המאמר היומי של מייק - 01.03.25  
LORA VS FULL FINE-TUNING: AN ILLUSION OF EQUIVALENCE

היום סקירה מספר 200 מאז שהתחלתי לכתוב סקירות יומיות לפני 9 חודשים. אז לרגל הציון העגול הזה נסתפק בסקירה קצרה של מאמר די קליל. המאמר משווה את השפעת של פיין-טיון עם LoRA לפיין-טיון רגיל שבו מעדכנים את כל המשקולות של המודל.

אזכיר עם אם LoRA אנו מאמנים מטריצות בעלות ראנק נמוך שמתווספות למטריצות המשקלים בכל שכבה. אזכיר שמטריצה בעלת רנק נמוך מגדול mxn ניתנת לייצוג באמצעות מכפלה של מטריצה A בגודל rxn ומטריצה B בגודל mxr כאשר . (ראנק נמוך). אז ב-LoRA מאמנים מטריצות A ו-B (לגל שכבת המודל) שכאמור מכפלתם מתווספת למטריצה המשקלים במקורית W\_0 של המודל שעובר פיין-טיון. למיטב זכרוני BA מתווסף למטריצות K ו-Q בשכבת ה-attention.

אז המחברים משווים את מטריצות המשקלות המאומנות אחרי פיין טיון מלא לבין פיין טיון על LoRA. זה נעשה באמצעות השוואות של וקטורים סינגולריים של מטריצות המשקולות המאומנות זוכרים SVD שזה קיצור של Singular Value Decomposition - זה הווקטורים שמופיעים באחת המטריצות עם עמודות אורתונורמליות. למי ששכח SVD הוא שיטה מתמטית שמפרקת כל מטריצה A למכפלה של שלוש מטריצות: A = U S V^TT, כאשר U VVV הן מטריצות אורתונורמליות, ו-S היא מטריצה אלכסונית המכילה את הערכים הסינגולריים.

המחברים הגדירו מושג intruder dimension או InDim בתור כזה ש״נעלם״ אחרי הפיין טיון. כלומר לכל וקטור סינגולרי של מטריצת המשקולות המקורית W\_0 מנסים להתאים וקטור סינגולרי של אותה המטריצה אחרי פיין-טיון. ההתאמה נמדדת באמצעות דמיון קוסיין - כלומר אם לווקטור סינגולרי נתון של W\_0 לא נמצא וקטור סינגולרי של המטריצה אחרי פיין טיון בעל דמיון קוסיין גבוה מספיק הוא נקרא משויך ל-InDim.

אז התברר שעבור פיין טיון עם LoRa מספר InDims הינו גבוה יותר מאשר עם פיין טיון מלא. מעניין כי עבור r (הראנק) של LoRa ממש נמוכים (קרובים ל 1) מספר InDims לא גבוה ואז הוא מתחיל לעלות כאשר מעלים את r ומתחיל לרדת כאשר r מגיע לערכים גבוהים יחסית (נגיד 64). המחברים טוענים שאם עושים פיין טיון עם LoRa המודל ״שוכח יותר מהידע שלו״ מאשר עם פיין טיון מלא. המחברים טוענים שזה קשור לעובדה שהמודלים שטויבו בצורה מלאה מציגים ביצועים טובים יותר על דאטה OOD כלומר על out-of-distribution data.

אציין שהמחברים התמקדו בטיוב מודלי אנקודר מבוססי RoBeRTA.

https://arxiv.org/abs/2410.21228