⚡️🚀המאמר היומי של מייק -04.12.24: ⚡️🚀  
KAN: Kolmogorov–Arnold Networks

האמת שזה די מחדל שב 7 חודשים מאז שהמאמר הזה התפרסם, לא סקרתי אותו. יש לו כרגע כבר 400 ציטוטים והיד עוד נטויה. אני באופן אישי מאוד אוהב מאמרים המבוססים על טענה מתמטית מוכחת ולצערי אין לנו הרבה כאלו בתקופה האחרונה.

המאמר הדי מדובר הזה מציג ארכיטקטורה חדשה המבוססת על משפט קולמוגורוב ארנולד שטוען שכל פונקציה רבת משתנים רציפה ניתנת לייצוג כסכום (כפול) של פונקציות של משתנה אחת. במילים פשוטות כל פונקציה ניתן לייצג בתור סכום של סכומים של פונקציות שכל אחת מהן היא של משתנה אחת בלבד.

משפט זה הוא ״מקביל״ ל- Universal Approximation Theorems (יש כמה כאלו) שאומרת שניתן לייצג כל פונקציה (המקיימת תנאי לא מגבילים במיוחד) על יד רשת נוירונים בעלת עומק 2 או יותר שכבות. רשתות נוירונים של היום בנויים בהתבסס על UAT (בגדול) והמאמר המסוקר מציע לבנות אותם בהתבסס על משפט KA. באופן די טבעי זה קיבל שם כן.

המודל KAN בנוי משכבות שכל אחד מהן סכום של פונקציות נלמדות (כלומר הפרמטרים בהם הם אלו שנלמדים על הדאטהסט). כל פונקציה נלמדת כזו מורכבת מצירוף לינארי של כמה b-splines (עוד פונקציה ללא פרמטרים הנקראת (silu(x.

ב-ספליין B זה פונקציה המוגדרת באינטרוול, המחולק לכמה מקטעים (נקרא grid) שמהווים פרמטרים של הבי-ספליין. B המורכב מכמה פולינומים (מדרגה 3 בד״כ) כך שלכל מקטע יש פולינום משלו. בי-ספליין משמשים לקירוב של פונקציות כאשר המקדמים לפולינום בכל מקטע נקבעים כדי למקסם את דיוק הקירוב. אז ב-KAN לומדים את את פרמטרי הגריד במטרה למזער את פונקציית הלוס של הבעיה.

וזהו זה - היה זמנו לא מעט התלהבות סביב הארכיטקטורה החדשה הזו אבל התברר שהאימון של KAN הוא לא פשוט בכלל ולא תמיד מתכנס. אבל זה לא הפריע לא לקבל 400 ציטוטים בחצי שנה עם עשרות רבות מאמרים המשך שכנראה אסקור כמה מהם. בינתיים אני לא איבדתי תקווה ב-KAN…

https://arxiv.org/pdf/2404.19756