

What is PEFT?

Instructor

Sourab Mangrulkar

Machine Learning Engineer at , Creator of PEFT



What is PEFT?

- Methods that allow us to fine-tune a large pretrained model on a specific downstream task while requiring significantly fewer parameters than full fine-tuning.

What is PEFT?

- Methods that allow us to fine-tune a large pretrained model on a specific downstream task while requiring significantly fewer parameters than full fine-tuning.
- The goal is to achieve comparable or even better performance than full fine-tuning, while requiring less computation and memory resources.

How does PEFT
works?

How does PEFT work?

- PEFT approaches only fine-tune a small number of (extra) model parameters while freezing most parameters of the pretrained LLMs, thereby greatly decreasing the computational and storage costs.

How does PEFT work?

- PEFT approaches only fine-tune a small number of (extra) model parameters while freezing most parameters of the pretrained LLMs, thereby greatly decreasing the computational and storage costs.
- Adapts to the downstream task while reducing the risk of catastrophic forgetting.

How does PEFT work?

- PEFT approaches only fine-tune a small number of (extra) model parameters while freezing most parameters of the pretrained LLMs, thereby greatly decreasing the computational and storage costs.
- Adapts to the downstream task while reducing the risk of catastrophic forgetting.
- Prevents overfitting