

**אוניברסיטת בן-גוריון בנגב**

הפקולטה למדעי הטבע

המחלקה למדעי המחשב

דו"ח מסכם

**נושאים בחזית מדעי המחשב**

***לסטודנטים מצטיינים***

**מגישים:**

**יונתן ששוני 205916265**

**יוסי כרמלי 204752406**

1. מבוא

**פרופ' משה זיפר העביר במהלך הקורס הרצאה שאחד מהנושאים בה היה למידת מכונה מונחית, וזהו הנושא בו בחרנו להתמקד. למידת מונחית** היא למידה שלומדת מניסיון, בהינתן dataset מתוייג יוצרים מודל שביכולתו לבצע פרדיקציה/הכללה. לאחר צפייה במספר הרצאות שמשה הפנה אותנו אליהן בחרנו להמתקד ולחקור מספר אלגוריתמי למידה שונים ולממש אותם. לאחר מכן בעזרת dataset בשם MNIST המכיל תמונות של ספרות ערכנו בדיקות לטיב המסווגים שקיבלנו מהאלגוריתמים השונים.

בנוסף, בתחילת הקורס הוסבר כי ניתן להרחיב את היקף הפרויקט על מנת לקבל תוספת לציון.

בחרנו להרחיב את היקף הפרויקט ולחקור גם את נושא מתחום הלמידה הלא מונחית, *.*

1. k-Nearest Neighbors

אלגוריתם השכנים הקרובים הוא אלגוריתם מאוד אינטואיטיבי, בהינתן מדגם של דוגמאות ודוגמא חדשה הוא יבחר את השכנים הדומים ביותר ל מתוך , ויחזיר את תווית הרוב מבין השכנים הקרובים.

באופן פורמלי: נקבל כקלט לאלגוריתם מדגם ושלם , ונחזיר כפלט את המסווג .

המסווג חוזה עבור כל *דוגמה במרחב הדוגמאות את התווית לפי הכלל הבא:*

כאשר הוא האינדקס במדגם של השכן ה הקרוב ביותר לדוגמא , ו- היא תוויות של שכן זה.

כאשר נרצה לעשות שימוש באלגוריתם השכנים הקרובים עולות שתי שאלות עיקריות:

* כיצד מוגדר המרחק בין שתי דוגמאות במרחב הדוגמאות?

אנחנו בחרנו לעבור עם dataset שמרחב הדוגמאות שלו כאמור הוא תמונות של ספרות, כלומר .

בהינתן מטריצה המייצגת תמונה נשטח אותה ונציג אותה כוקטור . כעת בהינתן שני וקטורים נגדיר את המרחק בניהם באמצעות הנורמה האוקלידית: *.*

* איזה ערך נבחר עבור הפרמטר ?

לשם כך פנינו לאלגוריתם **cross validation שיעזור לנו להעריך את הערך הטוב ביותר עבור מתוך קבוצת ערכים אפשריים***. האלגוריתם מחלק את המדגם לשני חלקים, בראשון הוא משתמש כמדגם כדי לבצע למידה. בשני הוא משתמש כדי להעריך את השגיאה של המסווג שנלמד, מדגם כזה נקרא מדגם ולידציה.*

*הקלט לאלגוריתם:*

* *מדגם מתוייג*
* *אלגוריתם למידת - במקרה שלנו*
* *קבוצת ערכים סופית עבור הפרמטר אותו נרצה לשערך – במקרה שלנו קבוצת ערכים עבור*
* *– במהלך האלגוריתם נחלק את המדגם ל חלקים שווים.*

*הפלט: מסווג של האלגוריתם הנתון לאחר ביצוע למידה עם הערך הפרמטר הטוב ביותר.*

*שלבי האלגוריתם:*

1. *חלק את המדגם ל חלקים שווים בגודלם .*
2. *לכל בצע:*
   1. *לכל* 
      1. *(מדגם הולידציה)*
      2. *(במקרה שלנו )*
      3. *(השגיאה היא אחוז הטעויות שביצע על הדוגמאות מ )*
3. *החזר את .*

להציג תוצאות

עבור מדגם בגודל XXX קבוצת ערכים עבור הפרמטר וחלוקה ל חלקים קיבלנו שהערך האופטימלי שהחזיר אלגוריתם cross validation הינו XXX

להסביר כמה ספרות לקחנו.. פה אפשר לקבל את כולן 0-9

1. עצי החלטה – ID3

*עץ החלטה הוא מודל שבא לתאר תהליך החלטה טבעי. קל להסביר כיצד המודל עובר גם לאנשים מחוץ לתחום מדעי המחשב. המסווג המתואר על ידי עץ החלטה מבצע חיזוי על ידי טיול משורש לעלה בעץ. כאשר בכל צומת פנימי בעץ מופיע מבחן עבור אחד מהתכונות של הדוגמא הנתונה, ההתקדמות לצד ימין או שמאל נעשית בהתאם לתוצאת המבחן בצומת הנוכחי. בעלי העץ נמצאות תוויות אפשריות ממרחב התוויות, וכאשר נגיע לסוף הטיול בעץ ונגיע לעלה נחזיר את התווית שהוא מייצג.*

*למען פשטות נניח כי מרחב הדוגמאות הוא ומרחב התוויות הוא .*

***יש לבצע התאמה של התמונות מתוך ה dataset ובעזרת טרנפורמציה להפוך אותם לוקטורים מעל .***

***צריך לבדוק אם זה בכלל אפשרי***

*ID3 הוא אלגוריתם חמדן, אשר בכל שלב בוחר את התכונה שתפריד את המדגם בצורה הטובה ביותר, בעזרת פונקציית .*

*קלט האלגוריתם: מדגם ותת קבוצה של פיצ'רים .*

*פלט האלגוריתם: עץ עבור עם פיצ'רים מתוך .*

*שלבי האלגוריתם:*

* *אם כל הדוגמאות ב מתוייגות באותה תווית , החזר עלה עם תווית זו.*
* *אם החזר עלה מתוייג עם תווית הרוב ב .*
* *(כאשר היא הקורדינטה ה בוקטור )*
* *החזר עץ בעל שורש עם תכונה , בן שמאלי המתקבל על ידי ,*

*ובן ימני המתקבל על ידי .*

*מהי הפונקציה ?*

*היא פונקציה בה אנו עושים שימוש כדי לבחור את הפיצ'ר עליו משתלם לנו לשאול בשלב הנוכחי, ובמילים אחרות את הפי'צר שמפריד את המדגם שלנו בצורה הטובה ביותר. נסמן ב- את השבר המתאר את אחוז הדוגמאות במדגם עם תווית , שאר הדוגמאות עם תווית. אם בשלב הנוכחי נעצור וניצור עלה נקבל שהשגיאה במסלול זה הינה*

. לפיכך נגדיר .

נניח כי נרצה לפצל את המסלול עם תכונה כלומר נשאל ונסמן:

* שבר של אחוז הדוגמאות המתויגות בצד השמאלי, כלומר עבור דוגמאות המקיימות .
* שבר של אחוז הדוגמאות המתויגות בצד הימני, כלומר עבור דוגמאות המקיימות .
* אחוז הדוגמאות שהלכו שמאלה, כלומר דוגמאות המקיימות .

כעת נשים לב כי השגיאה לאחר המבחן הינה: *.*

*לכן נגדיר .*

*קיימות פונקציות שונות, אנו נעשה שימוש בפונקציה שההבדל בינה לבין האמור לעיל הוא*

*.*

*לא ניתן לבצע למידה בצורה טובה כאשר נשקול להחזיר עץ החלטה מתוך קבוצה כל עצי ההחלטה האפשריים.*

*ניסיון כזה יגרום לתופעת , כלומר השגיאה על המדגם תהיה קטנה, אך השגיאה על דוגמאות חדשות תהיה גבוהה.*

*שני דרכים לפתור את הבעיה הן:*

* *הגבלת מספר הקודקודים בעץ – במקרה זה נצטרך דוגמאות במדגם שלנו.*

*ניתן לממש את הגבלה זאת על ידי סיום האלגוריתם כאשר הגענו למספר קודקודים מסויים, או לחלופין לגזום את העץ לאחר בנייתו. הרעיון בגזימה הוא הורדה של תתי עצים שלא פוגעים באופן משמעותי בשגיאה.*

נניח כי פונקצית המקרבת את השגיאה של עץ על , בהינתן עץ לכל קודקוד בעץ, החל מהשיכבה התחתונה, בדוק את 5 הערכים הבאים ובחר באופציה עם הערך המינימלי עבור : השאר את המצב הקיים, החלף את קודקוד j ב-עלה עם ערך 0/1, החלף את קודקוד j עם תת עץ שמאלי/ימני.

* *: יער הוא אוסף של כמה עצי החלטה, שמחזיר פרדיקציה לפי החלטת הרוב. כעת נשאלת השאלה כיצד נייצר מספר רב של עצים שונים? דרך אחת לעשות זאת היא לבצע למידה של כל עץ בעזרת מדגם שונה וקבוצת פיצ'רים מוגרלת.*

להציג תוצאות

1. Naïve bayes

נניח כי מרחב הדוגמאות שלנו מעל ונחשוב על דוגמה כעל סדרת המלצות של מומחים, שכל אחד מנחש את התווית הנכונה לדעתו מתוך .

ההנחה המתבצעת בשיטה זו היא ***אי תלות מותנת*** *בין המומחים, כלומר לכל שני מומחים כאשר אנחנו יודעים מהי התשובה הנכונה , אזי עצם הידיעה של הניחוש של מומחה לא עוזרת להבין את הניחוש של מומחה .*

*באופן פורמלי: .*

*נסמן את ההסתברות שמומחה צודק בניחוש שלו ב .*

*תחת הנחת האי תלות מותנת* המסווג האופטימלי *הוא: , כאשר .*

*האינטואיציה למקדמים הללו הוא כדלקמן:*

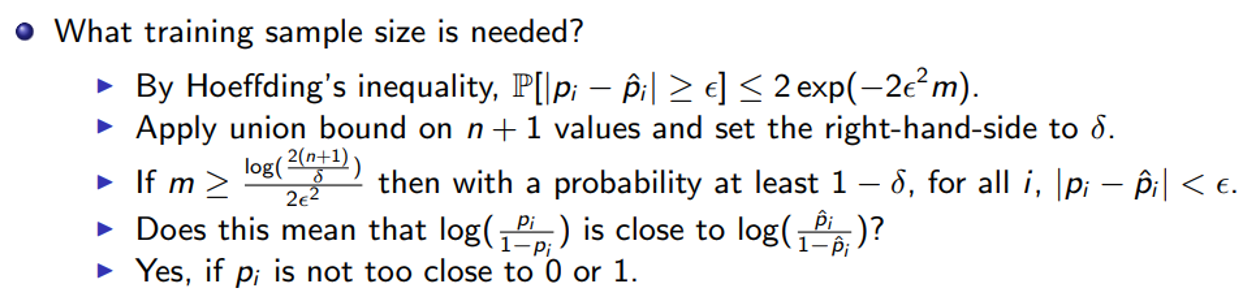
* *עבור מומחה שכמעט תמיד צודק () המשקל שהוא יקבל ישאף ל ולכן התשובה תהיה בהתאם להמלצתו .*
* *עבור מומחה שכמעט תמיד טועה () המשקל שהוא יקבל ישאף ל ולכן התשובה תהיה הפוכה להמלצתו.*
* *עברו מומחה שצודק ב (כלומר ) המשקל שהוא יקבל שווה ל*

*ולכן התשובה של המסווג לא תיקח בחשבון את המלצתו.*

כעת נשאלת השאלה כיצד נוכל למצוא את ערכם של עבור כל אחד מן המומחים?

הדרך לעשות זאת היא לקרב את ערכם באמצעות מדגם *.*

*נסמן ב את מספר הדוגמאות מתוך בהן צדק המומחה ה ונחשב .*



1. SVM

דרך נוספת לבצע למידה היא באמצעות מפרידים לינארים.

כיצד מפרידים לינארים מסווגים דוגמא חדשה? מפריד (היפר-מישור) כלשהו מחלק את המרחב לשני חלקים, אם הדוגמא החדשה ממוקמת מעל המפריד היא תקבל תוית מסויימת ואם היא מתחתיו היא תקבלת את התווית ההפוכה.

קיימים אלגוריתמים שונים המאפשרים ללמוד מפריד לינארי בצורה יעילה, כמו: תכנון לינארי, ו- *.*

*כל אלה מניחים כי המדגם הינו מדגם פריד, כלומר שאכן קיים מפריד לינארי שכל הדוגמאות עם תווית נמצאות בצד אחד של אותו מפריד, וכל הדוגמאות עם תוית נמצאות בצד השני של המפריד.*

*אנחנו פנינו לאלגוריתם אשר מוותר על הנחה זו, ומוצא מפריד לינארי גם עבור מדגם שאינו פריד.*

*תחילה נסביר את משמעות המושג של מפריד לינארי, ה הוא המרחק של המפריד מהנקודה הקרובה ביותר אליו מתוך הנקודות במדגם. ככל שה גדול יותר אפשר לחשוב שיש למפריד יותר מרווח לטעות בו, לכן נשאף למצוא מפריד עם גדול ככל הניתן.*

*כמו כן עבור מפריד לינארי ומדגם נגדיר פונקצית הפסד* **hinge loss** אותה נשאף למזער:

*, ועל המדגם: .*

*נשים לב שאפילו עבור מדגם פריד ו- מספיק גדול נקבל , ול- נורמה גודלה יותר משל w, אך*

*קטן יותר.*

*על מנת לאזן את הרצון מצד אחד למזער את פונקצית ההפסד ומצד שני להגדיל את ה נעשה שימוש ברגולטור : כאשר עבור גדול נעודד נורמה קטנה כלומר גדול, ועבור קטן נעודד הפסד קטן.*

*מצמצם את האובייקט הבא: .*

*את אלגוריתם ניתן לממש באמצעות תכנון ריבועי.*

*תכנון ריבועי הוא אלגוריתם הממזער את האובייקט תחת האילוצים .*

*נוכל להגדיר ולפתור את עבור הערכים הבאים של כקלט לתוכנית ריבועית:*

* *כאשר ישנם קורדינטות שערכן 0, ועוד קורדינטות שערכן .*
* *.*
* *כאשר ישנם קורדינטות שערכן 0, ואחריהן קורדינטות שערכן 1.*
* *.*
* *וקטור התוצאה שלנו יהיה , כאשר מבטאים את המפריד הלינארי שהתקבל מאלגוריתם .*

*כיצד בחרנו בעבודה של סיון את הרגולטור ?*

*להציג תוצאות*

1. למידת לא מונחית - Clustering(בונוס)

למידה לא מונחית היא למידה בה נקבל כקלט מדגם **לא** מתויג.

ניתן לבצע למידה לא מונחית כשלב מקדים ללמידה מונחית, על מנת לכווץ את הייצוג של הדוגמאות - .

כמו כן, ניתן לעשות בה שימוש על מנת לארגן את המידע שיש בידנו – .

אנחנו בחרנו להתמקד ב , שיטה זו מארגנת את המידע שיש בידנו, לצורך העניין את המדגם , לקבוצות שנקראות . דוגמה לשימוש היא חלוקה של מאגר תמונות לאלבומים שונים.

במקרה שלנו נרצה לחלק את התמונות של הספרות מ MNIST לעשר קבוצות, כך שכל קבוצה מייצגת ספרה שונה.

כדי לבצע את פעולת החלוקה לקבוצות נצטרך לעשות שימוש בפונקצית מרחק, אנו נעשה שימוש בפונקצית המרחק האוקלידי, כלומר .

האלגוריתם שלמדנו על מנת לבצע  *נקרא .*

*על מנת להציג אותו נסמן ב את הקלסטר ה- (כלומר קבוצה של דוגמאות), ונסמן ב את מרכז ה- (מרכז הוא איבר ממרחב הדוגמאות).*

*אלגוריתם הוא אלגוריתם איטרטיבי אשר עושה מעבר בין קלסטרים למרכזים ובכל שלב מעדכן את ערכיהם על מנת להשיג את המטרה.*

*האלגוריתם מתבסס על העבודה שניתן בקלות בעזרת מרכז לעבור להגדרה של קלסטר , ולהיפך.*

*קלט האלגוריתם: מדגם לא מתויג , ומספר הקלסטרים המבוקש .*

*פלט האלגוריתם: מרכזים (המגדירים קלסטרים).*

*שלבי האלגוריתם:*

* *איתחול רנדומלי של המרכזים .*
* *כל עוד האלגוריתם לא התכנס (כלומר כל עוד הרכבי הקלסטרים משתנה מסיבוב לסיבוב)* *בצע:*
* *לכל הגדר קלסטר .*
* *לכל הגדר את המרכז של הקלסטר להיות .*
* החזר .

*הצגת תוצאות*