

שאלה 1 – VAE

תחילה יצרתי מחלקה היורשת Mnist כן שאצור דאטה סט צבוע. בעזרת הגרלת צבעים וצביעת הספרה בתמונה. ולסיכום המודל הבדיד עובד הרבה יותר טוב מהרציף והמשולב, מכיוון שדרש הרבה פחות אפוקים וגנרט צבעים יותר בוהקים, וניתן היה לשלוט עליו. יתרון של הרציף שניתן להציג את המרחב הלטנטי, ולשלוט על הצבעים שלו באופן מובהק וברור. בנוסף על המודל המשולב לא היה ניתן לשלוט על המרחב הלטנטי שלו, אך הוא הציג תוצאות מגוונות וצבעוניות מאוד. ניסיונות השייפור היו בעזרת יותר שכבות קונבולוציה והסרת המקס פולינג שלא היה נחוץ וגרע. ובנוסף מספר אפוקים מתאים וניסויים על הייפר פרמטרים שונים.

מודל רציף:

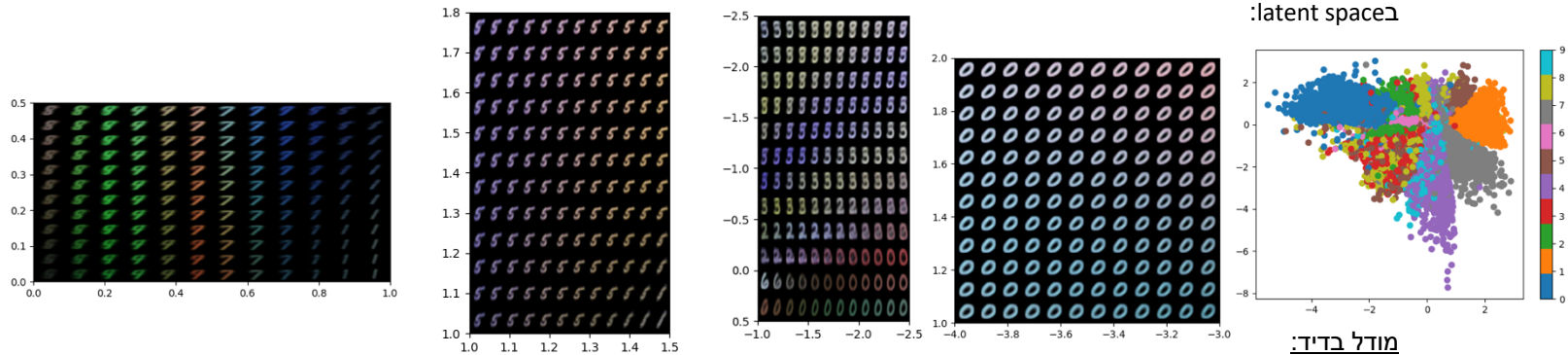
מודל זה מבוסס על התרגול בו ישנו encoder and decoder וביניהם ישנו משתנה מקרי נורמלי אשר הפרמטרים שלו נלמדים. **ארכיטקטורה – encoder:** שכבה לינארית מגודל (512, 28*3) כאשר הקלט הוא 28*28*3 מכיוון שהמרנו את התמונה לפורמט rgb. לאחר מכן אקטיבצית relu, ואז עוד שכבה לינארית מגודל (256, 512) לאחר מכן עוד אקטיבצית relu. ואז עוד שכבה לינארית מגודל (256, 3*2) לאחר מכן עוד אקטיבצית relu. מימד ה-3 בשביל 3 מימד הוקטור הנורמלי 2i כך ש החלק הראשון יהיה הפרמטרים של התוחלת והחלק השני הינו עבור לוג השונות.

לאחר מכן בדומה למה שעשינו בתרגול אנו עושים את reparameterization trick כדי ללמוד את הפרמטרים של ההתפלגות ולהגריל וקטור מהתפלגות נורמלית עם הפרמטרים הנלמדים ממימד 3 (latent dim).

ארכיטקטורה – decoder: מקבל את הוקטור הנורמלי שהוגרל מגודל 3, ונכנס לשכבה לינארית מגודל (256, 3), לאחר מכן אקטיבצית relu, ואז עוד שכבה לינארית מגודל (512, 256) לאחר מכן עוד אקטיבצית relu. ואז עוד שכבה לינארית מגודל (28*3, 512) כדי להעביר את הוקטור למימד של תמונה, אקטיבצית sigmoid כדי להגיע לערכים אי שליליים עבור פורמט התמונה (ובהצגת התמונה נכפול ב-255). ואז המרת הוקטור לתמונה.

תהליך האימון – 20 אפוקים, Adam optimizer with learning rate of 0.001, batch size = 128, BCE loss.

תרשימים – תחילה הצגתי את latent space של המספרים והצבעים, ודגמתי מכל קלסטר את ערכי ההתפלגות, וככה שלטתי על צבעים שונים. להלן התמונה הימנית היא latent space ושאר התמונות הן דגימות שונות לפי צבעים שונים latent space:



מודל בדיד:

מודל זה מבוסס על התרגול בו ישנו encoder and decoder וביניהם ישנו משתנה מקרי מסוג gumble-softmax אשר הפרמטרים שלו נלמדים.

ארכיטקטורה – encoder: שכבה לינארית מגודל (512, 28*3) כאשר הקלט הוא 28*28*3 מכיוון שהמרנו את התמונה לפורמט rgb. לאחר מכן אקטיבצית relu, ואז עוד שכבה לינארית מגודל (256, 512) לאחר מכן עוד אקטיבצית relu. ואז עוד שכבה לינארית מגודל (256, 3*20) לאחר מכן עוד אקטיבצית relu. מימד ה-20 היא בשביל 3 מימד הוקטור המוגרל ו-20 עבור מספר הקטגוריות של המשתנה המקרי gumble.

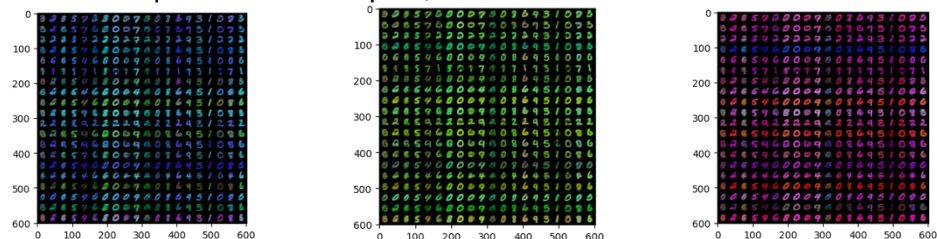
לאחר מכן בדומה למה שעשינו בתרגול אנו עושים את gumble-softmax כדי ללמוד את הפרמטרים של ההתפלגות ולהגריל וקטור מהתפלגות זו עם הפרמטרים הנלמדים ממימד 20*3 (latent dim).

ארכיטקטורה – decoder: מקבל את הוקטור שהוגרל מגודל 20*3, ונכנס לשכבה לינארית מגודל (256, 20*3), לאחר מכן אקטיבצית relu, ואז עוד שכבה לינארית מגודל (512, 256) לאחר מכן עוד אקטיבצית relu. ואז עוד שכבה לינארית מגודל (28*3, 512) כדי להעביר את הוקטור למימד של תמונה, אקטיבצית sigmoid כדי להגיע לערכים אי שליליים עבור פורמט התמונה (ובהצגת התמונה נכפול ב-255). ואז המרת הוקטור לתמונה.

תהליך האימון – 30 אפוקים, Adam optimizer with learning rate of 0.001, batch size = 128, ופונקציית לוס של BCE.

המחובר kl divergence כפי שראינו בתרגול המתאים למודל הבדיד.

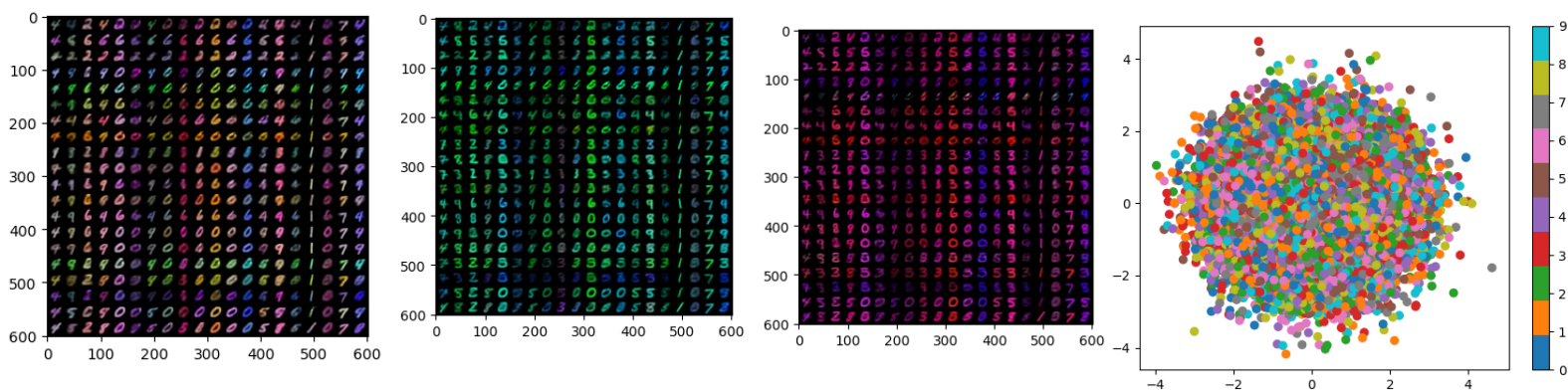
תרשימים – תחילה הצגתי את latent space של המספרים והצבעים, ודגמתי מכל שילוב של 3 קטגוריות, והצגתי את כל האפשרויות, שמתי לב כי עבור שילוב קטגוריות ספציפיות decoder מצליח להוציא צבע אחיד. לכן בעזרת שילובים אלו אני יכול לשלוט על צבע התמונה, להלן מספר שילובים של קטגוריות:



מודל משולב:

מודל זה מבוסס על שני המודלים שבנינו, הרציף והבדיד. כך שאנו מקבלים תמונה, שולחים אותה לencoder של כל אחד מהמודלים ומהם מגרילים שני משתנים מקריים, לפי encoders. אחד נורמלי ואחד gumbel, משרשרים אותם ולוקטור אחד, ואז שולחים לdecoder, עם ארכיטקטורה דומה לשל המודלים הקודמים מלבד השכבה הלינארית הראשונה שהיא מגודל של (256, 3*3), בעקבות הוקטור המשווה שהוגדרל משתי ההתפלגויות.

תהליך האימון – 30 אפוקים, Adam optimizer with learning rate of 0.001, batch size = 128. ופונקציית לוס של BCE המחובר לkl divergence כפי שראינו בתרגול המתאים למודל הבדיד, כדי שהמודל יתאים את עצמו לשני סוגי ההתפלגויות. **תרשימים** – בדומה למודלים הקודמים, במקרה הרציף הצגתי את המרחב הלטנטי אך לצערי הוא היה מאוד מבולגן, ולא אפשרי לדגום ממנו כדי לשלוט על הצבעים, לכן בחרתי להגריל מהתפלגות נורמלית עם פרמטרים של תוחלת 0 ושונות 5 בדומה למרחב הלטנטי. במקרה הבדיד דגמתי בצורה דומה, ולכן היה קשה לשלוט על הצבעים אך ניסיתי רק מתוך הבדיד כי על הרציף אין שליטה. להלן המרחב הלטנטי של הרציף ודגימות של צבעים שונים(אציין כי רב התמונות היו נורא צבעוניות ומעורבות בהמון צבעים שונים):



שאלה 2 – cycle GAN

תחילה בחרתי 10 צבעים שונים, ויצרתי שתי מחלקות היורשות מ-MNIST, כך שייצרו לי דאטה סט צבוע, כך שלכל מספר צבע קבוע. במדגם אחד לספרות היו צבעים שקבעתי, ובמדגם השני לספרות היו צבעים של ספרות שונות כנדרש. יצרתי 2 דיסקרימינטורים לכל מדגם(לשניהם אותה ארכיטקטורה), 21 גנרטורים(לשניהם אותה ארכיטקטורה), המעברים ממדגם אחד לשני ולהפך. כלומר צובעים את הספרות לפי המדגם האחר. ניסיונות השיפור היו משחקים עם היפר פרמטרים, בחירת 10 צבעים שמאוד שונים מתמטית אחד מהשני שמאוד השפיעו על הרשת, פונקציות לוס שונות שהתאימו, סדר האימון של כל רשת וכמובן מספר אפוקים שונים. **ארכיטקטורה** – discriminators: 2 שכבות קונבולוציה המגדילות את התמונה, לאחר מכן maxpooling מגודל 2, ולאחר מכן רשת fc לסיווג עם שתי שכבות לינאריות שבין כל אחד פונקציית אקטיבציה של relu, ולבסוף סיגמואיד לסיווג הסתברותי.

ארכיטקטורה – generators: 3 שכבות קונבולוציה המגדילות ומקווצות את התמונה(כדי שהרשת תלמד את האינפורמציה שהיא צריכה להזריק לתמונה, ולא לאבד אינפורמציה מהתמונה המקורית), שבין כל שכבה יש פונקציית אקטיבציה של relu.

תהליך האימון – 150 אפוקים, adam optimizer with lr of 0.0003 לכל רשת, batch size = 128, BCE loss. בכל אפוק אני מאמן תחילה את הדיסקרימינטור(נועל את הגנריטור) של המדגם הראשון בכך שאני מכניס לו תמונות מהמדגם ומסווג כמקוריות. לאחר מכן מגנרט מהמדגם השני לראשון ואז מכניס לו ומסווג כמזויפות. לאחר מכן אני מאמן את הגנריטור שמעביר המדגם השני לראשון(נועל את הדיסקרימינטור) בכך שאני מגנרט תמונות מהמדגם השני לראשון ונותן לדיסקרימינטור לסווג, ומכניס ומגדיר את הלוס כסיווג כמקוריות.

לאחר מכן בדומה אני מאמן את הדיסקרימינטור של המדגם השני ואת הגנריטור המעביר מהמדגם הראשון לשני. **תרשימים** – להלן שתי תמונות מסט המבחן כך שבתמונה הראשונה העמודות הזוגיות הינן מהמדגם הראשון והעמודות האי זוגיות הינן בהתאם הגינרט(הצביעה לצבע לאחר) למדגם השני של אותה ספרה. בדומה התמונה השנייה מתארת את המעבר מהמדגם השני לראשון.

