DoRA: Weight Decomposed Low-rank Adaptation

Introduction

- Various parameter efficient fine-tuning (PEFT) methods have been proposed to by-pass full parameter fine-tuning. Example: LoRA
- Significant learning capacity b/w LoRA and FT (full tuning)
  - Limited number of trainable parameters
- DoRA: decomposes pre-trained weight matrix W into MxD (magnitude and direction of updates)

LoRA : Low Rank Adaptation

- Updates made during fine-tuning have an intrinsic low-rank
  - $W' = W + B.A$. where $A \in R$ (rxk) and $B \in R$(dxr) and $r << \min(k,d)$
- W is freezed while B and A having much lesser parameters are updated
- Tends to either increase or decrease the magnitude and direction updates proportionally and lacks the nuanced capability for more subtle adjustments.

DoRA

- Decomposes original W = M.D and applies LoRA on D
- Intuitions:
  - LoRA to concentrate on D while also allowing the M component to be tunable simplifies LoRA's approach
  - Optimizing directional updates is made more stable through weight decomposition
- DoRA can be utilized to adjust the learning pattern, diverging from that of LoRA and aligning more closely with the pattern of FT

Experiments

- Performs better than LoRA:
  - Image/Video-Text Understanding
  - Commonsense Reasoning
  - Image/Video-Text Understanding
  - Visual Instruction Tuning
- Compatible with other variants of LoRA : VeRA(DVoRA), QLoRA(QDoRA)