מעבדה 4 – מבוא להנדסת נתונים

אליהו – ישראל עמר 329803506

שלום מעודה 206320772

שאלה 1:

קבלנו את הדיוק המקסימלי, כלומר 1 לאחר הרצת האלגוריתם.  
עובדה זו מוסברת היטב מאחר ועבור כל נקודה,החיזוי מתייחס אך לרק לנקודה הקרובה ביותר – כלומר לנקודה עצמה.

שאלה 2:

לפי הרצת האלגוריתם, המסווג המדויק ביותר הינו: 19NN

, בעזרתו מגיעים לרמת דיוק של 0.76~

שאלה 3:

ביצועים בעלי שונות נמוכה התקבלו עבור 2-Fold-Cross-Validation

במקרה שלנו, כאשר קיים לנו דאטה-סט מצומצם, עדיף לעבוד עם מספר מועט של folds

ע"מ שמקרי הקצה יתפזרו באופן מאוזן בין ה-folds ובכך נוכל להקטין את השונות בין ה-Folds  
אצלנו, נראה הגיוני לבחור מספר Folds שנע בין 2 ל-4.

שאלה 4:

נרמול לפי שיטת Min/Max הניב את התוצאות המדויקות ביותר.  
באופן כללי, ניתן לומר שעבור התרגיל שלנו, הנרמול שיפר את הקלסטרינג.