# Review 49: Exemplar VAE: Linking Generative Models, Nearest Neighbor Retrieval, and Data Augmentation

**Paper: https://arxiv.org/abs/2004.04795v3**

פינת הסוקר:

המלצת קריאה ממייק: חובה רק למי שמתעניין Exemplar Models וגם מבין קצת ב- VAE - לאחרים ניתן להסתפק בסקירה :).

בהירות כתיבה: בינונית.

רמת היכרות עם כלים מתמטיים וטכניקות של ML/DL הנדרשים להבנת מאמר: הבנה טובה בעקרונות VAE, ידע בסיסי ב- kernel density models.

יישומים פרקטיים אפשריים: יצירה של דוגמאות חדשות למטרת אוגמנטציה של דטאסטים קיימים למשימות שונות.

פרטי מאמר:

פורסם בתאריך: 04.03.21, בארקיב.

הוצג בכנס: NeurIPS 2020.

תחום מאמר:

variational autoencoder - VAE.

מודלים גנרטיביים לא פרמטריים שיוצרים דאטה "ישירות מהדוגמאות 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.

מודלים של תערובת גאוסיאנים (gaussian mixture models).

מבוא:

המאמר מציע לשלב שתי גישות ליצירת דאטה (גינרוט): VAE וגישה לא פרמטרית ליצירת דאטה ישירות מהדוגמאות מסט האימון, הנקראת EGM. שיטות ממשפחת EGM יוצרות דגימות חדשות ע'" בחירה באקראי של דוגמא מסט האימון והפעלת טרנספורמציה עליה. אחד היתרונות של שיטות אלו הינה הקלות של עדכון המודל: כאשר דאטה חדש נוסף לדאטה סט, אין צורך באימון נוסף. החיסרון המשמעותי של גישה זו הוא הצורך בהגדרת מטריקה במרחב הדאטה, הנדרשת להגדרת ״סביבה של נקודת דאטה״. למידת מטריקה כזו במרחבים בעלי מימד גבוה כמו בדומיין הויזואלי היא מאוד קשה. חיסרון נוסף של שיטות מסוג זה הוא הצורך לשמור את כל הדאטהסט 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 GAN ,VAE, זרימה מנרמלת (normalized flow) וגישות פרמטריות נוספות מבוססות על רשתות נוירונים עמוקות מסוגלות ללמוד התפלגויות מורכבות במרחבים במימד גבוה. במודלים גנרטיביים פרמטריים רשת נוירונים מאומנת ליצור פיסות דאטה חדשות ש״נראות טבעי״ מדגימות של וקטורים אקראיים בעלי רכיבים בלתי תלויים (הנקראים הווקטורים הלטנטיים) מהתפלגות נתונה לא פרמטרית (!!). התפלגות זו (הנקראת התפלגות פריורית) היא בדרך כלל (אך לא בהכרח) גאוסית עם מטריצת קווריאנס אחידה וקטור תוחלות אפס. אחרי שהאימון הסתיים אין לנו צורך לשמור את סט האימון. אולם אם נרצה להוסיף דוגמאות חדשות לדאטהסט, נצטרך סיבוב נוסף של אימון (צריך לציין שלהבדיל מהסיבוב הראשון לא נעשה את האימון מאפס אלא נעשה סוג של כיול (fine-tuning) של המודל שהתקבל מהסיבוב הראשון.

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

תמצית מאמר:

המאמר מציע לאמן VAE עם התפלגות הפריור (מעל המרחב הלטנטי) שהיא תערובת של גאוסיאנים (gaussian mixture) כאשר המרכז (וקטור תוחלות) של כל גאוסיאן הוא הייצוג הלטנטי של דוגמה מהדאטהסט. למעשה, ניתן לראות תערובת גאוסיאנים מעל הוקטורים הלטנטיים של דוגמאות מהדאטהסט בתור משערך צפיפות קרנלי (KDE) להתפלגות של המרחב הלטנטי של הדאטהסט. ל- VAE בעל פריור זה (הנקרא Examplar VAE או Ex-VAE בקצרה) יש יתרון משמעותי על מודלים גנרטיביים לא פרמטריים: לא צריך לשמור את הדוגמאות במרחב המקורי שלהם (מרחב בעל מימד גבוה) וניתן להסתפק רק בייצוגים 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:

נתחיל את הדיון מרענון של עקרונות VAE:

הסבר קצר על VAE:

VAE מורכב משתי רשתות נוירונים.

הרשת המקודדת N\_enc (אנקודר) שבונה ייצוג (וקטור) לטנטי של דאטה. הקלט ל-N\_enc הינו דוגמה x והפלט הינו פרמטרים (!!) של התפלגות פוסטריור של הוקטור הלטנטי של x, כלומר הפרמטרים של (P(z|x. למשל אם התפלגות פוסטריור היא גאוסית, הפלט של N\_enc הוא וקטור התוחלות ומטריצת קווריאנס של וקטור הייצוג). לאחר מכן מגרילים וקטור z עם פרמטרים אלו (למעשה עושים זאת דרך טריק של רפרמטריזציה ולא ע"י דגימה ישירה).

הרשת המפענחת N\_dec (דקודר) המקבלת כקלט וקטור ייצוג z והופכת אותו לדגימה מהמרחב המקורי. המטרה של הדקודר הינה לשחזר באופן כמה שיותר מדויק את הדוגמא x שממנו נוצר הקוד הלטנטי z.

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

לוס השחזור L\_rec המשערך עד כמה טוב הצלחנו לשחזר את פיסת הדאטה המקורית x.

מרחק KL בין התפלגות פריור (P\_pr(z נתונה לבין התפלגות הפוסטריורית (P(z|x הממודלת באמצעות הרשת המקודדת N\_enc. המטרה של איבר זה הינה לכפות על (P(z|x להיות קרובה ל-(P\_pr(z - ניתן לראות אותו כאיבר רגולריזציה. כאמור (P\_pr(z בדרך כלל נבחרת כגאוסית עם וקטור תוחלות אפס ומטריצת קווריאנס 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.

הערת לגבי התפלגויות הפריור והפוסטריור של VAE:

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.

כאמור Ex-VAE מהווה הכללה של ה-VAE המקורי כאשר התפלגות הפריור P\_pr הינה פרמטרית ומוגדרת כתערובת גאוסיאנים (P\_mix(z|x עם המרכזים בייצוגים הלטנטיים של הדוגמאות. נציין כי לכל גאוסיאן בתערובת זו יש מקדם כאשר N זה מספר הדוגמאות (examplars) המשמשות לאימון של Ex-VAE (המאמר מציין שלא חייבים להשתמש בכל הדאטהסט לאימון).

פונקציית הלוס של Ex-VAE היא מאוד דומה לזו של VAE המקורי ומכילה שני איברים:

לוס השחזור - זהה ל VAE

מרחק KL בין הקירוב הווריאציוני 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 של המרחב הלטנטי של הדאטהסט.

אז איך מאמנים את Ex-VAE? קודם כל נציין כי Ex-VAE מורכב מ-3 רשתות נוירונים:

הרשת המקודדת הרגילה N\_enc שהופכת דגימה מהדומיין המקורי לוקטור הלטנטי שלה.

הרשת המפענחת N\_dvar המיועדת לבניית קירוב וריאציוני של התפלגות הפוסטריור (q(z|x. נציין כי (q(z|x ממודלת עי״ גאוסיאן עם מטריצת קווריאנס אלכסונית כאשר כל איבר באלכסון הינו פונקציה של x (הממודלת עי״ הרשת).

הרשת המפענחת N\_dmix, בעלת אותם המשקלים הנלמדים כמו הרשת N\_dvar, המיועדת לשערוך של התפלגות תערובת הגאוסיאנים (P\_mix(z|x - ״התפלגות יעד״ עבור (q(z|x. למעשה N\_dmix משערכת את (P(z|x\_i עבור הדוגמאות x\_i (קרבות ל-x במרחב הלטנטי) המשמשות לבנייה של פיסת דאטה, הדומה ל-x. נציין כי (P(z|x\_i ממודלת עי״ גאוסיאין עם אותו וקטור תוחלות כמו (q(z|x אך עם מטריצת קווריאנס קבועה אלכסונית.

תהליך האימון:

מכיוון ש-Ex-VAE הינו סוג של VAE קלאסי ופונקציית הלוס שלו דומה לזו המקורית של VAE אתמקד רק בהבדלים החשובים בין האימון של VAE ושל Ex-VAE.

נציין כי החישוב של (P\_mix(z|x עלול להיות כבד חישובית אם N (מספר הדוגמאות המשתתפים באימון של Ex-VAE) גבוה. הסיבה לכך נעוצה בעובדה ש-(P\_mix(z|x הינו סכום של N גאוסיאנים (r(z|x\_i (עבור דוגמא x\_i) ויש צריך לחשב ערך של כל אחד מהם. המאמר מציע לקחת רק את הדוגמאות הכי קרובות ל- z במרחב הלטנטי מבחינת המרחק האוקלידי. מכיוון שאי אפשר לדעת לאיזה דוגמאות הייצוג הלטנטי z הכי קרובות בכל איטרציה של אימון, והמאמר מציע לשמור מערך של כמה דוגמאות קרובות מהאיטרציות הקודמות. מערך זה מתעדכן כאשר מתגלה דוגמא עם הוקטור לטנטי קרוב מספיק ל- z. שיטה זו נקראת במאמר k) kNN השכנים הכי קרובים) אבל שימו לב שלא מתבצע קליסטור אמיתי כלשהו במהלך האימון).

המאמר מציע לא להשתמש באיבר המתאים לדוגמא x\_i מתערובת הגאוסיאנים (P\_mix(z|x, כאשר מעדכנים את המשקלים של הרשתות לדוגמא x\_i. לטענת המאמר זה מונע התכנסות לפתרונות טריוויאליים המרוכזים מדי בוקטורים הלטנטיים של הדוגמאות מהדאטהסט.

הישגי מאמר:

המאמר משווה את איכות דגימות הנוצרות באמצעות Ex-VAE בשלושה תרחישים הבאים:

שערוך צפיפות ההסתברות: המאמר מראה כי ההסתברות הממוצעת של הדגימות הנוצרות באמצעות Ex-VAE הינה גבוהה יותר מאשר של שיטות המתחרות.

עבור דאטהסט מתויג, מאמנים את Ex-VAE ללא שימוש בתיוגים. המאמר מראה כי עם Ex-VAE, הקלאסטרים של קטגוריות שונות במרחב הלטנטי, יותר מופרדים מאשר עם השיטות המתחרות.

כאשר יוצרים דוגמאות חדשות עם Ex-VAE כדי להגדיל דאטהסט, המאמר מראה שיפור בביצועים במשימת סיווג ביחס לגישות המתחרות.

דאטהסטים:

MNIST, Fashion-MNIST, Omniglot, CelebA.

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, כלומר GANs. בנוסף אני מחכה לראות מחקרים נוספים בנושא שיטות גנרטיביות לא פרמטריות.