**תרגיל בית 3 – מבוא לאינטליגנציה מלאכותית מס' קורס: 236501**

**מגישים: גלעד ורנר, יאיר גלסמן.**

**חלק ב':**

3.2 הסיבה שעלינו להקפיד לקרוא ל-split\_crosscheck\_groups רק פעם אחת היא בכדי שנוכל להשוות בין ביצועי מסווגים שונים. מכיוון שחלוקת הדוגמאות לקבוצות היא אקראית, אילו נבצע חלוקה זו לפני כל מסווג, התוצאות שנקבל יכולות להיות מושפעות מחלוקות הדוגמאות – ולא מצורת הבנייה של המסווגים.

5.2 תוצאות הרצה של המסווגים:

5.3 עבור K=1 קיבלנו את התוצאות הטובות ביותר.

5.4 ניתוח תוצאות:

ראשית נשים לב שאת התוצאה הטובה ביותר מספק המסווג עם k=1, ולאחר מכן יש ירידה בדיוק ככל שמספר השכנים עולה, כשנקודת המינימום היא עבור מסווג הלוקח בחשבון 13 שכנים. ההסבר שלנו לתופעה זו שכנראה הדאטה שאיתו אנחנו עובדים הוא יחסית לא רועש. כלומר ההפרדה בין הקבוצות (יש הפרעת קצב/ אין הפרעת קצב) יחסית ברורה, לכן ע"י הגדלת מספר השכנים שאנחנו לוקחים בחשבון, אנחנו גורמים לגבולות להיות פחות מובהקים. נגיד דגימה שעבור k=1 שכנה בשטח של קצב לב תקין, עבור k=13 הגבול משתנה וכעת היא מסווגת בטעות להפרעת קצב.

7.4 גם בהשוואה לשאר המסווגים עבור K=1 קיבלנו את התוצאות הטובות ביותר.

**חלק ג':**

נפרט כעת את תהליך המחקר שבסופו בחרנו במסווג KNN עם k=1.

אנחנו יצאנו מנק' הנחה, שעבור כל סוג דאטה קיים מסווג שמתאים לתכונותיו, ולכן בתור התחלה בחנו רשימת מסווגים שונים ומשונים עם טווח של משקלים.

המסווג שניצח שוב ושוב מבחינת אחוזי הדיוק היה 1NN. לכן החלטנו לבנות מסווג שיתבסס במידה על 1NN אך יכיל שיפורים שיגדילו את אחוזי הדיוק.

"מועצת מסווגים"

החלטנו לנסות גישה בה נאמן מספר מסווגים על הדוגמאות, ואז במהלך המבחן כל מסווג יתן את התחזית לגבי הסיווג של דוגמת המבחן – והסיווג הסופי יקבע ע"פ הרוב. בחנו הרכבים שונים של מסווגים, וככל שהוספנו יותר מסווגים אחוז הדיוק ירד. התוצאות היו טובות, אך עדיין לא התעלו על 1NN, ואחוז הדיוק נע בין 0.89 ל-0.94.

הסיבה לכך, היא שככל הנראה יש מס' דוגמאות שקשה למסווגים להצליח איתן, וברגע שמוסיפים מסווגים אז גם אם 1NN מצליח לחזות נכון, השיטה הדמוקרטית שלנו גורמת לכך שבסוף תבחר תחזית שגויה.

לאחר ניסיונות רבים הצלחנו להגיע לתוצאה טובה של 0.952 אחוזי דיוק (יותר טוב מ1NN), וזו גם המועצה שאנחנו שולחים לתחרות. קבוצת המסווגים מורכבת מ1NN, MLPו – AdaBoost אשר מימוש שלושתם נלקח מספריית sklearn.