Analytics Vidhya

Limitations of Full Finetuning

Instructor

Sourab Mangrulkar

Machine Learning Engineer at

Creator of PEFT

The Secret Sauce behind LLMs & Limitations

Recipe to get state-of-the-art results in NLP

- Pretraining on web-scale data
- Finetuning on downstream task

Scaling Laws

The larger the Deep Learning/AI model, the better the performance.

Limitations of full fine-tuning approaches

- Compute
- Storage

How to Finetune very large models on consumer hardware?

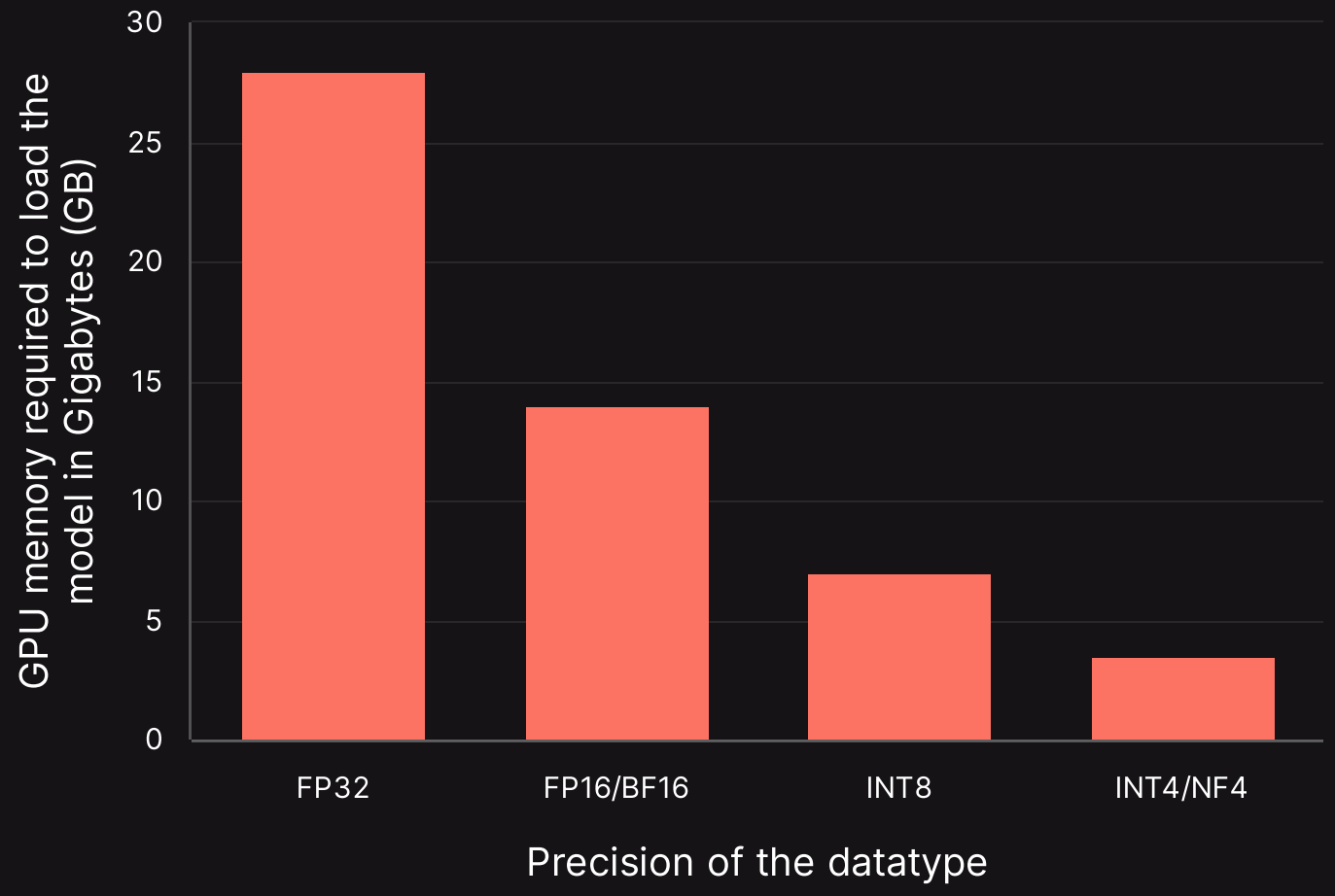
Analytics Vidhya

What makes a model large?

Two Factors

1. Number of parameters
2. Precision of the data type

Let’s put some numbers wrt Mistral-7B model with 7 Billion parameters



Analytics Vidhya

Why full finetuning is expensive?

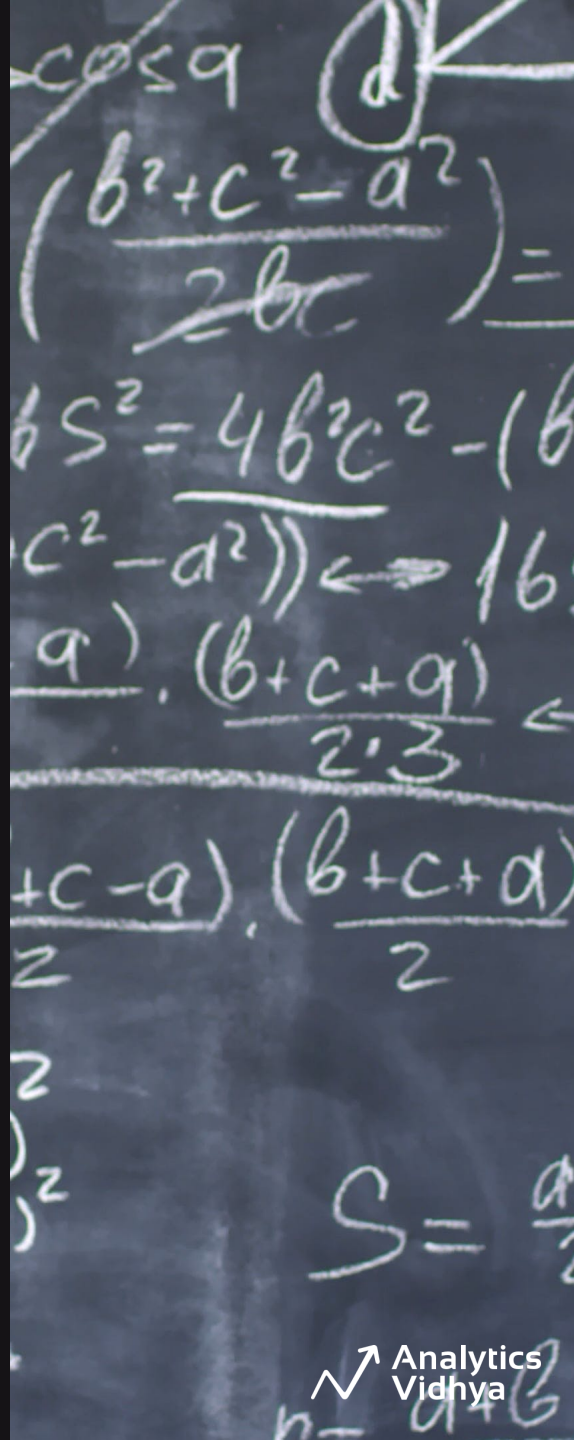
Finetuning Mistral-7B in mixed-precision using Adam Optimizer

Weights - 2 bytes / parameter

Gradients - 2 bytes / parameter

Optimizer state - 4 bytes / parameter (FP32 copy) +
8 bytes / parameter (momentum & variance estimates)

Total training cost: 16 bytes/parameter * 7 billion parameters = 112 GB



Analytics Vidhya

Analytics Vidhya

Hands On Project

Instructor

Sourab Mangrulkar

Machine Learning Engineer at

Creator of PEFT