

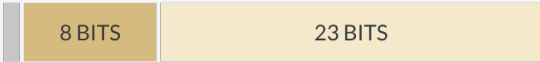







Foundations of LLM Mastery: Fine-tuning with multi GPUs

25 February 2025
ONLINE



Limited GPU memory

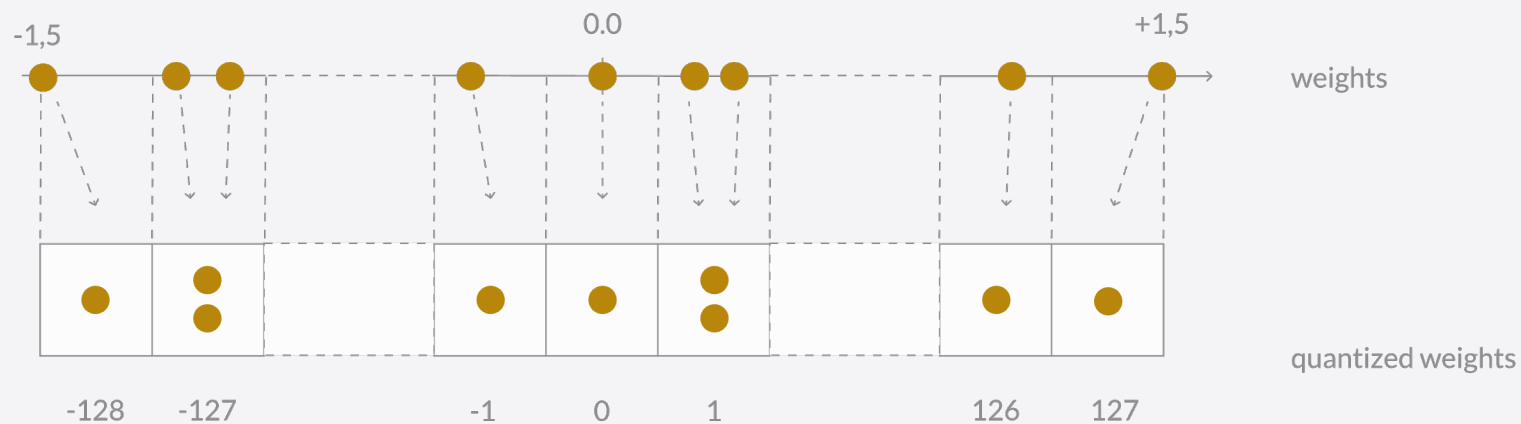
Bits per parameter		Data type		Largest possible number
32 bits	FP32			3.389×10^{38}
16 bits	BFLOAT16			3.389×10^{38}
16 bits	FP16			65504

 sign
 exponent
 mantissa

Fewer bits → Quantization

$$Value = (-1)^S * 2^{(E-15)} * (1 - M)$$

Quantization



Low rank adapters (LoRA)

