

DPO evita entrenar un modelo de recompensa explícito. Su pérdida es:

$$\mathcal{L}_{\text{DPO}} = -\mathbb{E} \left[\log \sigma \left(\beta \left[\log \frac{\pi_{\theta}(y_w | x)}{\pi_{\text{SFT}}(y_w | x)} - \log \frac{\pi_{\theta}(y_l | x)}{\pi_{\text{SFT}}(y_l | x)} \right] \right) \right].$$

Datos: $\log \frac{\pi_{\theta}(y_w | x)}{\pi_{\text{SFT}}(y_w | x)} = 0,5$ y $\log \frac{\pi_{\theta}(y_l | x)}{\pi_{\text{SFT}}(y_l | x)} = 0,3$.

- (a) Con $\beta = 0,1$: calcula el argumento de σ , aplica σ (usa $\sigma(0,02) \approx 0,505$) y obtén la pérdida.
- (b) Con $\beta = 1,0$: repite el cálculo (usa $\sigma(0,2) \approx 0,550$).
- (c) Compara las dos pérdidas. ¿Qué controla β ? ¿Qué ocurre si se elige un β demasiado pequeño o demasiado grande?

AMPLIACIÓN opcional · no se pide calculado en el examen

7. Árbol de decisión cuantitativo: adaptación de un LLM

Una empresa procesa 1 000 facturas al mes mediante un LLM que extrae campos clave (importe, fecha, proveedor). Cada extracción errónea requiere corrección manual con un coste de 20 €. Dispone de 200 facturas etiquetadas.

Se consideran cuatro estrategias de adaptación con las siguientes estimaciones:

Estrategia	Coste inicial	Precisión estimada
(a) Prompt engineering	0 €	70 %
(b) Few-shot (5 ejemplos)	0 €	80 %
(c) LoRA (200 ejemplos)	50 €	92 %
(d) RLHF completo	15 000 €	95 %

- (a) Para cada estrategia, calcula el número de errores mensuales y el coste mensual de corrección.
- (b) Calcula el ahorro mensual de LoRA frente a prompt engineering.
- (c) ¿En cuánto tiempo se amortiza el coste inicial de LoRA?
- (d) ¿En cuánto tiempo se amortiza RLHF frente a prompt engineering? ¿Y frente a LoRA?
- (e) ¿Qué estrategia recomendarías a esta empresa y por qué?