המאמר היומי של מייק: 11.06.25  
TRANSFORMER-SQUARED: SELF-ADAPTIVE LLMS

מזמן רציתי לסקור את המאמר הזה אך הוא הלך לי לאיבוד בפייפ המאמרים הבלתי נגמר שלי (כרגע עומד על 353 מאמרים העומדים להיסקר או להיפסל לסקירה מתישהו). המאמר נכתב על ידי מדענים (בתקווה 🙂) מחברת Sakana AI שעלתה לכותרות כשהרימה AI Data Scientist (שקיבל ביקורות די טובות למיטב זכרוני). המאמר מציע שכלול מאוד פשוט לתהליך האימון של מודלי שפה בתרחישי מולטיטסקינג. כאן מולטיטסקינג אומר שאנו מאמנים כמה מודלים-מומחים (לא לבלבל עם MoE) שכל אחד מהם מתמחה במשימה מסוימת מאיזה מודל בסיס חזק. ה

כל מודל כזה מאומן בצורה דומה לאדפטרים שזה סוג של PEFT שזה Parameter Efficient Fine Tuning כלומר כאשר מספר קטן של משקלים מאומן במהלך FT. המאמר מציע שיטת PEFT הנקראת SVF שזה למעשה Singular Value Fine Tuning שמטרת להתאים את המודל למשימה נתונה. כמו שניתן להבין משמה SVF מבוססת על הערכים הסינגולריים שמקרה הזה הם ערכים סינגולריים של מטריצות המשקלים בשכבת MLP. דרך אגב MLP מכיל שתי מטריצות משקלים בכל בלוק של טרנספורמר והמאמר לא מסביר (לפחות אני לא ראיתי) איך בדיוק נבנית מטריצת משקלים בכל בלוק (אולי עושים SVF לכל מטריצה בנפרד).

אז מה בעצם עושה SVF? הוא מבצע SVD (כלומר Singular Value Decomposition) עבור מטריצות משקלים בכל בלוק טרנספורמר במודל. אחת ממטריצות אלה היא אלכסונית ואילו שתיים האחרות הן אורתוגונליות (משמאל ומימין). המחברים מכניסים למכפלה זו מטריצה אלכסונית Z נלמדת ואתה מאמנים במהלך האימון. יש כאן איזשהי הנחה שמודל הבסיס למד את כל ״המשימות האפשריות״ ובמהלך פיינטיון אנו צריכים לחזק כאלו הרלוונטיות למשימה הנלמדת.

מעניין כי פיין טיון בוצע תוך שימוש בשיטה השייכת ללמידה עם חיזוקים או RL בקצרה הנקראת REINFORCE עם רגולריזציה רגילה שמשתמשים באימוני RL של מודלי שפה. שמעתם נכון הם לא השתמשו ב- PPO, לא ב- GRPO ולא ב- DPO ובנוסף המחברים עשו זאת עבור משימות עם verifiable rewards כלומר כאלו שניתן לדעת האם התשובה נכונה למשל שאלות מתמטיות או קידוד. במהלך אימון כזה מאמנים רק מטריצות Z בכל השכבות.

באינפרנס המחברים מציעים 3 שיטות. בשיטה הראשונה בשלב הראשונה שואלים מודל לאיזה משימה שייכת שאלה ל-LLM עם פרומפט מתאים. בהתבסס על התשובה מריצים מודל עם וקטורי Z עבור המשימה הנבחרת. השיטה השנייה היא לאמון מודל דיסקרימינטיבי המזהה מה סוג המשימה עבור שאלה נתונה. השיטה השלישית מניחה דאטהסט קטן עבור משימה מסוימת למאפשר אימון של וקטור המשקול עבור כל המודלים (עבור כל המשימות). כלומר במקום לשייך שאלה למשימה מסוימת מתארים אותה כצירוף לינארי בין כל המשימות. בסוף המשימה מקבלת את הייצוג שלה (באמצעות וקטורי Z משלו).

https://arxiv.org/abs/2501.06252