מבוא ללמידה עמוקה - תרגיל 3

עידן רפאלי ואנאל בן־סימון

22 בינואר 2021

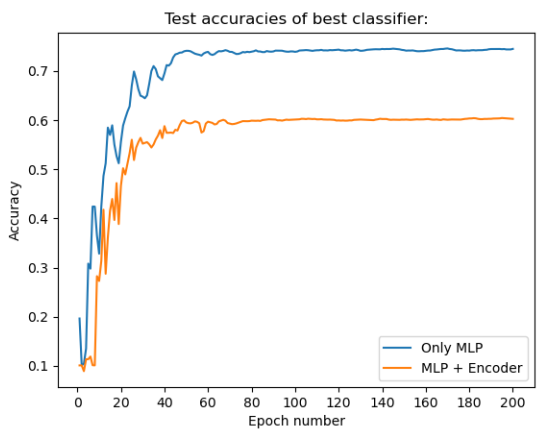
חלק תכנותי

1. סקרנו את קוד ה־Auto-Encoder, אימנו אותו והשתמשנו במודל המאומן במשימות הבאות.

2. לאחר שניסינו מספר ארכיטקטורות שונות למודל ה־classifier לסיווג ספרות ממרחב ה־latent של תמונות ה־MNIST שנלמד על־ידי ה־AE, עם קבוצת אימון בגודל של 50 דוגמאות בלבד, קיבלנו שהארכיטקטורה שהשיגה את הדיוק הגבוהה ביותר על קבוצת הולידציה היא הארכיטקטורה המתוארת להלן:

   - שכבה ראשונה עם מימד פלט של 128, ואקטיבציית sigmoid
   - שכבה שניה עם מימד פלט של 512, ואקטיבציית sigmoid
   - שכבה שלישית עם מימד פלט של 10, ואקטיבציית softmax

ניתן לראות את הארכיטקטורות של כל המודלים שניסינו בקובץ הקוד שהגשנו. המודל הטוב ביותר השיג אחוז דיוז של 0.745 באפוק האחרון. אותו מודל השיג אחוז דיוק של 0.607 כאשר הוא אומן לאחר שהוא הורכב על ה־encoder (מספר האפוקים הכולל בשני האימונים הוא 200). להלן גרף המתאר את הדיוק של המודל הטוב ביותר על פני ריצת האפוקים, בשני המקרים:

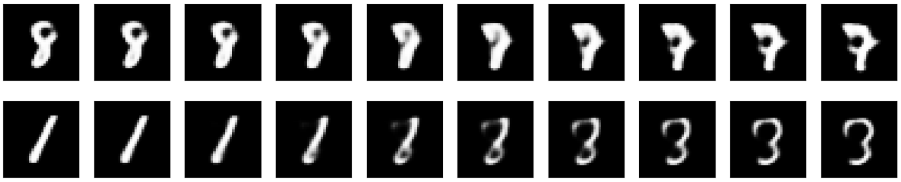


לפי הגרף, ניתן לראות כי עדיף לאמן את מודל ה־classifier לבדו, מבלי להרכיב אותו על ה־encoder מכיוון שהוא לבדו מצליח להשיג אחוזי דיוק גבוהים יותר (הבחנו בתופעה דומה גם במודלים האחרים שבדקנו).

3. להלן מספר דוגמאות לדגימות שנוצרו על־ידי מודל GAN שאימנו (לאחר שהשתמשנו ב־decoder כדי להמיר את הדגימות שנוצרו במרחב ה־latent למרחב התמונות:

הדגימות נראות לרוב כמו ספרות אמיתיות, ולכן אנו מסיקים שמודל ה־GAN למד את מרחב ה־latent של תמונות הספרות בצורה טובה יחסית.

לאחר שביצענו את משימת האינטרפולציה בין 2 תמונות, כאשר פעם אחת דגמנו 2 תמונות ממרחב ה־latent על־ידי הגנרטור, ובפעם השניה לקחנו 2 תמונות אמיתיות מקודדות במרחב ה־latent, קיבלנו את התוצאות הבאות:



השורה הראשונה היא אינטרפולציה בין 2 תמונות שנדגמו על־ידי הגנרטור, והשורה השניה היא אינטרפולציה בין 2 תמונות אמיתיות. לדעתנו התוצאות הטובות יותר התקבלו באינטרפולציה בין 2 תמונות שהתבצעו על־ידי הגנרטור. ניתן לראות בשורה הראשונה שלאורך כל האינטרפולציה, כמעט כל תמונה נראית קרובה יחסית לספרה אמיתית כלשהי (מתחילים מספרה שדומה ל־8, מתישהו עוברים לספרה שנראית דומה יותר ל־9 ולבסוף מסיימים בספרה שנראית כמו 7). לעומת זאת, בשורה השניה, התמונות באמצע תהליך האינטרפולציה לא נראות ברורות כל כך לדעתנו, ונראות כמו הכלאה בין 2 ספרות אולי, אך לא דומות לאף ספרה ספציפית. לכן אנו מסיקים שאיטרפולציה במרחב ה־latent הנוצר על־ידי הגנרטור טובה יותר מאינטרפולציה במרחב ה־AE.

4. בחרנו לאמן רשת מסוג Conditional GAN לצורך המשימה של ייצור תמונה לפי ספרה ספציפית שמתקבלת כקלט מהמשתמש. בהינתן ספרה כלשהי $0 \leq d \leq 9$ שתמונה שלה אנו מעוניינים לייצר על־ידי הגנרטור, נאמן גנרטור שמקבל כקלט וקטור שנדגם מהתפלגות גאוסיינית, ואליו נשרשר וקטור בגודל 10 שהוא ייצוג one-hot של הספרה (כלומר מכיל 1 בקורדינטה ה־d ו־0 בשאר הקורדינטות). בתהליך האימון, הדיסקרימינטור יקבל תמונות מקודדות על־ידי הגנרטור, שאליהן ישורשר גם כן וקטורים one-hot שהגנרטור השתמש בהן לייצור הקודים, ובנוסף הדיסקרימינטור נקבל קודים של תמונות אמיתיות שנוצרו על־ידי ה־AE, שאליהן משורשרים ייצוגי one-hot של הספרות המקוריות שהיו בתמונות מהן הקודים נוצרו (התיוגים מהדאטא של ה־MNIST).
לאחר שאימנו מודל Conditional GAN כמפורט לעיל, וביקשנו ממנו לייצר את כל אחת מהספרות מ־0 עד 9, קיבלנו את התוצאה הבאה:

התוצאות לדעתנו טובות, ודומות מאוד לספרות, ולכן אנו מסיקים שהמודל למד את ההתפלגות המותנית בצורה טובה.

שאלות תאורטיות

1. נשתמש ברשת דומה ל־BERT כדי לחקות התנהגות של שכבת קונבולוציה בצורה הבאה: אם למשל נרצה ללמוד פילטר בגודל $k \times s$, נחתוך את כל חלונות מתמונת הקלט בגודל $k \times s$ סביב כל אחד מהפיקסלים, כך שיווצרו לנו n תמונות בגודל $k \times s$, כאשר n זה מספר הפיקסלים בתמונת הקלט. כל חלון בגודל $k \times s$ ניתן לרשת במקביל. רשת BERT תלמד את הקשרים והקורלציה בין כל זוג פיקסלים בכל אחד מהחלונות, וכך הרשת תלמד על קורלציה מקומית בכל חלון של התמונה, ותצליח להשיג התנהגות דומה לזו שמשיגה שכבת קונבולוציה.

   כדי שרשת דומה ל־BERT תוכל לחקות התנהגות של שכבת FC (או רשת MLP שלמה), ניתן לרשת את הקלט בשלמותו. הרשת תלמד את הקשרים והקורלציה בין כל זוג פיצ'רים (למשל פיקסלים בתמונות), וזאת בדומה להתנהגות של שכבת FC.

2.

(א) **רשת GAN רגילה:** נשים לב כי, גם אם יש לנו זוגות תואמים של דגימות ממחלקות A ו־B, ברשת זו אין קשר ישיר בין דגימה ממחלקה A לדגימה המתאימה לה במחלקה B, מכיוון שהגנרטור מקבל רק דגימה של מחלקה A, ללא דגימה תואמת ממחלקה B, ואמורה לייצר תמונה כלשהי ממחלקה B שאינה דומה בהכרח לדגימה התואמת. בנוסף הדיסקרימנטור מקבל בכל פעם דגימה ממחלקה B, או פלט של הגנרטור ואמור להפריד ביניהם (וגם כאן לא מתבצע קישור לדגימה התואמת ממחלקה A). לכן אסטרטגיה זו לא תרוויח משדרוג ה־dataset המתואר ב־iii ל־ii (וכמובן שגם לא מהשדרוג ל־i). לכן אנחנו לא מצפים שהרשת תצליח על אף אחד משלושת ה־datasets, שכן, עבור דגימה מ־A היא יכולה לצייר דגימה מ־B שלא תואמת כלל לדגימה מ־A, וייתכן אף שהרשת תכנס למצב של mode collapse ותייצר את אותה דגימה מ־B לכל קלט דגימה מ־A.

(ב) **רשת GAN מעגלית:** עבור dataset מספר iii, כאשר אין לנו זוגות תואמים מ־A ו־B, אנחנו מצפים שיהיו תוצאות לא טובות כי בדומה לרשת GAN רגילה, אין דרך לכוון לדגימה הרצויה מ־B עבור דגימה מ־A. כן ראוי לציין שייתכן בכל זאת שהתוצאות יהיו מעט יותר טובות לעומת רשת GAN רגילה עבור iii, כי לדעתנו הסיכוי שהרשת תכנס ל־mode collapse במקרה זה נמוכה יותר, בגלל ה־loss שדורש שחזור של המקור, אך ייתכן שהדגימה מ־B שממנה הגנרטור השני משחזר את הדגימה המקורית מ־A לא תהיה תואמת לה. אסטרטגיה זו תרוויח לדעתנו מהשדרוג ל־dataset מספר ii, כי ברגע שיש לנו זוגות תואמים, נוכל בכל פעם לשלוח דגימה מ־A לגנרטור הראשון ואת הדגימה התואמת לה מ־B נשלח לגנרטור השני, וכך, על־ידי מזעור ה־loss, נגרום לגרנטור הראשון לייצר לכל דגימה ב־A דגימה תואמת מ־B. לדעתנו אסטרטגיה זו לא תרוויח בצורה משמעותית משדרוג

מ־dataset מספר ii ל־dataset למספר i, וזאת מכיוון שהקשר בין זוגות תואמים לא מתבטא בצורה ישירה במהלך האימון, אלא רק בצורה עקיפה דרך ה־loss. כלומר, עבור זוג תואם של דגימות מ־A ו־B, הגנרטור הראשון יקבל את הדגימה מ־A וייצר דגימה כלשהי מ־B, ללא קשר לדגימה התואמת, וכנ"ל באופן הפוך על הגנרטור השני שיקבל את הדגימה מ־B.

(ג) **רשת GAN מותנית:** עבור dataset מספר iii, כאשר אין לנו זוגות תואמים מ־A ו־B, אנחנו מצפים שיהיו תוצאות לא טובות, כי בדומה לרשת GAN רגילה, אין דרך לכוון לדגימה הרצויה מ־B עבור דגימה מ־A. אסטרטגיה זו תרוויח לדעתנו מהשדרוג ל־dataset מספר ii, ותביא לתוצאות טובות יותר, מכיוון שברגע שיש לנו זוגות תואמים, הדיסקרימנטור בתהליך האימון יוכל לקבל זוגות תואמים, ובכך לגרום לגנרטור לייצר עבור דגימה מ־A דגימה תומאמת מ־B כדי שהוא יוכל לבלבל את הדיסקרימנטור. לדעתנו אסטרטגיה זו תרוויח גם משדרוג ל־dataset מספר i ותביא לתוצאות טובות אף יותר משימוש ב־ii, כי כדי שהגנרטור יצליח לבלבל את הדיסקרימנטור, הגנרטור יצטרך לייצר עבור דגימה מ־A, דגימה מ־B שהיא יותר מפורטת ומדוייקת, וכך על דגימות חדשות מ־A (שלא בהכרח יש להן דגימה תואמת מ־B), הנגטור יצליח לייצר דגימה מ־B שתואמת ל־A ברמת פירוט גבוהה.