לילה טוב חברים, היום אנחנו שוב בפינתנו deepnightlearners עם סקירה של מאמר בתחום הלמידה העמוקה.
היום בחרתי לסקירה את המאמר שנקרא:
Exemplar VAE: Linking Generative Models, Nearest Neighbor Retrieval, and Data Augmentation

שיצא לפני בערך חודש
הוצג בכנס: NeurIPS 2020

תחומי מאמר:

- אוטו-אנקודר וריאציוני ((variational autoencoder - VAE
- מודלים גנרטיביים לא פרמטריים שיוצרים דאטה "ישירות מדוגמאות של סט האימון" (exemplar generative models -EGM)

כלים מתמטיים, מושגים וסימונים:

- משערך חלון של פארזן (parzen window - PW) או שערוך צפיפות בעזרת קרנל (kernel density estimation - KDE)
- חסם תחתון של ELBO evidence
- מרחק KL בין מידות הסתברות
- תערובת גאוסיאנים ((gaussian mixture

תמצית מאמר: המאמר מציע לשלב שתי גישות ליצירת דאטה (גינרוט): אוטו-אנקודר וריאציוני VAE וגישה לא פרמטרית ליצירת דאטה ישירות מהדוגמאות מסט האימון, הנקראות EGM. שיטות ממשפחת EGM יוצרות דגימות חדשות ע"" בחירה באקראי של דוגמא מסט האימון והפעלה של טרנספורמציה עליה. אחד היתרונות של שיטות אלו הינה הקלות של עדכון המודל: כאשר דאטה חדש נוסף לדאטה סט אין צורך באימון נוסף. החיסרון הבולט של גישה זו הינו צורך בהגדרת מטריקה במרחב הדאטה הנדרשת בשביל להגדיר סביבה של נקודת דאטה. למידת מטריקה כזו במרחבים בעלי מימד גבוה כמו בדומיין הויזואלי היא מאוד קשה. חיסרון נוסף של שיטות מהסוג הזה הינו צורך לשמור את כל הדאטה סט בשביל ליצור דגימות חדשות שעלול להיות די יקר מבחינת גודל הזיכרון (עבור משימות מסוימות זה גם עלול להיות בעייתי בהיבט הפרטיות)

לעומת זאת מודלים גנרטיביים פרמטריים לדוגמא אוטו-אנקודר וריאציוני VAE, GAN, זרימה מנרמלת (normalized flow) ושיטות אחרות ממשפחה זו מבוססות על רשתות נוירונים עמוקות מסוגלות ללמוד התפלגויות מורכבות במרחבים במימד גבוה. במודלים גנרטיביים פרמטריים רשת נוירונים מאומנת ליצור דוגמאות ש"נראות טבעי" מדגימות של וקטורים אקראיים עם רכיבים בלתי תלויים(מהמרחב הלטנטי) מהתפלגות נתונה לא פרמטרית (!!). התפלגות זו הינה גאוסית עם מטריצת קוואריאנס אחידה ווקטור תוחלת אפס ברוב המקרים. אחרי שהאימון מסתיים אין לנו צורך לאחסן את סט האימון. עולם אם נרצה להוסיף עוד דוגמאות לדאטה סט נצטרך לסיבוב נוסף של אימון (צריך לציין שלהבדיל מהסיבוב הראשון לא נעשה את האימון מאפס אלא נעשה סוג של כיול (fine-tuning) של המודל מהסיבוב הראשון.

המאמר מציע לשלב את שתי גישות אלה במטרה ליהנות מיתרונותיה של כל אחד מהם.

רעיון של מאמר במשפט אחד: לאמן VAE עם הפריור (prior) מעל המרחב הלטנטי שהוא תערובת של גאוסיאנים GM (או חלון פארזן PW) מעל הייצוגים(הקודים) הלטנטיים (שמהווים את המרכזים של הגאוסיאנים) של הדוגמאות מהדאטה סט.

למעשה תערובת גאוסיאנים מעל הקודים הלטנטיים של דוגמאות מהדאטה סט ניתן לראות בתור משערך צפיפות קרנלי (KDE) מעל המרחב הלטנטי של הדאטה סט. ל- VAE בעל פריור זה (הנקרא Examplar VAE או Ex-VAE בקצרה) יתרון משמעותי על מודלים גנרטיביים לא פרמטריים: לא צריך לשמור את הדוגמאות במרחב

המקורי שלהם (במרחב בעל מימד גבוה) וניתן להסתפק רק את הייצוגים הלטנטיים שלהם שדורשים הרבה פחות מקום אחסון. מצד שני כאשר עוד נוספות נקודות לדאטה סט, לא מוכרחים לאמן את המודל מחדש. אציין שבמקרה זה הייתי עושה fine-tuning לרשת המקודדת (שבונה קוד לטנטי של דוגמא) מכיוון שהדוגמאות שנוספו עשויים לתרום הייצוגים הלטנטיים שהיא יוצרת. דרך אגב ניתן לאמן את Ex-VAE על חלק מהדאטה סט וליצור דוגמאות חדשות על שאר הדוגמאות (שלא השתתפו באימון).

הסבר של רעיונות בסיסיים: אז כל העסק הזה עובד? נתחיל את הדיון מרענון לגבי מה זה VAE:

<u>הסבר קצר על VAE:</u>

VAE מכיל שתי רשתות נוירונים

- הרשת המקודדת N_enc שבונה ייצוג(קוד) לטנטי של דאטה. הקלט ל N_enc הינו נקודת דאטה **x** והפלט הינו פרמטרים(!!) של התפלגות פוסטריור של P(z|**x.**)למשל אם התפלגות פוסטריור הוגדרה כגאוסית אז הפלט של הרשת הוא וקטור התוחלות ומטריצת קוואריאנס וקטור הייצוג). לאחר מכן מגרילים וקטור z עם פרמטרים אלו (למעשה עושים זאת דרך טריק של רפרמטריזציה ולא עי" דגימה ישירה).
- הרשת המפענחת N_dec (דקודר) המקבלת כקלט קוד ייצוג z והופכת אותו לדגימה מהמרחב המקורי. המטרה של דקודר הינה לשחזר כמה שיותר מדויק את הדוגמא **x** שממנו נוצר הקוד הלטנטי z.

פונקציית לוס של VAE נגזרת מהחסם התחתון של evidence (נקרא ELBO) ומורכבת משני איברים:

1. לוס השחזור L_rec המשערך עד כמה טוב הצלחנו לשחזר את **x**
2. מרחק KL בין התפלגות פריור P_pr(z) נתונה ולבין התפלגות הפוסטריורית P(z|**x**) המיוצגת עי" הרשת המקודדת N_enc (באופן לא מפורש). המטרה של איבר זה הינה לכפות על P(z|**x**)להיות קרובה ל- P_pr(z). ב VAE הסטנדרטי ההתפלגות P_pr(z) בדרך כלל נבחרת כגאוסית עם וקטור תוחלות אפס ומטריצת קוואריאנס יחידה. הקירוב של P(z|**x**)המחושב עי" N_enc נקרא הקירוב הוריאציוני - נסמן אותו ב- q(z|**x.**)

מי שצריך הסבר יותר מפורט על VAE מוזמן להביט ב- פוסט המעולה הזה על VAE.

<u>הערת לגבי התפלגויות הפריור והפוסטריור של VAE:</u> ניתן לראות את P_pr(z) ב - VAE גם בתור "התפלגות יעד" בשביל P(z|**x**). זה נובע מהעובדה שאחת המטרות של אימון VAE הינו מזעור של מרחק KL בין P(z|**x**) ל- P_pr(z).

כמו שכבר הזכרנו Ex-VAE מהווה שפצור של ה- VAE הסטנדרטי כאשר הפריור P_pr הינו פרמטרי המוגדר כתערובת גאוסיאנים P_mix(z|**x**) עם המרכזים בייצוגים(קודים) הלטנטיים של הדוגמאות. נציין שלכל גאוסיאן בתערובת זו יש מקדם$1/N$כאשר N זה מספר הדוגמאות (examplars) המשמשות לאימון של Ex-VAE (כבר ציינתי שלא חייבים להשתמש בכל הדאטה סט לאימון). פונקציית בלוס של Ex-VAE מאוד דומה לזו של VAE הרגיל ומכילה שני איברים: לוס השחזור - זהה ל VAE והשני הינו מרחק KL בין הקירוב הוריאציוני של הפוסטריור q(z|**x**) לבין P_mix(z|**x.**) ברוח ההסבר הניתן בהערה לגבי הפריור והפוסטריור, אחת המטרות של האימון היא "לכפות" על התפלגות הפוסטריור להיות קרובה ככל האפשר לתערובת גאוסיאנים P_mix(z|**x**) המהווים שערוך קרנלי של הצפיפות מעל המרחב הלטנטי (ראה הסבר בפסקה הקודמת).

אז איך מאמנים את הדבר הזה? קודם כל נציין שכאן אנו מאמנים 3 רשתות נוירונים:

- הרשת המקודדת הרגילה N_enc שהוכפת את הקוד הלטנטי דגימה מהדומיין המקורי
- הרשת המפענחת N_dvar המיועדת לבניית קירוב וריאציוני של התפלגות הפוסטריור q(z|**x.**) נציין ש q(z|**x**) ממודלת עי" גאוסיאן עם מטריצת קוואריאנס אלכסונית כאשר כל איבר באלכסון הינו פונקציה של **x** (הממודלת עי" הרשת)

- הרשת המפענחת N_dmix המיועדת לשערוך של התפלגות תערובת הגאוסיאנים (**x**|z)P_mix- ״התפלגות יעד״ עבור (**.x**|z)qנוסיף כאן כי (**x**|z)P_mixממודלת עי״ גאוסיאין עם אותו וקטור תוחלות כמו (**x**|z)qאך עם מטריצת קוריאנס קבועה (אלכסונית עם ערך קבוע על האלכסון)

<u>תהליך האימון:</u> מכיוון ש Ex-VAE הינו סוג של VAE קלאסי ופונקציית הלוס שלו דומה לזו המקורית של VAE אתמקד רק בהבדלים החשובים של באימון בין VAE ל- Ex-VAE

1. חישוב של (**x**|z)P_mixבנקודה z עלול להיות כבד חישובית אם N (מספר הדוגמאות המשתתפים באימון של Ex-VAE) גבוה. הסיבה לכך נעוצה בעובדה (**x**|z)P_mixהינו סכום של N גאוסיאנים (**x_**i|z)r (עבור דוגמא **x_**i) וצריך לחשב ערך של כל אחד מהם. המאמר מציע לקחת רק את הדוגמאות הכי קרובות ל- z במרחב הלטנטי מבחינת המרחק האוקלידי (כמובן שאי אפשר לדעת לאיזה דוגמאות הייצוג הלטנטי z הכי קרוב בכל איטרצית אימון והם בונים איזשהו קאש על סמך האיטרציות הקודמות ומעדכנים אותו כאשר מגלים דוגמא עם הקוד לטנטי קרוב מספיק ל- z). הם קוראים לשיטה הזו kNN (k השכנים הכי קרובים אבל שימו לב שלא מתבצע קליסטור אמיתי כלשהו במהלך האימון).
2. הם מוחקים איבר המתאים לדוגמא **x_**i מתערובת הגאוסיאנים (**x**|z)P_mixכאשר מעדכנים את המשקלים של הרשתות לדוגמא **x_**i. לטענת המאמר זה מונע התכנסות לפתרונות טריוויאליים המרוכזים מדי בקודים הלטנטיים של הדוגמאות מהדאטה סט.

הישגי מאמר: המאמר משווה את איכות דגימות הנוצרות באמצעות Ex-VAE בשלושה רובדים שונים:

1. שערוך צפיפות ההסתברות: הם מראים שההסתברות הממוצעת של הדגימות הנוצרות עם Ex-VAE מהטסט סט הינה גבוהה יותר מאשר השיטות המתחרות.
2. הם מראים שעבור כמה דאטה סטים הקלאסטרים (לקטגוריות שונות של דוגמאות) במרחב הלטנטי שנוצרים עם Ex-VAE הם יותר מופרדים מאשר המתחרות. נציין כי Ex-VAE מאומן ללא שימוש בלייבלים כלל (!!)
3. בם מראים שאם יוצרים דוגמאות חדשות q1עם Ex-VAE כדי להגדיל דאטה סט, שיפור בביצועים במשימת סיווג המושג, יותר גבוה מהגישות המתחרות

דאטה סטים:MNIST, Fashion-MNIST, Omniglot, CelebA

לינק למאמר: https://arxiv.org/abs/2004.04795

לינק לקוד: https://github.com/sajadn/Exemplar-VAE

נ.ב. מאמר נחמד עם רעיון למודל גנרטיבי שלא נתקלתי בו בעבר. מסקרן האם גישה כזו או השכלול שלה מסוגלת להתחרות באיכות התמונות עם SOTA בתחום הזה, קרי GANs. אני גם מחכה למחקרים חדשים בנושא שיטות גנרטיביות לא פרמטריות

#deepnightlearners