⚡️🚀המאמר היומי של מייק -28.10.24: ⚡️🚀  
HEAVY-TAILED DIFFUSION MODELS

המאמר עם השם הקצר הזה משך את עיניי כי יש לי חיבה גם למודלי דיפוזיה גנרטיביים וגם להתפלגויות בעלות תכונות מעניינות למשל זנבות כבדים. בגדול התפלגות נקראת בעלת זנב כבד או ארוך כאשר התפלגות לזנב שלה (כלומר המסה ההסתברותית מימין לנקודה) מקשל (הסתברות) הינה גבוה יותר מאשר להתפלגות מעריכית. נשמע קצת מסובך אבל במילים פשוטות ניתן להגיד כי להתפלגויות בעלות זנב כבד(HT) יש יותר מסה בקצוות.

למשל התפלגות נורמלית אינה בעלת זנבות כבדים והתפלגות סטודנט t וגם התפלגות קושי הן כן. אוקיי, למה אני בכלל מדבר על זה? הסיבה היא די פשוטה - ההנחה שנוכל להניח התפלגות גאוסית על כל סוג של דאטה אינה נכונה. יש סוגי דאטה שלא ניתן לאפיין אותם בצורה טוב עם התפלגות בעלות זנבות קלים. עקב גם אנו נתקשה לגנרט דאטה מהתפלגויות אלו אם נמדל אותו (הדאטה) עם מודלי הבנויים על הנחות גאוסיות גם אם המודלים האלו הם בעלי expressiveness גבוהה כמו מודלי הדיפוזיה. עדיין יהיה מאוד בעייתי ליצור באמצעותם דאטה בעלת התפלגות HT במיוחד בקצוות ההתפלגות.

אז המאמר, שהוא אחד הכבדים ביותר מתמטית מאלו שראיתי לאחרונה, מציע להחליף את התפלגויות גאוסיות שיש לנו במודלי דיפוזיה בהתפלגות סטודנט שהיא התפלגות HT. כלומר כל מה שהיה בעלת התפלגות גאוסית במודל דיפוזיה מקורי יהיה מהתפלגות t. דרך אגב אחד הפרמטרים של התפלגות t (שהיא כמובן וקטורית עבור מודלים אלו כי אנו רוצים לגנרט דאטה בעלת מימדים רבים) שהוא שולט ב״כבדות הזנב״ שלה וכאשר היא שואפת לאינסוף אנו מקבלים את ההתפלגות הגאוסית האהובה עלינו. כלומר המודלים המוצעים במאמר הם הכללה של מודלי דיפוזיה גאוסיים שאנו מכירים ואוהבים.

כמובן שלא מספיק סתם להחליף התפלגות גאוסית במודל דיפוזיה בהתפלגות t - זה דורש להגדיר לא מעט התפלגויות מותנות הנדרשות לנו להגדרת הלמידה של תהליך denoising. זה די לא טריוויאלי אבל העקרון נשאר דומה -מאמנים את המודל להסיר רעש (שהוא מפולג עם t) באופן הדרגתי. במקום KL divergence המוכר לנו ממודלי דיפוזיה המחברים משתמשים ב-γ-Power divergence כדי למדוד מרחק בין ההתפלגות הדאטה אחרי הסרת רעש לזה של הדאטה האמיתי (לכל איטרציה).

גם תהליך הגנרוט מוגדר דומה עקרונות למודלי דיפוזיה גאוסיים אבל כמובן כל ה-hyperparameters מותאמים להתפלגות t. יש גם רפרמטריזציות שאנו כה אוהבים במודלי דיפוזיה, ייצוג באמצעות משוואות דיפרנציאליות חלקיות, גם באמצעות טכניקה חדשה הנקראת flow matching (הבונה מסלול מיטבי בין ההתפלגות ההתחלתית והתפלגות הדאטה). כאמור מאמר די כבד מתמטית ומקווה שהצלחתי להסביר לכם את העקרונות לפחות.

https://arxiv.org/pdf/2410.14171