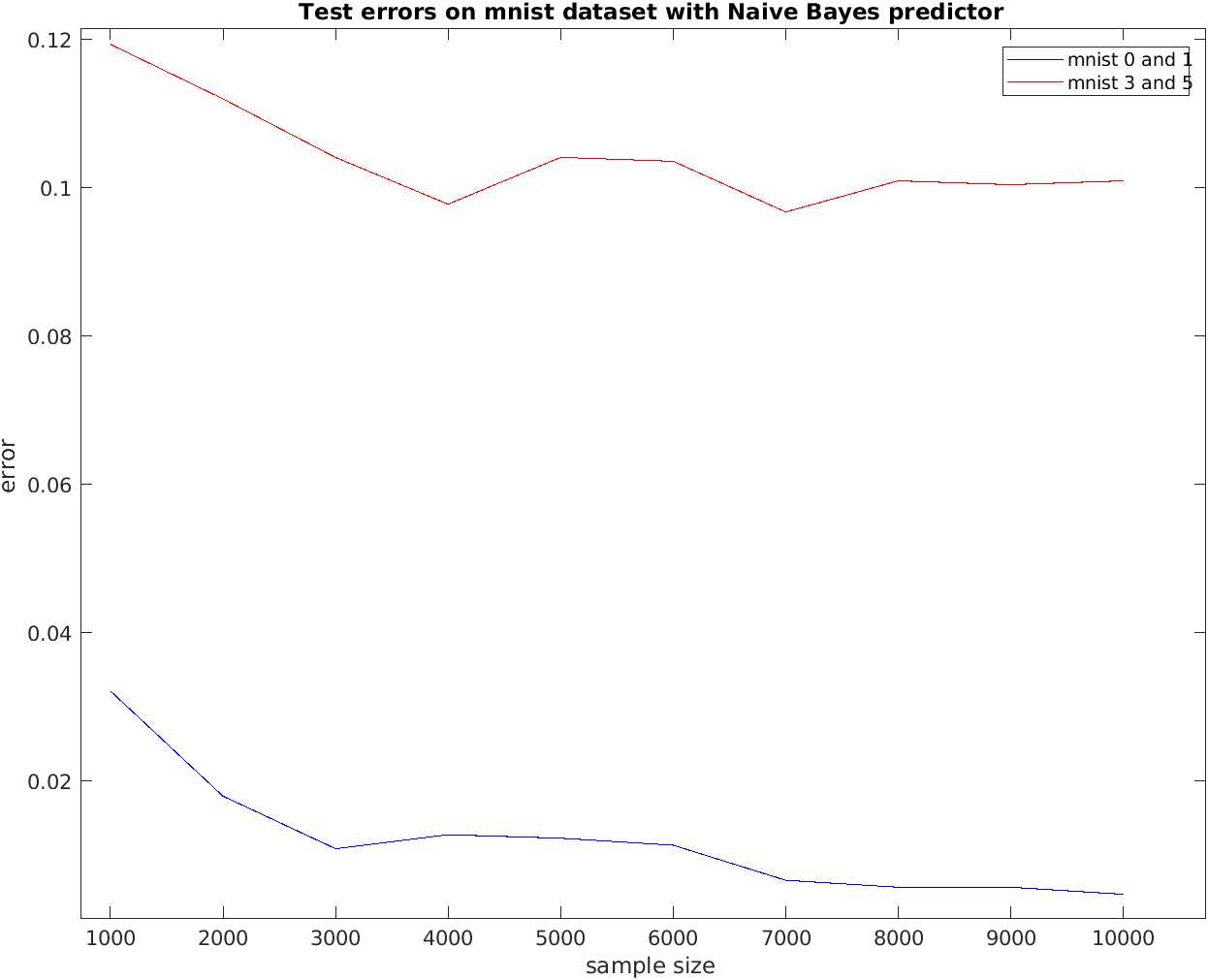
**עבודה 3**

**מבוא ללמידה וניתוח של מידע רב**

**מגישים: רעי וייס-ליפשיץ ועומרי אטל**

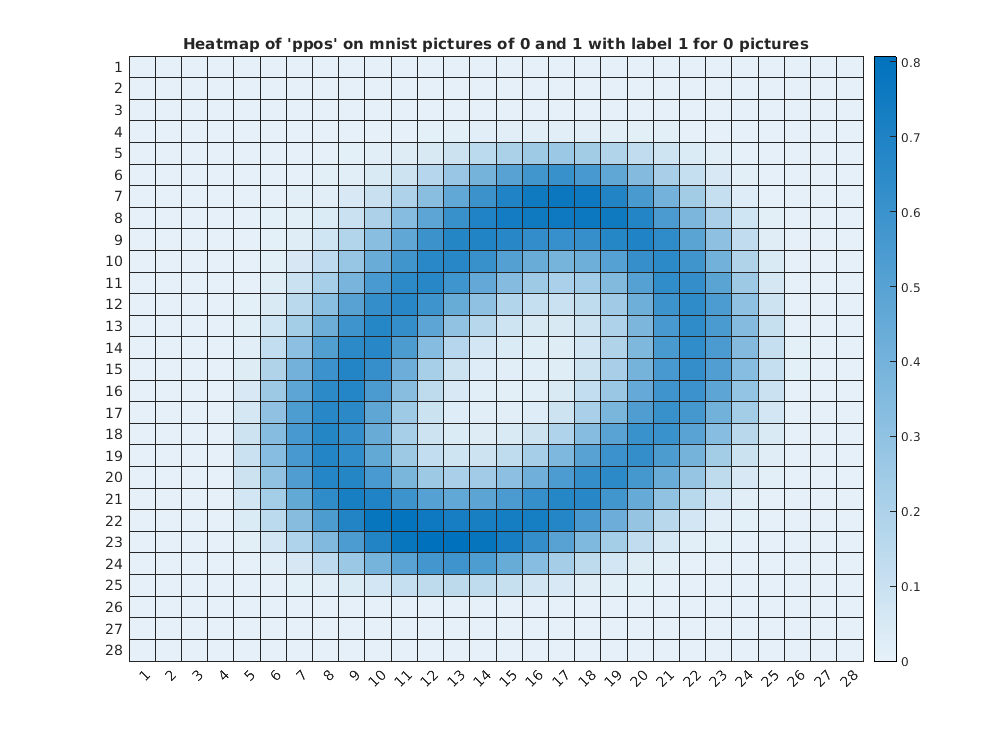
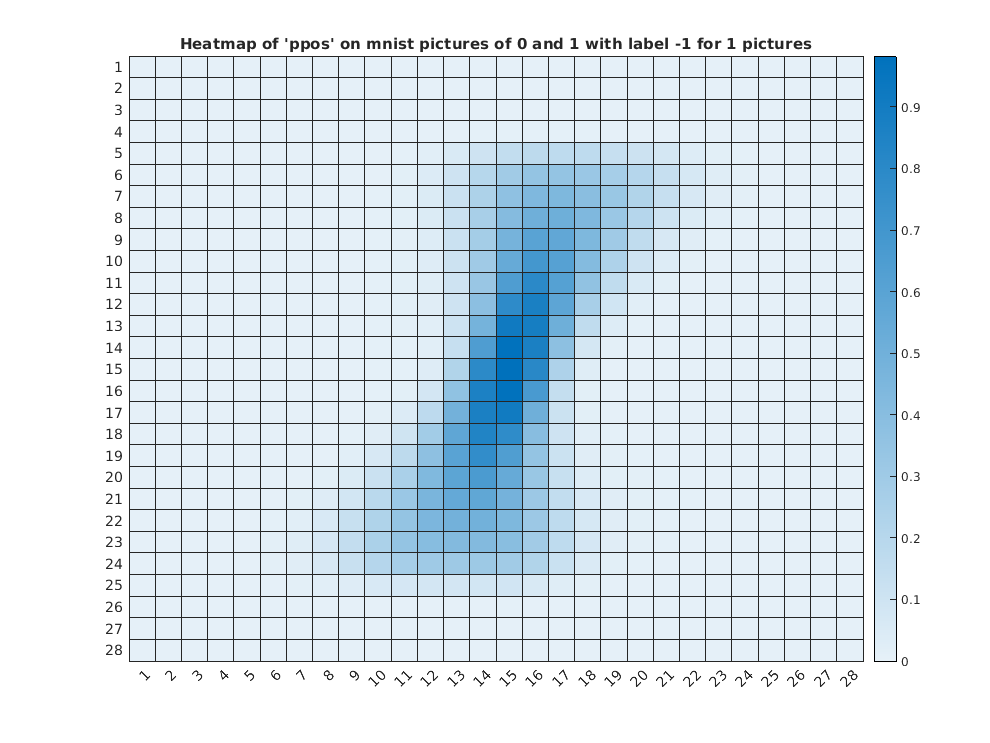
שאלה 2

סעיף א



סעיף ב

* עבור קלסיפיקציה בין הספרה 0 ו-1 קיבלנו תוצאות טובות מאוד אפילו ב-training size נמוך וקיבלנו ירידה יחסית משמעותית ככל שה-training size עלה.
* עבור קלסיפיקציה בין הספרה 3 ו-5 קיבלנו תוצאות פחות טובות ובנוסף קיבלנו ירידה פחות תלולה (ויותר רועשת) ביחס לבעיה הקודמת.

סעיף ג

* לפי הגדרה, , כאשר בעל קואורדינטות. במקרה שלנו, תמונות עם ספרה 0 הן בעלות תווית 1, ולכן זו ההסתברות לקבל בקואורדינטה 1 כאשר הספרה היא 0. בהינתן שהספרה שקיבלנו היא 0 אז הסיכוי הגבוה ביותר לקבל במשבצת כלשהי 1 נמצא במשבצות שיוצרות את הספרה.
* עבור משמעותו היא ההסתברות לקבל במשבצת 1 כאשר נתון לנו שהספרה היא 1 (אצלנו הספרה 1 בעלת תווית 1-). לכן באופן דומה, הסיכוי הגבוה ביותר לקבל 1 במשבצת כלשהי נמצא במשבצות שיוצרות את הספרה 1.

סעיף ד

עבור בדיקה על תמונות של ו- אין כלל שינוי, ועבור תמונות של 3 ו-5 השינוי הוא קל ביותר, אחוז מהמדגם השתנה מתווית ל-.

נשים לב כי דבר זה הגיוני, מכיוון שהחישוב להחלטת התווית נראה כך:

כאשר ההיפותזה שקובעת את התווי של .

נציין כי ערכו של שחושב בשני המקרים הוא כ-, וכל הוא בעל קואורדינטות. לכן השינוי בין ל- הוא זניח ביחס לסך הגודל הנסכם, אשר פרט ל- נשאר זהה בשני המקרים. לכן השפעת השינוי הזה היא מועטה עד לא קיימת, פרט למקרים בהם הפרדיקציה הייתה ברמת ביטחון נמוכה (). זאת הסיבה שאנו רואים את השינוי רק בפרדיקציה עבור ו-5 שם השגיאה הייתה יותר גבוהה, ולכן רמת הביטחון הייתה ככל הנראה נמוכה יותר.

שאלה 3

סעיף א

תיאור הגרף G:

* Input layer – מכיל d נוירונים, כאשר ערך כל נוירון קלט מתאים לקואורדינטה של .
* Output layer – עם נוירון output יחיד בעל פונקציית אקטיבציה sign. כל הנוירונים של   
  ה-Input layer מחוברים לנוירון של ה-Output layer.

נוכיח ש- . נראה זאת על ידי הכלה דו-כיוונית.

*: יהי . נוכיח שקיים עבורו לכל . נגדיר , כאשר הם המשקלים של הצלעות בין ה-Input layer ל-Output layer.  
יהי , לפי הגדרת , מתקיים ש- . לפי הגדרת הרשת מתקיים שה-output של ה-Output layer הינו  
.  
 תהא . באופן דומה נגדיר   
ומשום ש- אז .*

*סעיף ב*

*טענת עזר: לא קיים מפריד לינארי עבור שמחזיר 1 אם ורק אם לכל יש ב-x יותר קואורדינטות חיוביות משליליות.*

*הוכחה: יהי ונניח בשלילה שלכל אם ורק אם ב- יש יותר קואורדינטות חיוביות משליליות.*

* *מקרה א: אם w מכיל רק אפסים אז ברור שההנחה שגויה.*
* *מקרה ב: w מכיל רק קואורדינטה אחת שהיא לא 0, בה"כ אז ניקח  
   ו בסתירה להנחה.*
* *מקרה ג: w מכיל שתי קואורדינטות שהן לא 0, בה"כ .*

1. *אם נגדיר אזי ולכן בסתירה להנחה.*
2. *אחרת, לפחות אחד מ- ואז נגדיר ונקבל ש- ולכן .*

* *מקרה ד: מכיל 3 קואורדינטות שהן לא 0.*

*אם כל או שאחד בלבד חיובי אז באופן דומה למקרה ג2  
נגדיר .*

*אחרת, שניים בלבד מהקוארדינטות חיובית, בה"כ אז באופן דומה למקרה ג1   
נגדיר .*

*ולבסוף, אם כל אז נגדיר   
ואז ולכן בסתירה.*

*כעת נגדיר את מבנה רשת הנוירונים ונראה שקיימת היפותזה כנדרש.*

תיאור הגרף G:

* Input layer – מכיל 3 נוירונים , כאשר ערך כל נוירון קלט מתאים   
  לקואורדינטה של .
* Hidden layer – מכיל 3 נוירונים, נסמנם כך שכל נוירון קלט מחובר ל , ולכל יש פונקציית אקטיבציה .
* Output layer –מכיל נוירון output יחיד בעל פונקציית אקטיבציה sign. כל הנוירונים של   
  ה- Hidden layerמחוברים לנוירון של ה-Output layer.

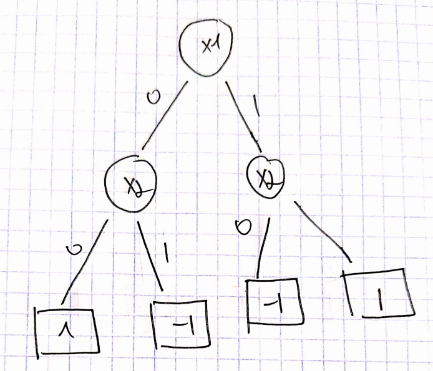
נראה שקיימת היפותזה כך שלכל אם ורק אם ב-x יש יותר קואורדינטות חיוביות משליליות. נבחר כאשר הוא וקטור המשקלים של נוירון , וקטור המשקלים של , – כנ"ל עבור ו- וקטור המשקלים של נוירון ה-output.

אז נשים לב ש- ומכאן , שזוהי בדיוק ההיפותזה הנדרשת.

שאלה 4

סעיף א

נראה שקיים עץ החלטה שמסווג במדויק את המדגם בעומק 2 ונראה שלא קיים עץ כנ"ל בעומק .

הסבר לנכונות העץ: לכל , אם אז יורד במורד תת העץ השמאלי, ושם העץ מחזיר 1 אם ורק אם , בדומה עבור ותת העץ הימני.

עבור עץ החלטה בעומק 1, ניתן לשאול רק עבור קואורדינטה אחת. אם השורש שואל לפי אז ברור שהעץ לא מסווג נכון את המדגם. אחרת, בה"כ השורש שואל עבור , מכיוון שבמדגם כל דוגמה מופיעה לפחות פעם אחת, אז שתי הדוגמאות נמצאות במדגם. אם התשובה של תת העץ המתאים ל- מחזירה 1, אז הוא טועה לגבי , ואם תשובתו היא 0 אז הוא טועה לגבי . סה"כ נקבל שעץ החלטה בעומק 1 לא יכול לסווג נכון את המדגם.

סעיף ב

ניקח מדגם S בגודל 8 כך שכל ערך אפשרי לדוגמה ב- נמצא במדגם פעם אחת בדיוק. נראה ריצה אפשרית של אלגוריתם ID3 על המדגם בעל פונקציית Gain   
עם עבור ופונקציית err שהוגדרו בהרצאה. *כלומר   
איטרציה ראשונה (עומק 0):*

נשים לב ש-

ולכן ונניח שהאלגוריתם בחר לפצל את השורש לפי ה-attribute ה-3.

איטרציה שניה (עומק 1):

* בענף השמאלי- .

משקילות הקואורדינטות ו- .

*מכיוון ש- עבור הענף השמאלי, נניח שהאלגוריתם בחר לפצל לפי .*

* *מגיעים לשלב 2 של האלגוריתם ומחזירים לייבל 1 או 0. בכל מקרה השגיאה היא בשני הענפים שפוצלו לפי ומגיעים לעומק 2 בתת העץ השמאלי.*

*איטרציה שלישית (עומק 1):*

* *בענף הימני- נקבל אותן התוצאות באופן זהה לענף . כלומר מגיעים לשגיאה של בכל אחד מהעלים.*

*סה"כ נקבל שבכל עלה יש לנו שגיאה של ולכן שגיאת עץ ההחלטה שנבנה ע"י אלגוריתם ID3 בריצה זו מגיע לשגיאה של .*

*שאלה 5*

*סעיף א*

*תחילה ניזכר בתכונה ידועה של מטריצות: לכל מטריצה ממשית A, . נתון שעבור בעיית הרגרסיה יש פתרון אופטימלי w יחיד שהוא כאשר הוא המטריצה שדוגמאות המדגם S פרוסות בשורותיה. אז בהכרח הפיכה, ומשום שממדיה דרגתה d. לפי התכונה הנ"ל של מטריצות ממשיות .*

*סעיף ב*

*נסמן ונראה שזוהי קבוצת הפתרונות האופטימליים עבור בעיית linear regression על כך שהווקטור הוא הפתרון לבעיית הרגרסיה על . כאשר הינו וקטור באורך ש- הקואורדינטות הראשונות שלו הן .   
אנחנו נראה שכל וקטור הינו פתרון אופטימלי לבעיית הרגרסיה על .*

*סימון: .*

*יהי , נראה ש- מביא למינימום את . נשים לב שמשום שהקואורדינטה ה - של היא 0, אז לכל כאשר . אזי מתקיים לכל ש- ולכן מביא למינימום את ומהווה פתרון אופטימלי לבעיית הרגרסיה על .*

*שאלה 6*

*סעיף א*

*נסמן את המטריצה כאשר . כלומר*

*למדנו בכיתה שה- distortion במקרה זה יהיה סכום שני הערכים העצמיים הקטנים ביותר של המטריצה A. אחרי שימוש בפונקציהnumpy.linalg.eig בפייתון על A נקבל שסכום זה הוא .*

*סעיף ב*

*בשביל לחשב את עלינו למצוא את שני הו"ע המתאימים לשני הע"ע הגדולים ביותר של A ולפרוס אותם בשורות ולנרמל אותם. שוב, נשתמש בפונקציהnumpy.linalg.eig בפייתון ונקבל ש-*

*כך ששורה 1 מתאימה לע"ע הגדול ביותר ושורה 2 מתאימה לע"ע .*

*סעיף ג*

*בשביל לקבל את הווקטורים לאחר ביצוע PCA בעזרת U, נבצע*

*ונקבל ש-*

*נשתמש בנוסחה ונקבל שה- distortion הוא .*

* *נזכור שגם בסעיף א קיבלנו distortion 21.27, כלומר קיבלנו תוצאות זהות. לפי מה שנלמד בכיתה, תוצאת הביטוי שקולה לסכום 2 הע"ע הקטנים ביותר של המטריצה A שהוגדרה למעלה. מכאן שהתוצאות חייבות להיות זהות.*

*שאלה 7*

*סעיף א*

*נטען שאלגוריתם single-linkage clustering לא מספק את אקסיומת ה-scale invariance. נראה דוגמה נגדית. ניקח ו- . עבור מטריקה אוקלידית ב- ונסמנה האלגוריתם יחבר בין האשכול ל- שכן מרחקם מינימלי, ואז נקבל אשכול ואשכול . לפי הגדרת  
single-linkage ולפי מטריקה אוקלידית ולא נחבר אותם שכן מרחקם גדול מ-7. נגדיר מטריקה חדשה . כעת עבור מטריקה זו, נקבל שבתחילת האלגוריתם, המרחק בין ל- הוא , וברור שהמרחק בין ל- ול- גדול מ-7. מתנאי העצירה של האלגוריתם לא נחבר אף אשכול ונקבל clustering שונה למרות ש-  
.*

*סעיף ב*

*נטען שאלגוריתם single-linkage clustering אכן מספק את אקסיומת richness. יהי clustering C לקבוצה S, נראה שקיימת מטריקה כך שהאלגוריתם יבחר את*

*נגדיר: . נראה תחילה ש- מקיימת את 3 האקסיומות. יהיו .*

1. *סימטריות – אם נמצאים באותו אשכול אז נקבל ש- ואם הם לא נמצאים באותו אשכול אז .*
2. *פוזיטיביות – ברור שכן .*
3. *אי שוויון המשולש: יהיו .*
4. *אם אז ברור שכן 0 קטן או שווה לכל מספר חיובי.*
5. *אם נמצאים באותו אשכול כלשהו, אז . יהי , נשים לב שאם הוא נמצא באותו אשכול ביחד עם אז נקבל ש  
   , ובדומה אם לא נמצא באף אשכול (מרחקו מ- יהיה . אם לא נמצאים באותו אשכול אז מרחקם לפי הגדרת המטריקה. בנוסף, לכל בהכרח מתקיים שהוא לא נמצא איתם ביחד באותו אשכול (או בכלל נמצא באשכול), שכן אחרת שניהם היו ביחד באותו אשכול. לכן סה"כ נקבל שבמקרה זה .*

*כעת, לפי מה שנלמד בהרצאה, אם נגדיר גרף , כך ש- וגם  
 אז באלגוריתם single-linkage רכיבי הקשירות יהיו הקלאסטרים. נותר להראות שרכיבי הקשירות של הגרף הם בדיוק האשכולים . לפי הגדרת , לכל שני קודקודים ב- הנמצאים באותו אשכול יש קשת ב-E שכן מרחקם . בנוסף לכל שלא נמצאים באותו אשכול, כלומר , מתקיים ש- ולכן לפי הגדרת E לא קיימת קשת ביניהם. כך למעשה הוכחנו שהאשכולים הם רכיבי הקשירות בגרף G. כתוצאה מכך, ממה שנלמד בהרצאה, האשכולים שיבחרו הם באלגוריתם single-linkage.*