LLM Refinement

Natural Language Processing

University of Maryland

LoRA

Starting Point

We begin with a matrix W:

$$W = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Starting Point

We begin with a matrix W:

$$W = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Goal: adapt W (without retraining it entirely) using a **low-rank update**.

Target Matrix

We want to reach:

$$W_{\text{target}} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 5 \\ 3 & 5 & 7 & 9 \end{bmatrix}$$

Target Matrix

We want to reach:

$$W_{\text{target}} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 5 \\ 3 & 5 & 7 & 9 \end{bmatrix}$$

What change do we need to apply?

Compute the Difference

$$\Delta W = W_{\text{target}} - W =$$

Compute the Difference

$$\Delta W = W_{\text{target}} - W = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \end{bmatrix}$$

Compute the Difference

$$\Delta W = W_{\text{target}} - W = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \end{bmatrix}$$

Goal: approximate ΔW using a **low-rank factorization**.

What Rank is Needed? (Think Like LoRA)

Question: What is the **minimum rank** needed to represent ΔW?

What Rank is Needed? (Think Like LoRA)

Question: What is the **minimum rank** needed to represent ΔW?
Observation: Row 4 is just 2$\times$ row 3.

What Rank is Needed? (Think Like LoRA)

Question: What is the **minimum rank** needed to represent ΔW?
Observation: Row 4 is just 2$\times$ row 3.

$$\Rightarrow \Delta W \text{ has rank } 1$$

Rank-1 Factorization

We write:

$$\Delta W = a \cdot b$$

Rank-1 Factorization

We write:

$$\Delta W = a \cdot b$$

$$a = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 2 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}$$

Rank-1 Factorization

We write:

$$\Delta W = a \cdot b$$

$$a = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 2 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}$$

Check:

$$a \cdot b = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \end{bmatrix}$$

LoRA Update View

In LoRA we adapt W like:

$$W' = W + AB$$

LoRA Update View

In LoRA we adapt W like:

$$W' = W + AB$$

Here:

$$A = a, \quad B = b$$

LoRA Update View

In LoRA we adapt W like:

$$W' = W + AB$$

Here:

$$A = a, \quad B = b$$

$$W' = W + \Delta W = W_{\text{target}}$$