

מעבדה בלמידה עמוקה מפגש 4

מגישים

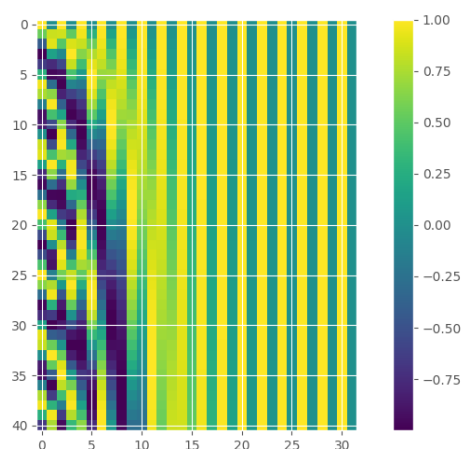
אביב שם-טוב ואיזאר חסון

חלק 1

מטלה 3

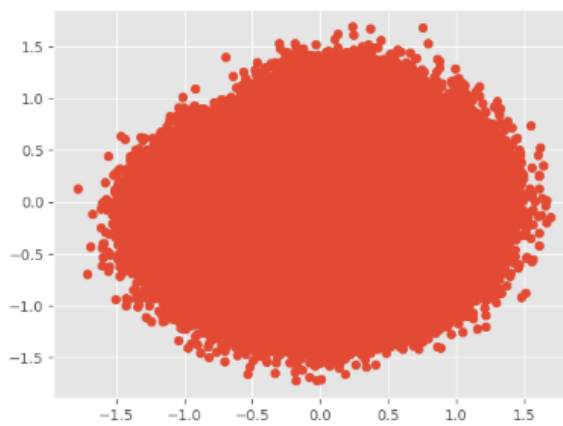
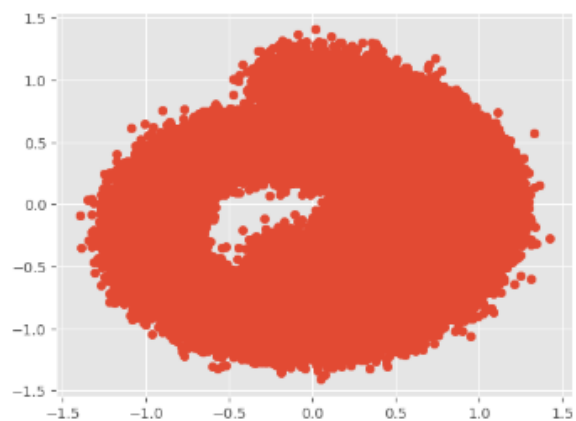
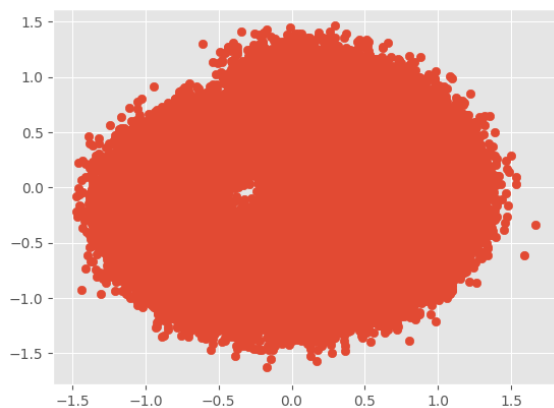
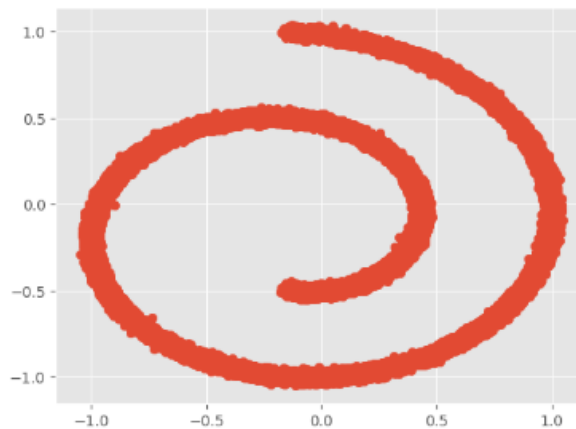
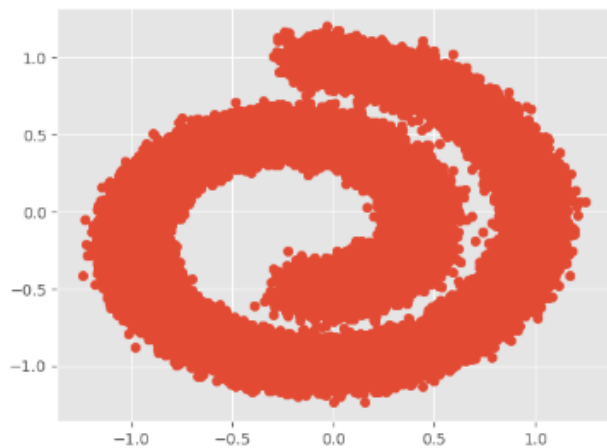
ראינו כי במודלי שפה, מטרת ה-positional embedding היא לשמור על הסדר של המילים במשפט. הוא היה דו מימדי בגלל שכבת ה-embedding והרצף של הטוקנים.

בתמונות לעומת זאת, הפיקסלים עצמם אינם רצף חד-מימדי (לא ממוקמים בשורה ישרה, אלא על משטח). לכן, positional embedding במודלי דיפוזיה של תמונות נועד לתת למודל מידע על המיקום היחסי של כל פיקסל ביחס לפיקסלים האחרים בתמונה.

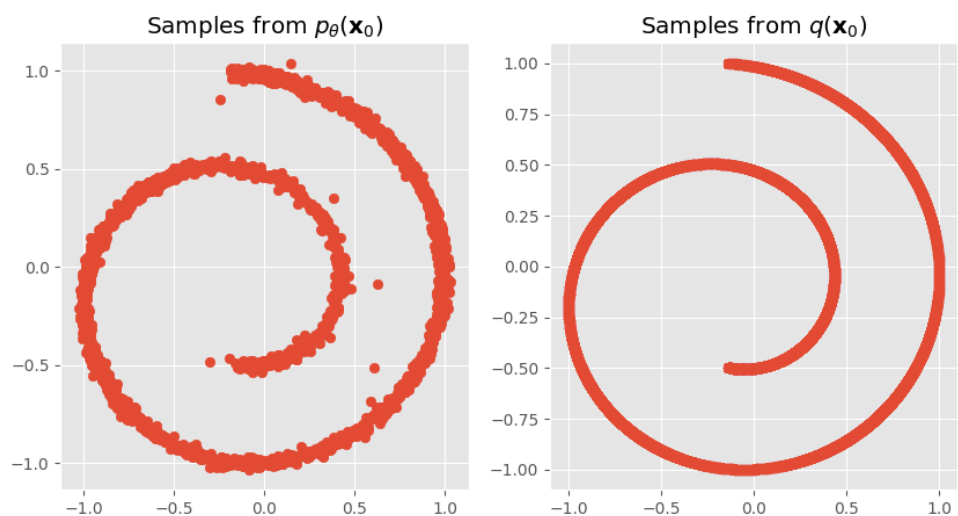


מטלה 4

ככל ש t -גדל אנו רואים את הרעש סביב הספירלה המקורית גדל (השונויות גדלה). כאשר עבור $t = 37$ אנחנו כבר רואים עיגול מלא של נקודות ולא ספירלה, כלומר התמונה כבר מאוד רועשת ואיבדה את המשמעות המקורית שלה. עבור t קטן מאוד (1) אנחנו רואים את הספירלה עם רעש קל מאוד.



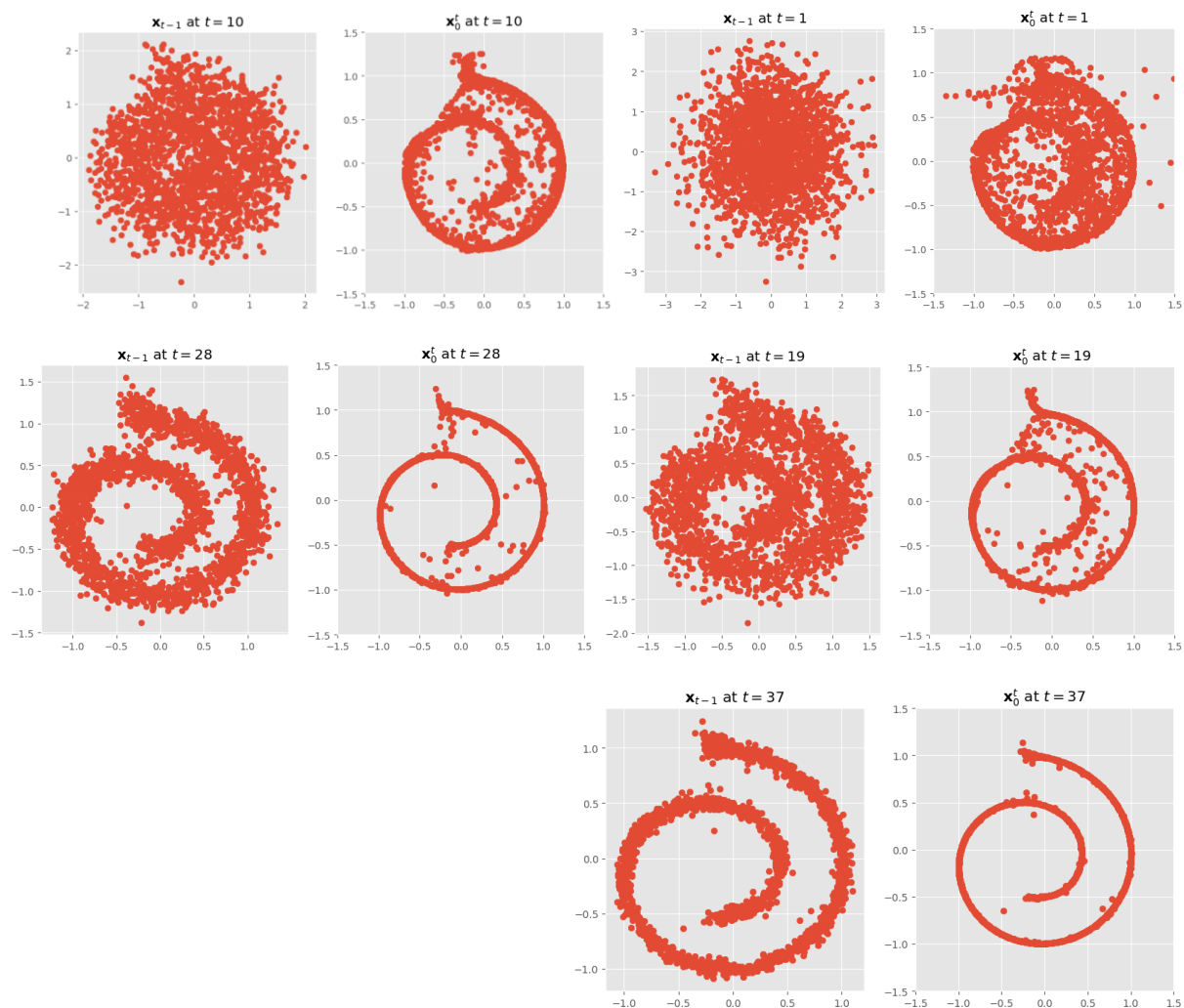
מטלה 7



כפי שניתן לראות עבור $p_{\theta}(x_0)$ קיבלנו בקירוב תוצאה מאוד דומה ל- $q(x_0)$ וזאת על אף שהתחלנו מרעש טהור. כלומר הוסר רעש מהנקודות כך שהן תעבורנה לאזור מתאים מבחינת הפילוג של הדוגמאות בדאטה שעליו בוצע האימון.

מטלה 8

אנחנו רואים שבתוצאות, כפי שראינו בתאוריה, ככל שמבצעים יותר צעדים התוצאה x_t בצעד הבא (x_{t-1}) וכן של x_0 בנקודה הזו משתפרות. בכל צעד אנחנו לא ממשיכים עם x_0 של הצעד הקודם, כי אנחנו עדיין רוצים שתהיה לנו שונות לרעש שנוכל להפחית אותה בצעדים הבאים.



חלק 2:

מטלה 2:

תיקון תמונות לאחר המעבדה: (רעש ההתחלתי לא היה נכון בקובץ הקוד)



ניתן לראות שבכל צעד הרעש נעלם בהדרגתיות, זה אנלוגי למקרה הסיפורי הדו מימדי, שם "תמונה של פרצוף אדם" היה מבנה הספירלה, והרעש בתמונה היה רעש בתזוזות ב X - Y , אך התהליך בעצם זהה.

מטלה 5:

גנרציית תמונות ללא טקסט:



אם לא ניתן למשתמש גישה לשנות את הטקסט שאליו המודל מנסה להתקרב, נקבל תמונות רנדומליות, ללא שום משמעות.

הבעיה עם הדבר היא שהמשתמש צריך להכווין את המודל באופן ספציפי מאוד, כלומר המשתמש לא יכול להגיד למודל "התמונה השמחה בעולם", המודל אינו יכול לענות על שאלות, הוא צריך את "התשובה" ואותה לצייר.