⚡️🚀המאמר היומי של מייק -02.11.24: ⚡️🚀  
Learning to Compress: Local Rank and Information Compression in Deep Neural Networks

היום סוקרים מאמר כחול לבן למחצה (אחד המחברים משניים הוא ישראלי רביד שוורץ זיו) והם חוקרים נושא שמעניין אותי מאוד באופן אישי. הנושא הוא דחיסה של דאטה באמצעות רשתות נוירונים והוא גם מאוד קשור לעבודות של נפתלי תשבי האגדי בנושא צוואר בקבוק מידעי (information bottleneck או IB) וגם השערת יריעה (manifold hypothesis או MH) בנוגע לרשתות נוירונים עמוקות.

MH טוענת שדאטה מהעולם האמיתי (כגון תמונות או טקסט) אינם מפוזרים באופן אחיד במרחב בעל מימד גבוה, אלא שוכנים על יריעה בעל מימד נמוך יותר. רשתות נוירונים עמוקות מצליחות היטב עם הדאטה הז כי הן לומדות לזהות ולנצל את המבנה של אותה יריעה, מה שמאפשר להן לבצע הכללה טובה למרות המורכבות העצומה של המרחב המקורי.

כמובן שזה קשור לדחיסה כי ניתן לראות במיפוי ממרחב בעל מימד גבוה למרחב בעל מימד נמוך שהרשתות עושות בהתאם ל MH סוג של דחיסה. ניתן לראות ״שהמימד האמיתי״ של מרחב הפיצ'רים של שכבה ברשת נוירונים קשורה לראנק(=דרגה) של היעקוביאן שלהם (הפיצ'רים) ביחס לקלט. למה זה קורה בעצם? הרי מרחב האפס של היעקוביאן מייצג כיוונים שבהם האקטיבציות של השכבה לא משתנות (כפונקציה של הקלט). ככל שמימד של מרחב האפס גדול יותר הדרגה של יריעת הפיצ'רים בשכבה נמוכה יותר. משמעות הדבר היא שהתרחשה יותר "דחיסה" או הפחתת מימדי הקלט.

נציין שמטריצות עם דרגה לא מלאה מהוות מרחב בעל מידה אפס במרחב של כל המטריצות (כמו הסתברות של כל מספר עם דוגמים יוניפורמית בין 0 ל 1). עקב כך המאמר מגדיר robust local rank או RLR שזה מספר ערכים סינגולריים (הכללה של הערכים העצמיים) של היעקוביאן שהם גדולים ממספר קטן אפסילון אך חיובי (נזכור עבור דרגה אמיתית צריך להחליף אפסילון ב 0).

אוקיי, מקווה ששרדתם את זה אז עכשיו מגיעים שני המשפטים העיקריים של המאמר. הם טוענים שברשתות עמוקות (מספר שכבות גבוה) בבעיות סיווג תמיד יהיה שכבה l שה-RLR יהיה נמוך מ-(פרופורציונלי לאפסילון בחזקה מינוס 2 ובנורמת אופרטור של מטריצת השכבה l (נורמת אופרטור זה הערך הסינגולרי הגבוה ביותר). הכוונה כאן לרשת שעושה התאמה מושלמת לדאטה האימון (עם מרג'ין 1 כלומר מצליחה להפריד בין הקטגוריות השונות בבטחה). משמעות המשפט היא שהרשת המאומנת דוחסת את הדאטה בשכבה l באופן אפקטיבי.

המחברים מוכיחים משפט דומה בנוגע לבעיות רגרסיה.

תמיד כיף לצלול למתמטיקה חמודה עם המאמרים של רביד 🙂

https://arxiv.org/abs/2410.07687