

## "דימות נתונים" – עבודת גמר תשפ"ב

### הקדמה:

במסגרת הקורס, מצ"ב פרטים על עבודת הגמר אשר יש להגיש. מטרת העבודה היא להכיר לסטודנט בצורה אישית ובלתי אמצעית את מגוון הנושאים אשר נלמדו בקורס. בין היתר, יש להבין את הנושאים שעברנו עליהם בקורס הקודם בצורה משמעותית וטובה.

במסגרת הקורס, מצ"ב פרטים על עבודת הגמר אשר יש להגיש. מטרת העבודה היא להכיר לסטודנט בצורה אישית ובלתי אמצעית את מגוון הנושאים אשר נלמדו בקורס. בין היתר, יש להבין את הנושאים שעברנו עליהם

1. עבודה מתקדמת עם pandas
2. למידת ensemble
3. למידת unsupervised
4. עבודה עם PCA והורדת מימדים

### עם אילו dataset נעבוד?

הראשון הינו fashion-mnist המכיל כ-70,000 תמונות של הספרות 0-9 ברזולוציה של 28\*28 או 784 פיקסלים בצבע אחד. הדאטא השני נקרא cats-vs-dogs והוא מכיל כ-25,000 תמונות של כלבים וחתולים בגדלים שונים.

1. השתמשו בכל מה שלמדתם על מנת לייצר מודל המסווג טוב את fmnist. חשוב מאד שלא לגעת ב-10,000 דוגמאות של test. ניתן להשתמש בכל הטכניקות שלמדנו (וגם במה שלא). אסור להשתמש ברשתות נוירונים.
2. חזרו על סעיף 1 אבל הפעם עם הדאטא של dogs-vs-cats. שימו לב שהפעם מדובר בתמונות בעלות גודל משתנה ובצבעים. ניתן להשתמש בספריית matplotlib או opencv כדי לסדר את הגודל (חפשו את פונקציית resize) וגם להפוך את התמונות לשחור לבן או לסקאלה אחרת.

3. חזרו על הדאטא מסמסטר א ובדקו איך אתם יכולים לשפר אותו באמצעות השיטות והאלגוריתמים שלמדנו במהלך הסמסטר. אילו תובנות אתם יכולים לקבל רק דרך ויזואליזציה שלו.

## עבודת pandas.

בעבודה זו תעבדו עם דאטא של לוגים של ציורי תמונות. ניתן לראות דוגמה של הדאטא בציור הבא:

X	Y	Pressure	TiltX	TiltY	Time
520.616237440813	532.149208915579	0.352285027503967	2296	2204	00:03.4235438
517.825153019979	529.726758286176	0.374013870954514	2296	2204	00:03.4382849
518.088462871001	529.568772375563	0.332753479480743	2296	2204	00:03.4382849
518.615082573045	529.568772375563	0.329579621553421	2296	2204	00:03.4532941
520.194941679178	529.568772375563	0.320057988166809	2296	2204	00:03.4532941
520.721561381222	529.463448435154	0.320057988166809	2296	2204	00:03.4752784
524.197251414713	528.515532971475	0.323720157146454	2296	2204	00:03.4752784
524.934518997575	528.304885090657	0.326649874448776	2296	2204	00:03.4908033
527.356969626978	527.883589329022	0.335439085960388	2296	2204	00:03.4908033
528.252223120453	527.830927358817	0.338612943887711	2296	2204	00:03.5084337
531.411941332717	527.778265388613	0.344472408294678	2296	2204	00:03.5191494
533.676406051507	527.988913269431	0.346913874149323	2296	2204	00:03.5246387
536.151518651114	528.094237209839	0.348867028951645	2296	2204	00:03.5356515
541.891673403395	529.042152673519	0.355703055858612	2296	2204	00:03.5429720
543.155560688301	529.252800554337	0.358632802963257	2296	2204	00:03.5585786
546.578588751588	530.200716018016	0.364492267370224	2296	2204	00:03.5727545
550.949532278554	531.2012934519	0.372304886579514	2296	2204	00:03.5878052
556.005081418178	531.991223004966	0.382803082466125	2296	2204	00:03.5948259

הדאטא מכיל את נקודות x-y, לחץ של העט, מנח העט וזמן.

הדאטא מתאר צורות שונות שציירו ילדים. ניתן לראות דוגמאות של ציורים בעמוד הבא:

מה נדרש (שימו לב - בנקודה זו אני יכול להוסיף כמה דרישות. לא לשנות אלא להוסיף)

1. לספור כמה קווים שונים יש בתמונה (ללא עבודה עם אלגוריתמים של עיבוד תמונה)
2. לחשב את האורך של כל קו ואת הלחץ הממוצע המופעל בלחיצה על כל קו.
3. לספור כמה הרמות יד היו בכל הציור. שימו לב שמדובר על משימה שונה מאשר משימה 1
4. לבדוק האם הצורה פתוחה או סגורה (בנוס - רק לגבי צורות מסוימות)

