⚡️🚀המאמר היומי של מייק -09.11.24: ⚡️🚀

CROSS-ENTROPY IS ALL YOU NEED TO INVERT THE DATA GENERATING PROCESS

מאמר המשך של המאמר שסקרתי אתמול שהציע שיטה חדשה ל-SSL או Self-Supervised Learning באריכות. מטרת SSL היא לבנות מודל המפיק ייצוג דאטה עוצמתי שיהיה קל לבנות ממנו מודלים downstream לביצוע משימות שונות על הדאטה הזה בתור backbone (למשל על ידי הוספת שכבות, LoRA, אדפטרים או שיטות פיין טיון אחרות הבנויות על ה-backbone הזה). כלומר הייצוג הזה צריך להיות מסוגל לזקק את כל התכונות המהותיות של הדאטה הזה כלומר לדחוסו בצורה יעילה.

משימת downstream הפשוטה ביותר היא משימת סיווג ובמקרה הזה מודל ייצוג טוב צריך להיות מסוגל להבדיל בין דאטה שייך לקטגוריות שונות (למרות שהמודל עצמו מאומן על דאטה לא מתויג). המאמר של אתמול הציע לאמן מודל שיודע לזהות פיסת דאטה מהייצוג שלה. כלומר כל פיסת דאטה מקבלת קטגוריה משלה (כלומר אם יש לנו דאטהסט עם 10L דוגמאות אז יש לנו 10K קטגוריות). בגדול מאמנים שכבה לינארית בנוסף לאנקודר (מודל הייצוג) שממפה (השכבה הלינארית) את וקטור הייצוג לקטגוריות עם לוס cross-entropy.

אז המאמר של אתמול טען שניתן להגיע לייצוגים חזקים עם השיטה הזו (למשימות downstream מסוג סיווג) והמאמר המסוקר הוכיח כמה טענות לגבי הרעיון שנדון במאמר (טוב זה לא בדיוק אבל קרוב) שסקרנו אתמול תחת הנחות די הגיוניות. המאמר די מתמטי ואנסה להסביר את הרעיון העיקרי בלי לצלול לנוסחאות וללא התעמקויות יתר לפרטים מתמטיים לא מהותיים.

המחברים מניחים כמה הנחות שעוזרות להם לחקור את הגישה הזו. ההנחה הראשונה מניחה שיש תהליך גנרטיבי המגנרט פיסות דאטה השייכים לכמה קטגוריות (מספרם ידוע). בפרט היא מדברת על כך שקיים מודל גנרטיבי g המגנרט דאטה מייצוגו הלטנטי z. המשתנה הלטנטי z בהינתן קטגוריה C מוגרל מהתפלגות von Neumann-Fisher או vMF בקצרה. vMF היא התפלגות רב מימדית על ספירה בעלת רדיוס אחת המוגדרת על ידי וקטור z\_c תוחלת ופרמטר ריכוז (סקלר המגדיר את מידת המריחות של ההתפלגות).

עכשיו המשפט הראשון במאמר טוען אם מאמנים ייצוג f (האנקודר) עלי ידי מקסום פונקציה שדומה לזאת מהמאמר הקודם רק שהקטגוריות יהיה קטגוריות של הדאטה(המיוצגות במרחב הלטנטי) ולא כל פיסת דאטה שייכת לקטגוריה משלה(נכון זה לא אותו הדבר אבל עדיין), יש פירוש די יפה לוקטורים w המרכיבים מטריצת W שהיא המיפוי הלינארי שאנו לומדים מהמרחב הלטנטי למרחב הדאטה.

במקרה הפשוט - משפט אחד מגדיר 4 מקרים, התלויים האם וקטורים w (המרכיבים את W) ווקטורי ייצוג אחרי (f(x, וקטורי w מהווים טרנספורמציה אורתוגונלית של מרכזי הקטגוריות z\_c שממנו הוקטורים הלטנטיים מוגרלים (כלומר זה אותם הווקטורים תחת סיבוב רב מימדי כלשהו). כלומר קיבלנו w\_i עם מאוד קשורים למבנה של הדאטה. בנוסף במקרה הזה ההרכבה של האנקודר f (מה שאנו מאמנים) והדקודר g הינה לינארית כלומר הצלחנו למצוא את ההופכית של הגנרטור g - וזה תוצאה די חזקה (משפט 2 מנסח את זה בצורה די טובה).

ההוכחות לא פשוטות בכלל ועם זאת המאמר הזה מאוד חשוב ואני מקווה שהצלחתי לפחות להסביר לכם את מהותו.

https://arxiv.org/abs/2410.21869