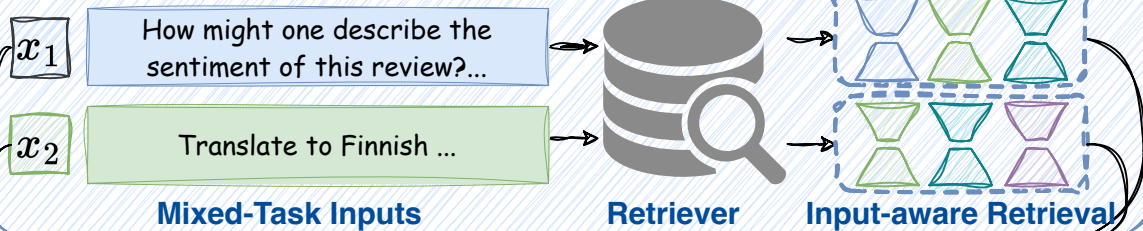
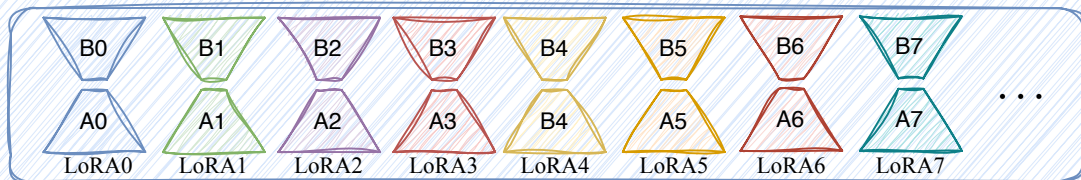
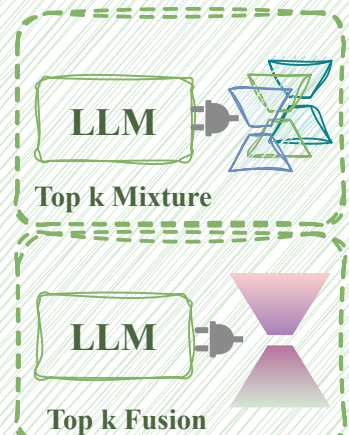


a) Input-Aware LoRA Retrieval

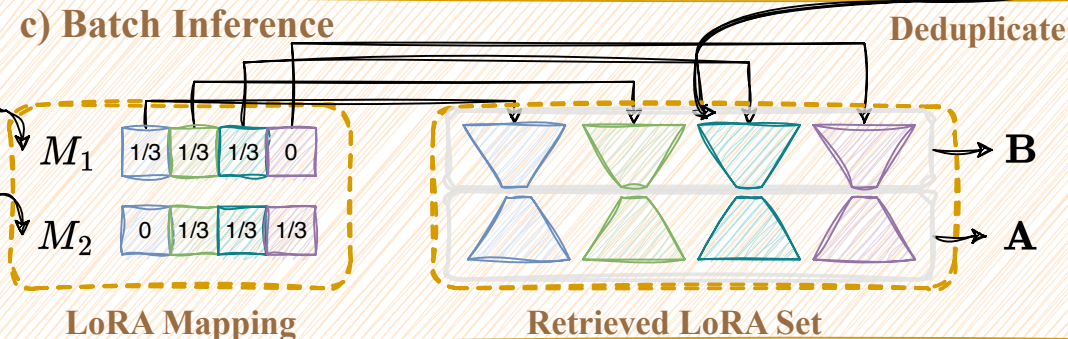
Candidate LoRAs



b) LoRA Composition



c) Batch Inference



Top k Mixture

$$M \circ (B \circ A \circ X)$$

Top k Fusion

$$(M \circ B)(M \circ A) \circ X$$

Batch Calculation