לילה טוב חברים, היום אנחנו שוב בפינתנו DeepNightLearners עם סקירה של מאמר בתחום הלמידה העמוקה
היום בחרתי לסקירה את המאמר שנקרא:
VAEBM: A symbiosis between autoencoders and energy-based models
שיצא לפני שבועיים.

תחומי מאמר: מודלים גנרטיביים, energy-based models(EBM), variational autoencoder(VAE)

מאמר הוצג בכנס: לא הצלחתי לאתר

כלים מתמטיים במאמר:
reparameterization trick, Langevin dynamics, markov chain monte-carlo(MCMC)

תמצית מאמר: המאמר מציע לבנות מודל גנרטיבי עי" שילוב VAE ו- EBM בשביל ליהנות מהיתרונות של שניהם: היכולת של EBM לייצג התפלגויות מורכבות בצורה מדויקת והיכולת של VAE לגנרט דגימות בצורה מהירה ויעילה. השילוב שלהם נותן מענה לחולשות העיקריות של שתי השיטות האלו: עבור EBM זו דגימה מאוד איטית המגבילה שימוש בגישה זו רק לגינרוט תמונות בגודל קטן ועבור VAE זה יכולת מידול לא מדויקת של התפלגות הדאטה המתבטא ביצירת תמונות מטושטשות. המאמר מציע ארכיטקטורה הנקראת VAEBM, המורכבת מרכיבי VAE ו- EBM, הנקראת VAEBM. היא מנצלת את היכולת של רכיב ה- VAE שלה בשביל ללמוד את המבנה הכללי של המרחב הלטנטי מחד ומנצלת את רכיב ה- EBM שלה בשביל "לתקן" את ה- VAE ב"איזורים שאין בהם דאטה אמיתי". בנוסף VAE מאפשר להאיץ את יכולת הדגימה של EBM עי" רפרמטריזציה שלה במרחב הלטנטי.

רעיון בסיסי: הם מגדירים את המודל המגנרט h(x,z) כמכפלה של p_vae(x,z), המודל המגנרט הרגיל של VAE (כאשר x שייך למרחב התמונות ו z הינו המשתנה הלטנטי) ו- p_ebm(x) המהווה את הביטוי הסטנדרטי עבור פונקציה התפלגות של EBM (שזה בעצם התפלגות גיבס). כמובן של p_vae ו- p_ebm יש את הפרמטרים שלהם שאותם אנו מאמנים בשביל למקסם את ה- log של h(x) על הדאטה סט כאשר h(x) זה פונקצית ההתפלגות של x אחרי המרגינליזציה של המשתנה הלטנטי z מהביטוי עבור h(x,z). כמו שמראים במאמר המיקסום הישיר של ביטוי זה מחייב דגימות של ההתפלגות האפוסטריורית של x המגונרט עם VAE ויש לזה סיבוכיות חישובית גבוהה (MCMC). נציין שפונקצית לוס כאן הינה סכום של הלוסים הסטנדרטיים של VAE ושל EBM (כל אחד עם הפרמטרים שלו). אז המאמר מציע לבצע את המיקסום בשני שלבים: קודם כל על הלוס של VAE כאשר הפרמטרים של EBM מוקפאים ורק אז למקסם את רכיב הלוס של EBM כאשר הפרמטרים של VAE מוקפאים. בנוסף בשביל להקל על הדגימה מ- h(x,z) הנחוצה בשביל שערוך הגרדיאנט של רכיב EBM, המאמר מציע לעשות רפרמטריזציה משותפת של x ושל המשתנה הלטנטי z. בעצם גישה זו מונעת דגימה דו שלבית: דוגמים את z ואז את x בהינתן z שעלול להיות מאוד בעייתי כ- p(x|z) עשויה להיות מרוכזת באיזור מאוד קטן ו- MCMC מתקשה להתמודד עם המצב הזה

תקציר מאמר: קודם כל בואו נרענו את זכרוננו וניזכר מה זה בעצם VAE ו EBM

<u>אוטו אנקודר וריאציוני (VAE):</u> זה מודל גנרטיבי שהומצא ב 2014 ומהווה הכללה סטוכסטית של אוטו אנקודר רגיל כאשר אנו מגדירים מראש את ההתפלגות על המרחב הלטנטי. VAE מורכב משתי רשתות: אנקודר ודקודר שלכל אחד יש המשקלים שלו. המטרה של האנקודר הינה לקחת תמונה x (לא חייב להיות תמונה כמובן אך המאמר מתמקד רק בדומיין הזה) ולבנות ייצוגה במרחב הלטנטי כאשר המטרה של הדקודר הינה לשחזר את התמונה מייצוג זה. כבר אמרנו ש VAE זו הגרסא הסטוכסטית של AE רגיל אז אחרי שמקבלים וקטור ייצוג לטנטי מהאנקודר, דוגמים מהתפלגות המוגדרת עי" וקטור זה ואת הדגימה הזו מעבירים הקלט לאנקודר לגינרוט התמונה. בשביל לבצע אופטימיזציה (GD) לפי הפרמטרים של האנקודר משתמשים בטריק של רפרמטריזציה - דוגמים מהתפלגות

קבועה ולא תלויה בפרמטרים ו"הופכים" את הדגימות כאילו הן שנדגמו מההתפלגות התלויה בווקטור הייצוג עי" פעולה אריתמטית גזירה (לדוגמא מגרילים מהתפלגות גאוסית סטנדרטית ו"הופכים" את הדגימות לכאלו שנדגמו מהתפגלות גאוסית עם תוחלת ומטריצת קוואריאנס נתונה עי" פעולה לינארית פשוטה). הלוס של VAE מורכב מלוס השחזור (עד כמה טוב אנו יודעים לשחזר את התמונת הקלט לאמקודר) והלוס על המרחק בין ההתפלגות המטרה של המרחב הלטנטי ובין ההתפלגות המושרת עי" דגימות מהאנקודר (הנמדד במרחק KL).

<u>מודלים מבוססי אנרגיה ((EBM</u>: זה גם מודל גנרטיבי כאשר הרעיון כאן זו בנייה מפורשת של פונקציה התפלגות p_ebm על מרחב התמונות עי" מקסום של פונקצית מטרה. בשביל לגנרט תמונות צריך לדגום מ p_ebm שבדרך כלל עושים זאת עם אחד הסוגים של MCMC. בגלל הסיבוכיות הגבוהה של דגימה זו, כרגע ניתן לגנרט עם EBM רק תמונות קטנות (עד 64x64).

פונקציית מטרה כאן זו תוחלת של הלוג של p_ebm על הטריין סט. p_ebm מוגדרת עי" התפלגות גיבס שזה בעצם אקספוננט שלילי של E(x) (תלויה בפרמטרים ונקראת פונקצית אנרגיה) מוכפלת בקבוע נרמול (בשביל ש p_ebm יהוה פונקציית התפלגות). EBM מאומן בעזרת גרדיאנט דסצנט כאשר הגרדיאנט של הלוס מורכב מהפרש תוחלות של E(x) על הדגימות מ- p_ebm (השלב השלילי) ועל הטריין סט (השלב החיובי). מכיוון שלא ניתן לדגום מ p_ebm(x) בצורה מפורשת משתמשים באחד הסוגים של MCMC - בדרך כלל בדינמיקה של לנגווין (LD). LD זה תהליך איטרטיבי הבונה דגימות של p_ebm עי״ הזזת דגימות בכיוון הגרדיאנט (על x) של פונקצית אנרגיה (שזה לוקח את רוב הזמן באימון של EBM).

<u>איך בעצם משלבים את EBM ו- VAE:</u> אפשר לראות שאם לוקחים את המודל המגנרט h(x,z) אז את פונקצית הלוס של VAEBM ניתן לחסום מלמטה עי" הסכום של הלוס הסטנדרטי של VAE והלוס של EBM. נציין שהאופטימיזציה של L_ebm כאשר כוללת שיערוך של גרדיאנט של L_ebm לפי הפרמטרים של VAE. שערוך זה הינו מאוד כבד מבחינה חישובית כי זה כולל דגימות מההתפלגות פוסטריורית על דגימות של VAE. אז מה שמציעים במאמר זה לאפטם L_ebm ו- L_vae לסירוגין שמונע את הצורך לגזור את הראשון לפי הפרמטרים של VAE. הרעיון השני של המאמר זה שימוש בטריק של רפרמטריזציה על x ו- z (המשתנה הלטנטי) <u>בו זמנית</u> שמונע את הצורך לדגום מההתפלגות המותנת x|z שעלול להיות בעייתי כאן (הוסבר בפרק "רעיון בסיסי").

כאחד ההרחבות של תהליך האימון הם מציעים לבצע כמה איטרציות GD בשביל לקרב את פונקציית ההתפלגות של x אחרי שלב האופטימיזציה של VAE לפונקצית האנרגיה E (מנסים להביא למינימום את מרחק KL ביניהם). זה מזכיר את העדכון של הגנרטור ב Wasserstein GAN.

<u>לסיכום</u>: בשלב הראשון VAE מנסה לשערך את ההתפלגות האמיתית ובשלב השני EBM מתקן אותה. <u>במאמר טוענים שמכיוון ש VAE מצליח לבנות קירוב לא רע ואז לא נדרשת מספר צעדים גבוה עבור הפרמטרים של EBM בשביל לאפטם אותה</u>. חוץ מזה VAE בונה את המרחב הלטנטי (מימד נמוך) של ההתפלגות האמיתית שמתבטא בהתפלגות יותר "חלקה" מההתפלגות האמיתית שגורם לדגימה יותר יעילה עם MCMC.

הישגי מאמר: הם משווים את הביצועים של VAEBM מול מודלים גנרטיביים רבים אחרים מסוגים שונים ומראים את עליונותו במונחי inception score(IS) ו- freshet inception distance (FID) על רובם (האמת שהיחיד שמנצח אותם ב FID על 10CIFAR זה 2StyleGAN שיש לו ארכיטקטורה מורכבת בהרבה) . נציין (וזה די חזק למרות שהם מראים זאת רק לדאטה סט אחד) ש- VAEBM מצליח לשחזר את כל המודים (modes) של StackedMNIST (כל תמונה בדאטה סט הזה הינו שילוב של 3 תמונות ב MNIST המקורי אז יש 1000 מודים ו- VAEBM מצליח לשחזר את כולם) להבדיל מכמה מודלים עדכניים של GAN (למרות הייתי רוצה לראות שם עוד כמה...). המאמר גם משווה את יעילות דגימה מול מודל גינרוט חזק denoising score matching ומציין ש- VAEBM יעיל יותר מפי 12 ממנו כאשר איכות התמונות המגונרטות היא די קרובה.

דאטה סטים: SVHN, CIFAR100, CelebA, StackedMNIST

לינק למאמר: https://arxiv.org/abs/2010.00654

לינק לקוד: אין קוד רשמי בארקייב

נ.ב. לדעתי הרעיון של המאמר די חזק בתחום המודלים הגנרטיביים ואני מניח שזו רק ההתחלה ויבואו עוד שיפורים ושיפצורים. בינתיים לא ברור האם רעיון זה יוכל להעמיד מתחרה רציני ל GANs.

#deepnightlearners