המאמר היומי של מייק - 20.02.25  
Losing dimensions: Geometric memorization in generative diffusion

המאמר מציג מסגרת תיאורטית חדשה להבנת האינטראקציה בין הכללה לזיכרון במודלי דיפוזיה גנרטיביים מנקודת מבט גיאומטרית. המחקר משתמש בטכניקות של פיזיקה סטטיסטית, גיאומטריה דיפרנציאלית ותורת המטריצות האקראיות כדי לנתח כיצד מודלי דיפוזיה לומדים ומאבדים תת-מרחבים של הדאטה בהתאם לגודל הדאטה, קיבולת המודל הגנרטיבי, ושיטת האימון.

### חדשנות מרכזית:

קונספט הזיכרון הגיאומטרי:  
המאמר מציג את הרעיון שזיכרון(memorization) במודלי דיפוזיה גנרטיביים אינו תופעה בינארית (או הכללה או זיכרון מוחלט של הדאטה), אלא משהו שמתרחש באופן סלקטיבי בתת-מרחבים מסוימים של מרחב הדאטה. תובנה זו מרחיבה תיאוריות קיימות שפירשו את תופעת הזיכרון בתור ״תהליך קריסה מוחלטת״ של זיכרון נקודות דאטה מסוימות. המחברים מצאו כי תהליך הזיכרון מוביל לאובדן ממדים ביריעה הלטנטית, ולא קריסה לנקודות דאטה פרטניות.

אובדן ממדים סלקטיבי ותלות בשונות:  
המחברים מראים שתת-מרחבים עם שונות גבוהה נעלמים בשלב מוקדם יותר של האימון, באופן פרדוקסלי המצביע על כך שהתכונות הבולטות ביותר של דאטה הן הראשונות להיפגע מהשפעות תופעת הזיכרון. ממצא זה חשוב משום שהוא מצביע על כך שזיכרון אינו משפיע באופן אחיד על כל הממדים, אלא מתרחש באופן מבני בהתאם לפיזור השונות של הדאטה.

שימוש בכלים מפיזיקה סטטיסטית ותורת המטריצות האקראיות:  
המחקר מבוסס על תיאוריות של מעבר פאזה הלקוח מפיזיקה סטטיסטית, במיוחד מעבר פאזה זכוכיתי (glassy phase transition) במודלי זכרון אסוציאטיביים (dense associative memory) המווה וריאנט מודרני של מודל הופפילד (זה שקיבל פרס נובל לפני כמה חודשים). המאמר מספק ביטויים אנליטיים עבור דינמיקה להפרשים בין ערכים סינגולריים של מטריצת היעקוביאן של פונקציית ה-score במהלך גנרוט הדאטה (עם מודל מאומן), ומדגים כיצד פערים אלו נסגרים באופן הדרגתי כאשר המודל מתחיל ״לזכור״ (memorize) את דוגמאות האימון. המחברים עושים שימוש באנלוגיה למודל האנרגיה האקראי (REM) לניתוח מעבר העיבוי (condensation transition), המתאר את הזמן הקריטי(האיטרציה של הדיפוזיה) שבו זיכרון משתלט.

הבחנה בין score אמפירי ל-score מדויק:  
המאמר מבחין בין פונקציית ה-score תיאורטית (מדויקת) לבין הפונקציה שמשוערכת מתוך דאטהסט. פונקציית ה-score האמפירי מציגה תנודות סטטיסטיות התלויות בגודל הדאטה ובזמן הדיפוזיה, באופן הקשור ישירות לזיכרון הגיאומטרי. המחקר מכמת כיצד מספר דרגות החופש הגנרטיביות האפקטיביות מתפתח כפונקציה של גודל הדאטהסט וזמן הדיפוזיה(איטרציות של הדיפוזיה).

אימות אמפירי על דאטה סינתטי וריאלי:  
המחברים מאשרים את התיאוריה שלהם באמצעות ניסויים על דאטה סינתטיים (למשל, יריעות ליניאריות המכילים תתי-מרחב בעלי שונות משתנים) ועל הדאטהסטים האמיתיים (MNIST, CIFAR-10, CelebA). תוצאות הניסוי מתיישבות היטב עם התחזיות התיאורטיות שפותחו במאמר, ומדגימות כי מודלי דיפוזיה מאומנים מציגים את האובדן ההדרגתי החזוי של הממדיות ככל שגודל הדאטהסט פוחת. המאמר מציג שיטה חדשה להערכת הממדיות הפנימית של דגימות שנוצרו, מה שמאשר את ההשערות התיאורטית.

https://arxiv.org/abs/2410.08727