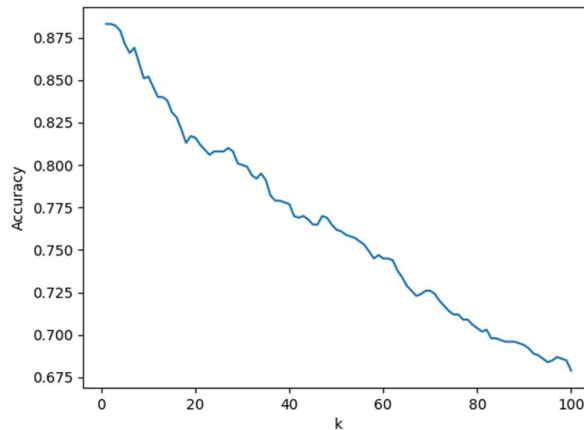


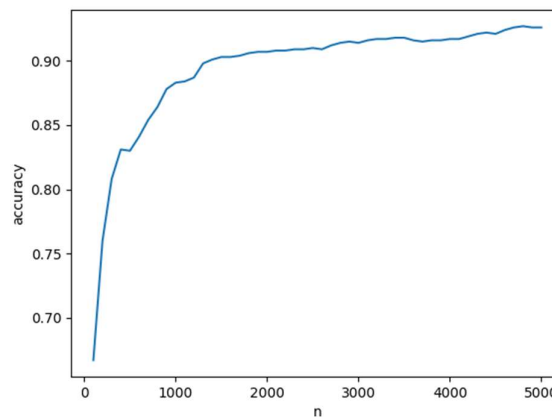
(b) הדיוק של האלגוריתם הינו 85.2% - `n=1000 k=10: percentage of correct labels is 0.852`

מאלגוריתם אקראי נצפה לקבל דיוק של $\frac{1}{\#labels}$.
במקרה שלנו ישנם 10 תיוגים ולכן נצפה לקבל דיוק של 10%.



(c)

נראה ש- $k=1$ מביא לדיוק הרב ביותר, קרי התמונה הדומה ביותר (לפי נורמה l2) ב-training data היא גם בעלת אותו התיוג כמו של ה-query image.
ככל ש- k גדל ישנן תמונות דומות בעלות תיוג שונה (שאוילי נפוצות יותר ב-training לדוג') והן מורידות את אחוז הדיוק. גם בהנחה שההתפלגות של תמונות מתיוגות ב-training data היא אחידה, אז כיוון שבחרנו לאמן על 1000 תמונות, לכל תיוג אמורות להיות כ-100 תמונות ומכאן הגיוני שבכל ש- k יתקרב ל-100 הדיוק ירד (ובוודאי כאשר נבחר $k > 100$).



(d)

בהתאם למה שראינו בכיתה, ניתן לראות שהגדלתו של ה-training data משפרת את הדיוק של האלגוריתם באופן לוגריתמי. הרחבת ה-training data מגדילה את ההסתברות למצוא תמונה דומה יותר ל-query image ובכך לתייג אותה כהלכה (נזכיר כי בבדיקה זו $k=1$). לאחר כ-1000-2000 תמונות ב-training data ה"מרחב" נעשה צפוף יותר, והסיכוי למצוא תמונה חדשה שדומה יותר מהתמונות שכבר קיימות ב"מרחב" הולך וקטן - מכאן נובע האופי הלוגריתמי של הגרף.