⚡🚀 המאמר היומי של מייק 30.05.2024⚡🚀

2BP: 2-Stage Backpropagation

אנו יודעים שהמודלים העמוקים גדולים היום מדי כדי להיכנס לזיכרון ram של gpu אחד. עקב כך מחלקים את משקלי המודל בין הgpus השונים (sharding). זה פותר צוואר בקבוק אחד (זכרון) אבל כתוצאה מכך נוצר צוור בקבוק אחר בחישוב של backprop, המאמר הנסקר פיתח שיטה למקבל את חישוב הגרדיאנטים במהלך backprop ובכך מקל על צוואר הבקבוק הזה.

מאמר: https://arxiv.org/pdf/2405.18047