⚡️🚀המאמר היומי של מייק 07.07.24:⚡️🚀

The Road Less Scheduled

היום סוקרים מאמר שלא נראה כמו מאמר למידה עמוקה רגיל. בהתחלה זה אולי יכול להיראות שהמאמר מציע עוד שכלול מי יודע מה ל-ADAM או שיטה אופטימיזציה של לוס אחרת. אבל זה לא בדיוק. המאמר כן מציע שיטת אופטימיזציה (מציאת מינימום) לפונקציות קמורות אבל זה בא ממטרה לשפר את Adam או משהו כזה אלא מציע שיטה לשיפור קצב ההתכנסות של אלגוריתם מורד הגרדיאנט (GD) הידוע.

המאמר מתחיל מכך שמבחינה תיאורטית האלגורית של (Polyak-Ruppert (PR הוא זה שאמור להביא התכנסות אופטימלי אבל בפרקטיקה זה פחות קורה (לא ברור לאיזה פרקטיקה הם מתכוונים כי התוצאות שהם נתנו מתייחסות לרשות עמוקות הלא קמורות). PR בעצם עושה אותו GD אבל העדכון האמיתי המוחלק מעריכית עם העדכון האחרון. כלומר באיטרציה t העדכון של GD נכנס עם המקדם 1/t (אפשר לשחק עם זה לפי המאמר אבל קשה להגיע לקצב החלקה אופטימלי).

המאמר מציע שיטה חדשה (3 שלבים במקום 2 ב-PR) שמשפרת ההתכנסות של PR ללא צורך בבחירה של פרמטר ההחלקה.

https://arxiv.org/abs/2405.15682