⚡️🚀המאמר היומי של מייק 18.07.24: ⚡️🚀

Trainable Highly-expressive Activation Functions

 ממשיכים את קו הגיוון וסוקרים מאמר לא קשור ישירות למודלי שפה. היום נסקור מאמר של כמה חוקרים ישראלים המציע דרך חדשה לבנות פונקציות אקטיבציה ברשת נוירונים. היום פונקציות אקטיבציה הן לא נלמדות לרוב (ReLU, GeLU, tanh וכדומה). לפעמים פונקציות אקטיבציה מכילות hyperparameter שלא נלמד במהלך האימון אלא נקבע מראש (Leaky ReLU, Swish וכדומה).

המאמר מציע פונקציות אקטיבציה שהן(הפרמטרים שלהן) אשכרה נלמדות במהלך האימון. ד״א לאחרונה ראינו דוגמא נוספת לפונקצית אקטיבציה נלמדת ראינו לא מזמן במאמר המפורסם Kolmogorov-Arnold network או KAN - שם אלו היו ספליינים נלמדים. במאמר המסוקר אימצו שיטה אחרת לבנייה של פונקציות אקטיבציה נלמדות. הבנייה נעשה דרך שדות וקטורים שמגדירות את המסלול של נקודה במרחב.

במקרה הזה אנו מתחילים מנקודה x ובעזרת נגזרת של כיוון תנועת הנקודה(=שדה וקטורי) ב״זמן״ (שמתחיל ב t=0 ומסתיים ב- t=1) נבנה המסלול של נקודה x. המסלול מסתיים ב t=1 לכל x שלמעשה מגדיר לנו פונקציית אקטיבציה (a(x. ניתן לתאר את התקדמות נקודה באמצעות משוואה אינטגרלית (כמו שיטת אוילר לפתרון משוואות דיפרנציאליות).

המאמר מתבונן במקרה של שדה וקטורי נתון על ידי פונקציה רציפה המורכבת מפונקציות אפיניות (לינארית מוזזת) באינטרוול נתון. פונקציית זו מכיל פרמטרים נלמדים המגדירים את הפונקציות האפיניות. ניתן להראות כי פונקציות אקטיבציה היוצאות מהתהליך הזה הם diffeomorphism, כלומר פונקציה גזירה בעלת פונקציה הופכית גזירה גם כן. פונקציות כאלו נקראות CPAB. דרך אגב פונקציות אלו שימשו בעבר לטרנספורמציות ״לוקאליות״ של דאטה בסדרות זמן או של תמונות (למשל ל-time warping דינמי של סדרות זמן).

המאמר מציע לשכלל את פונקציית אקטיבציה שתיארנו קודם ומגדירים אותה לכל x ולא באינטרוול נתון. הם מגדירים באינטרוול ״הרגיל״ פונקציית אקטיבציה שהרחבנו עליהם לפני תוכפל ב-GeLU (שזה התפלגות קומולטיבית של גאוסיאן המוכפל ב- x) ובשאר האינטרוול תהיה שווה ל-x. יש גם עוד גרסה שבה במקום x פונקציית אקטיבציה שווה ל LReLU מעבר לאינטרוול שלה.

בנוסף יש איבר רגולריזציה על הפרמטרים של CPAB של פונקציית האקטיבציה המוצעת. כדי לזרז את החישובים (הרי כל פעם צריך לפתור משוואה אינטגרלית לכל אקטיבציה) המחברים מציעים לבצע קווינטוט ולחשב את ערך הפונקציה רק ב-n נקודות באינטרוול ה-CPAB שלה.

מאמר כיפי וכתוב היטב - נהניתי לקרוא.

https://arxiv.org/abs/2407.07564