המאמר היומי של מייק - 08.02.25  
Rejection Sampling IMLE: Designing Priors for Better Few-Shot Image Synthesis

היום עושים הפסקה קלה עם LLMs וסוקרים מאמר המציע שיטה מעניינת לאימון מודלי גנרטיביים במקרה שיש לכם מעט דאטה לאימון. כידוע מודלים גנרטיביים מודרניים כמו מודלי דיפוזיה, גאנים, VAEs מצריכים כמות עצומה של דאטה אבל לפעמים אין לנו את הלוקסוס הזה ואנו צריכים לאמן על כמות קטנה של דאטה. האם זה אפשרי בכלל?

התשובה על כך חיובית (לפחות לפי המאמר). המחברים מציעים שיטה הנקראת RS-IMLE לאימון מודל גנרטיבי עם מעט דאטה שמשכלל שיטת IMLE שזה Implicit Maximum Likelihood Estimation. בגדול מאוד IMLE די דומה לשיטה גנרטיבית סטנדרטית - היא דוגמת משתנה בעל התפלגות קלה לדגימה (גאוסית) z ומאמנת מודל גנרטיבי(רשת נוירונים) כדי לגנרט פיסת דאטה. ההבדל הוא בפונקציית לוס: עם IMLE לכל דגימה x מהדאטהסט אנו ממזערים את רק המרחק בינה לבין נקודה z\_i אחת בלבד: כזו ש-(T(z\_i שלה הינו קרוב ביותר אליה. כאן (T(z\_i היא פיסת דאטה שגונרטה מ-z\_i ו- T זה המודל שאנו מאמנים.

כלומר בשלב הראשון של IMLE אנו דוגמים m נקודות ומעבירים אותם דרך מודל T(נקרא לו מיפוי בהמשך) ובונים m פיסות דאטה מגונרטות. לאחר מכן לכל דגימה x\_j מדאטהסט האימון אנו בוחרים את z\_i הקרובה ביותר ל-x\_j. בסוף רק נקודות כאלו משתתפות במזעור של פונקצית לוס. כמובן שמספר הנקודות m המגונרטות בשלב הראשון צריך להיות גבוה משמעותית מאשר גודל הדאטהסט לאימון n. המטרה של שיטת אימון זו היא לאפטם את המודל רק עבור הנקודות במרחב הלטנטי (z) שהן הממופות קרוב לנקודת מהדאטהסט.

הבעיה עם הגישה הזו שההתפלגות של הנקודות ״הנבחרות״ במהלך האימון כבר לא גאוסית שעלול ליצור לנו בעיות באינפרנס כי אנו כן רוצים לדגום את z מהתפלגות גאוסית. המרחק בין מיפוי T של דגימה גאוסית מנקודה מהדאטהסט שונה בהתפלגות מזה של הדגימה z הממופה הכי קרוב לקודה זו (האמת זה די ברור). דרך אגב המאמר מוכיח את הטענה הזו ומציע שיטה להתגבר על זה.

השיטה שהמאמר מציע נראית ממש פשוטה אך מבוססת על ניתוח מתמטי די מעמיק של התפלגויות המרחקים. בשלב הראשון של האימון (אחרי הדגימה מהתפלגות גאוסית) בוחרים את z\_i כאשר נופלים במרחק יותר גדול מבוע אפסילון מכל נקודות בדאטהסט האימון אחרי המיפוי (כלומר יש לנו rejection sampling). לאחר מכן, בדומה ל-IMLE, לכל נקודה x בדאטהסט בוחרים את z שהמיפוי שלו עם T נופל הכי קרוב אליה ומאמנים את T למזער את המרחק הממוצע בין z-s הנבחרים לנקודות העוגן שלהם. הייפרפרמטרים החשובים כאן זה אפסילון ומספר נקודות z שנדגמות.

אינטואיטיבית זה עובד כי מלכתחילה אנו בוחרים נקודות רחוקות יותר (לאחר המיפוי) מהנקודות בדאטהסט שמאפשר לשמור התפלגות של הנקודות הנבחרות בשלב לאחר מכן קרובה לגאוסית.

https://arxiv.org/abs/2409.17439