⚡️🚀המאמר היומי של מייק 22.08.24: ⚡️🚀

Approaching Deep Learning through the Spectral Dynamics of Weights

היום נסקור מאמר החוקר מה הסיבות לתופעה של גרוקינג. למי שלא מכיר גרוקינג זו תופעה די מעניינת המתרחשת כאשר ממשיכים לאמן רשת נוירונים (למרות שזה קורה גם במודלים אחרים) גם אחרי לוס הוולידציה מתחיל לעלות (כלומר אנו נכנסים למשטר אוורפיט). מתברר אם לא עוצרים וממשיכים לאמן לוס הוולידציה מתחיל לרדת כלומר המודל נכנס למשטר ההכללה כלומר לומד את ה״חוקיות האמיתית״ מאחורי הדאטה.

התופעה הזו היא מקרה פרטי של double descent (יש גם multiple descent) שמתרחש גם אם אנו מוסיפים פרמטרים למודל בצורה עקבית ומגיעים למצב שיש לנו over-parametrization. כלומר יש המודל שלנו לכאורה מתחיל ״יותר מדי פרמטרים״ כדי ״להבין את הדאטה״. וגם שם זה קורה בצורה בלתי רציפה כלומר יש אינטרוול של פרמטרים שביצועי המודל יורדים עבורם ורק אז מתחילים לרדת.

המאמר חוקר מה קורה עם משקלי המודל כאשר הוא נכנס למשטר הגרוקינג. מתברר שתופעה הגרוקינג קשורה לירידה בראנק של מטריצות המשקלים של המודל. בשבילי זה די אינטואיטיבי כי לדעתי במהלך גרוקינג המודל מצליח להתכנס ל״פתרון פשוט ביותר עבור הדאטהסט. פתרון פשוט הכוונה הוא מודל שאפקטיבית הוא קטן, כלומר רוב וקטורי המשקלים בו או אפס או תלוים לינארית זה בזה.

https://arxiv.org/abs/2408.11804