⚡️🚀המאמר היומי של מייק -16.11.24: ⚡️🚀

NON-NEGATIVE CONTRASTIVE LEARNING

מאמר מעניין בנושא הלמידה הניגודית (contrastive learning) או CL בקצרה. נזכיר שמטרת CL היא לבנות ייצוג יעיל לדאטה לא מתויג שנוכל להשתמש בו לאחר מכן לאימון מודלי לשמישות downstream שונות (למשל על ידי הוספה של כמה שכבות ייעודיות למשימה למודל שבונה את הייצוג). השיטה הפופולרית ביותר ל-CL (שלה יש וריאציות ושכלולים רבים) היא InfoNCE הוצעה לראשונה במאמר של Oord et al כבר בשנת 2018 הרחוקה.

השיטה מנסה לקרב ייצוגים של דוגמאות דומות (כגון אוגמנטציה של אותה התמונה) מבחינה דמיון קוסיין (מכפלה פנימית מנורמלת) ובאותו הזמן היא מנסה להרחיק ייצוגים של דוגמאות לא דומות (הנבחרות בד״כ באקראי). זה נעשה (בגדול) עלי ידי אימון מודל שממזער את היחס בין מרחקי הקוסיין (מעלים אותו באקספוננט) של זוגות דוגמאות שליליים (כלומר לא דומים) לזה של זוגות דוגמאות חיוביים (דומים). נציין שבכל באץ לוקחים מספר גבוה של זוגות שליליים (את הסיבות הסברתי בסקירות הקודמות בנושא).

המאמר מציע שיטה המשפרת את איכות הייצוגים הנלמדים, למשל כאלו שבהם הקטגוריות השונות של דאטה (אזכיר שמדובר באימון עם דאטה לא מתויג) יהיו מרוכזות ב״חלקים מסוימים״ (תת-וקטורים) של וקטורי הייצוג כאשר שאר הערכים יהיו אפסים או מאוד קרובים ל-0. וקטורים כאלו יהיו נוחים יותר משימות downstream הקשורים לסיווג דאטה. המאמר טוען ששיטת CL עם פונקציית לוס בסגנון InfoNCE לא מצליחות להפיק ייצוגים עם תכונות כאלו והסיבה העיקרית היא האינווריאנטיות שלהם לסיבוב הנובעת מהצורה של פונקציית הלוס שלהם (הסבר מפורט בפרק 2.1 במאמר).

המחברים מציעים שני חידושים עיקריים. קודם כל הם מציעים לאמן ייצוגים שהם לא שליליים (ב-InfoNCE אין שום מגבלה כזו). החידוש השני הוא פונקציית לוס שאכן מכילה מכפלות פנימיות של וקטורי ייצוג הדאטה אבל בלי אקספוננטים ויחסים (כבר הוצע קודם אבל ללא אי שליליות). הפעם פונקצית הלוס היא הפרש בין המרחק הריבועי בין הדוגמאות השליליות לבין המרחק בין הדוגמאות החיוביות.

מהחברים מצטטים מאמר שהראה שהייצוגים המופקים על ידי המודל הממזער לוס זה ללא הגבלה של אי שליליות הינם שקולים לאלו המתקבלים מפקטוריזציה סימטרית (מייצגים מטריצה כמכפלה של מטריצה F והשחלוף שלה) של מה שנקרא מטריצת co-occurrence A. לקח לי קצת זמן להבין מה זה בדיוק אבל בגדול זה מטריצה המכילה סוג של ״הסתברויות״ של שתי דוגמאות יהיו חיוביות (אוגמנטציה של אותה הדוגמא).

כלומר אם יש לנו דאטהסט של 1000 דוגמאות ו-10 אוגמנטציות שונות פר דוגמא מטריצה A בגודל 10K x 10K מכילה 1/10 לזוגות חיוביים (כאשר תמונות i ו- j הן אוגמנטציות של אותה התמונה) 0 בשאר המקומות. מדובר כאן בפקטוריזציה למטריצה F שהיא low-rank כלומר אחד המימדים שלה (מימד הייצוג של דאטה) הוא הרבה יותר קטן מהמימדים של מטריצה A (שהיא עצומה לדאטהסטים בגודל רציני, מיליוני תמונות).

אז המאמר משתמש באותו הלוס אבל מחפש וקטורי ייצוג שהם אי שליליים (מפעילים עליהם פונקציות כגון ReLU, sigmoid, softplus וכדומה). בנוסף המחברים שמו לב כי בייצוגים המתקבלים יש נוירונים מתים כלומר כאלו שמאוד קרובים ל-0 עבור כל הדוגמאות). המחברים משתמשים בטריקים נחמדים כמו stop-gradient כדי להתמודד עם התופעה הזו.

בסוף מקבלים ביצועים משופרים כאשר הייצוגים המתקבלים הינם יותר disentangled ויותר קרובים לאורתוגונליות לדאטה מקטגוריות שונות.

https://arxiv.org/abs/2403.12459