

סקירה זו היא חלק מפינה קבועה בה אני סוקר מאמרים חשובים בתחום ה-ML/DL, וכותב גרסה פשוטה וברורה יותר שלהם בעברית. במידה ותרצו לקרוא את המאמרים הנוספים שסיכמתי, אתם מוזמנים לבדוק את העמוד שמרכז אותם תחת השם [deepnightlearners](#).

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

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:

לינק למאמר: [זמין להורדה](#).

לינק לקוד: [כאן](#).

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](#))
- [\(density estimation - KDE\)](#).
- [סום 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"י 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 VAE, GAN, 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 פרירית) 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 כיוול (fine-tuning) של המודל שהתקבל מהסיבוב הראשון.

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

תמצית מאמר:

המאמר מציע לאמן VAE עם התפלגות הפריר (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 Exemplar 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).

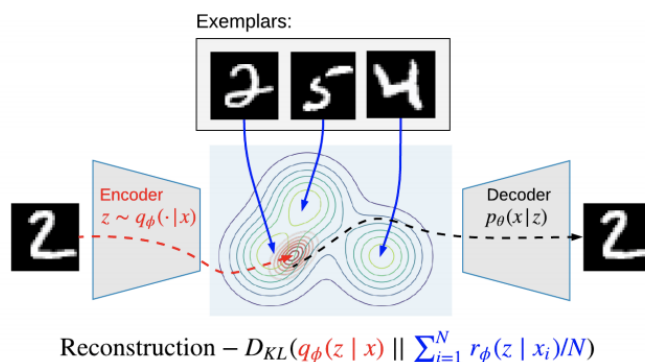


Figure 1: Exemplar VAE is a type of VAE with a non-parametric mixture prior in the latent space. Here, only 3 exemplars are shown, but the set of exemplars often includes thousands of data points from the training dataset. The objective function is similar to a standard VAE with the exception that the KL term measures the disparity between the variational posterior $q_\phi(z|x)$ and a mixture of exemplar based priors $\sum_{n=1}^N r_\phi(z|x_n)/N$.

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 פוסטריר של הוקטור הלטנטי של x , כלומר הפרמטרים של $P(z|x)$. למשל אם התפלגות פוסטריר 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:

1. לוס השחזור L_{rec} המשערך עד כמה טוב הצלחנו לשחזר את פיסת הדאטה המקורית x .

2. מרחק KL בין התפלגות פריור $P_{pr}(z)$ נתונה לבין התפלגות הפוסטרירור $P(z|x)$ 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 ע"י 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 והפוסטרירור 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 $1/N$ כאשר N זה מספר הדוגמאות (examples) 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 הוריאציוני של הפוסטרירור $q(z|x)$ לבין $P_{mix}(z|x)$. ברוח ההסבר הניתן בהערה לגבי הפריור והפוסטרירור, אחת המטרות של האימון היא "לכפות" על התפלגות הפוסטרירור להיות קרובה ככל האפשר לתערובת גאוסיאנים $P_{mix}(z|x)$, המהווים שערך 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 הפוסטרירור $q(z|x)$. נציין כי $q(z|x)$ ממודלת ע"י גאוסיאן עם מטריצת קווריאנס אלכסונית כאשר כל איבר באלכסון הינו פונקציה של x (הממודלת ע"י 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 ע"י גאוסיאן 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.

1. 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 kNN (k השכנים הכי קרובים) אבל שימו לב שלא מתבצע קליסטור אמיתי כלשהו במהלך האימון).

2. 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.

Method	Dynamic MNIST	Fashion MNIST	Omniglot
VAE w/ Gaussian prior	-84.45 \pm 0.12	-228.70 \pm 0.15	-108.34 \pm 0.06
VAE w/ VampPrior	-82.43 \pm 0.06	-227.35 \pm 0.05	-106.78 \pm 0.21
Exemplar VAE	-82.09 \pm 0.18	-226.75 \pm 0.07	-105.22 \pm 0.18
HVAE w/ Gaussian prior	-82.39 \pm 0.11	227.37 \pm 0.1	-104.92 \pm 0.08
HVAE w/ VampPrior	-81.56 \pm 0.09	-226.72 \pm 0.08	-103.30 \pm 0.43
Exemplar HVAE	-81.22 \pm 0.05	-226.53 \pm 0.09	-102.25 \pm 0.43
ConvHVAE w/ Gaussian prior	-80.52 \pm 0.28	-225.38 \pm 0.08	-98.12 \pm 0.17
ConvHVAE w/ Lars	-80.30	-225.92	-97.08
ConvHVAE w/ SNIS	-79.91 \pm 0.05	-225.35 \pm 0.07	N/A
ConvHVAE w/ VampPrior	-79.67 \pm 0.09	-224.67 \pm 0.03	-97.30 \pm 0.07
Exemplar ConvHVAE	-79.58 \pm 0.07	-224.63 \pm 0.06	-96.38 \pm 0.24
PixelSNAIL w/ Gaussian Prior	-78.20 \pm 0.02	-223.68 \pm 0.03	-89.59 \pm 0.07
PixelSNAIL w/ VampPrior	-77.90 \pm 0.02	-223.45 \pm 0.02	-89.50 \pm 0.13
Exemplar PixelSNAIL	-77.95 \pm 0.01	-223.26 \pm 0.01	-89.28 \pm 0.12

Table 3: Density estimation on dynamic MNIST, Fashion MNIST, and Omniglot for different methods and architectures, all with 40-D latent spaces. Log likelihood lower bounds (nats), estimated with IWAE with 5000 samples, are averaged over 5 training runs. For LARS [2] and SNIS [36], the IWAE used 1000 samples; their architectures and training procedures are also somewhat different.

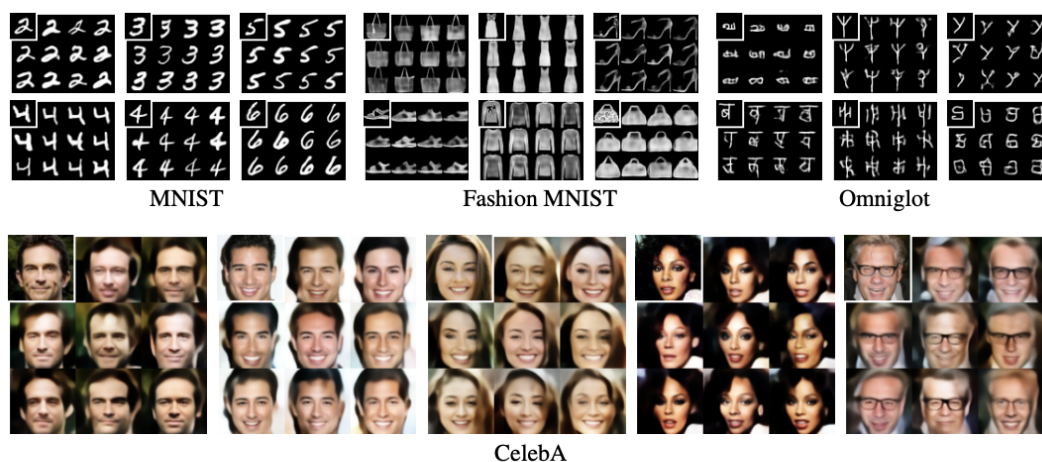


Figure 3: Given a source exemplar on the top left of each plate, Exemplar VAE samples are generated, showing a significant diversity while preserving properties of the source exemplar.

הישיג 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.

3. 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.

#deepnightlearners

הפוסט נכתב על ידי מיכאל (מייק) ארליכסון, PhD, Michael Erlihson. מיכאל עובד בחברת סייבר Salt Security בתור Principal Data Scientist. מיכאל חוקר ופועל בתחום הלמידה העמוקה, ולצד זאת מרצה ומנגיש את החומרים המדעיים לקהל הרחב