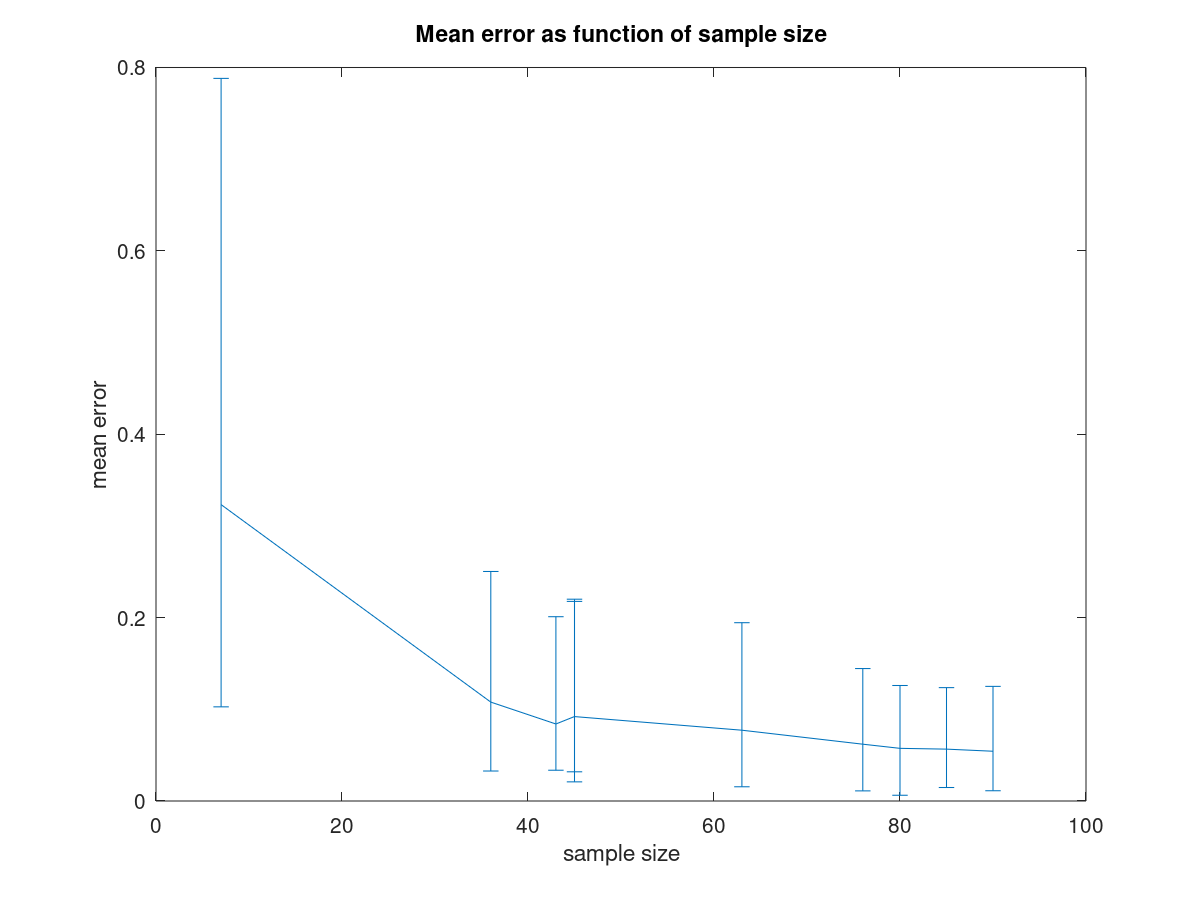
**עבודה 1 – מבוא ללמידה וניתוח של מידע רב**

**מגישים: עומרי אטל 208625103, רעי וייס-ליפשיץ 69696969**

שאלה 2

 סעיף א

סעיף ב

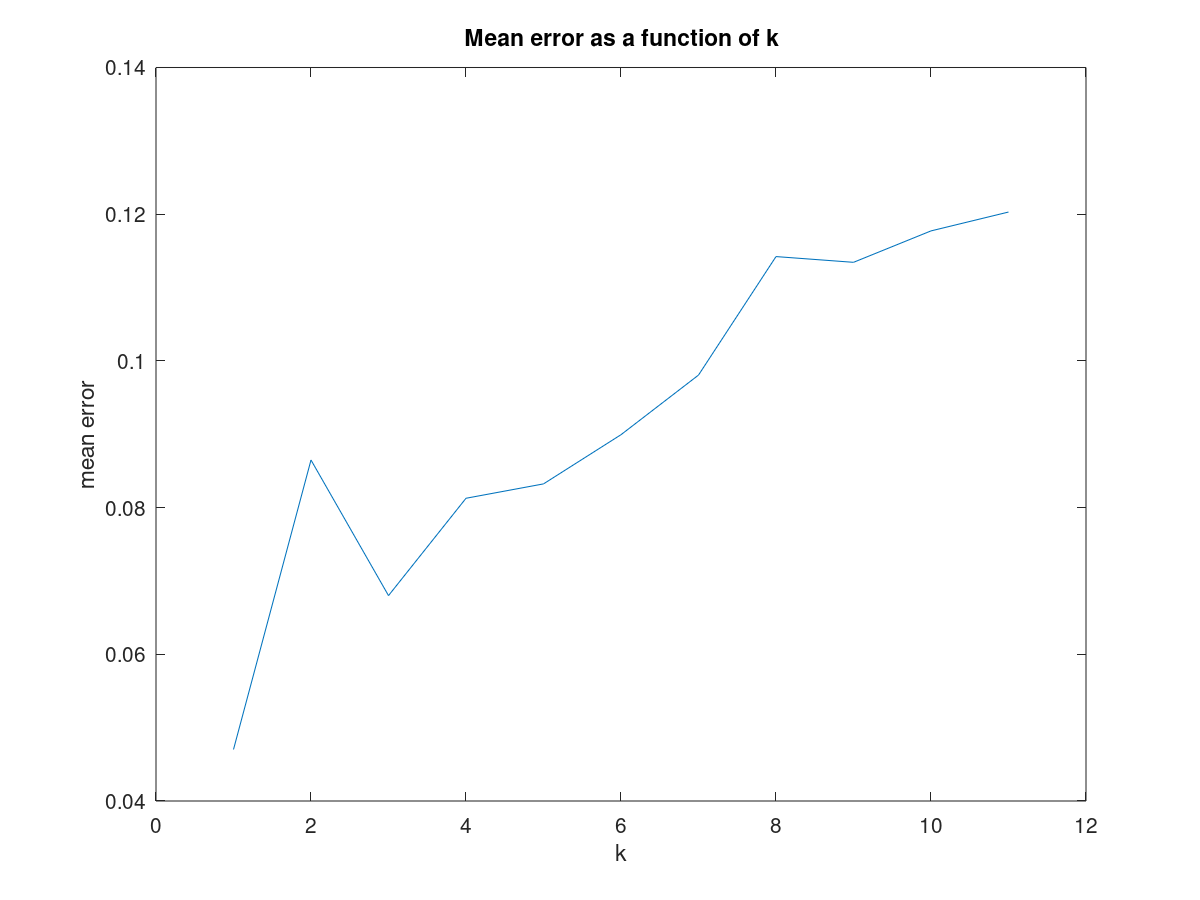
השגיאה הממוצעת יורדת ככל שה sample size גדל.  
הסבר: נשים לב תחילה שתמונות בעלות אותו לייבל יהיו דומות יחסית אחת לשנייה, כלומר המרחק האוקלידי שלהן יהיה קרוב ברוב המקרים. עבור k=1 אנו מריצים את למעשה את שיטת השכן הקרוב. ככל שהמדגם גדל, יש ל-predictor יותר דוגמאות מהלייבלים השונים, כך שלמעשה נוצרים גושים של נקודות בעלי אותו לייבל שגדלים יחד עם המדגם. מכאן, כאשר נפעיל את השכן הקרוב על כל דוגמה מה-test, יש סיכוי יותר גבוה שהדוגמה הזו "תיפול" בגוש הנכון (עם אותו הלייבל).

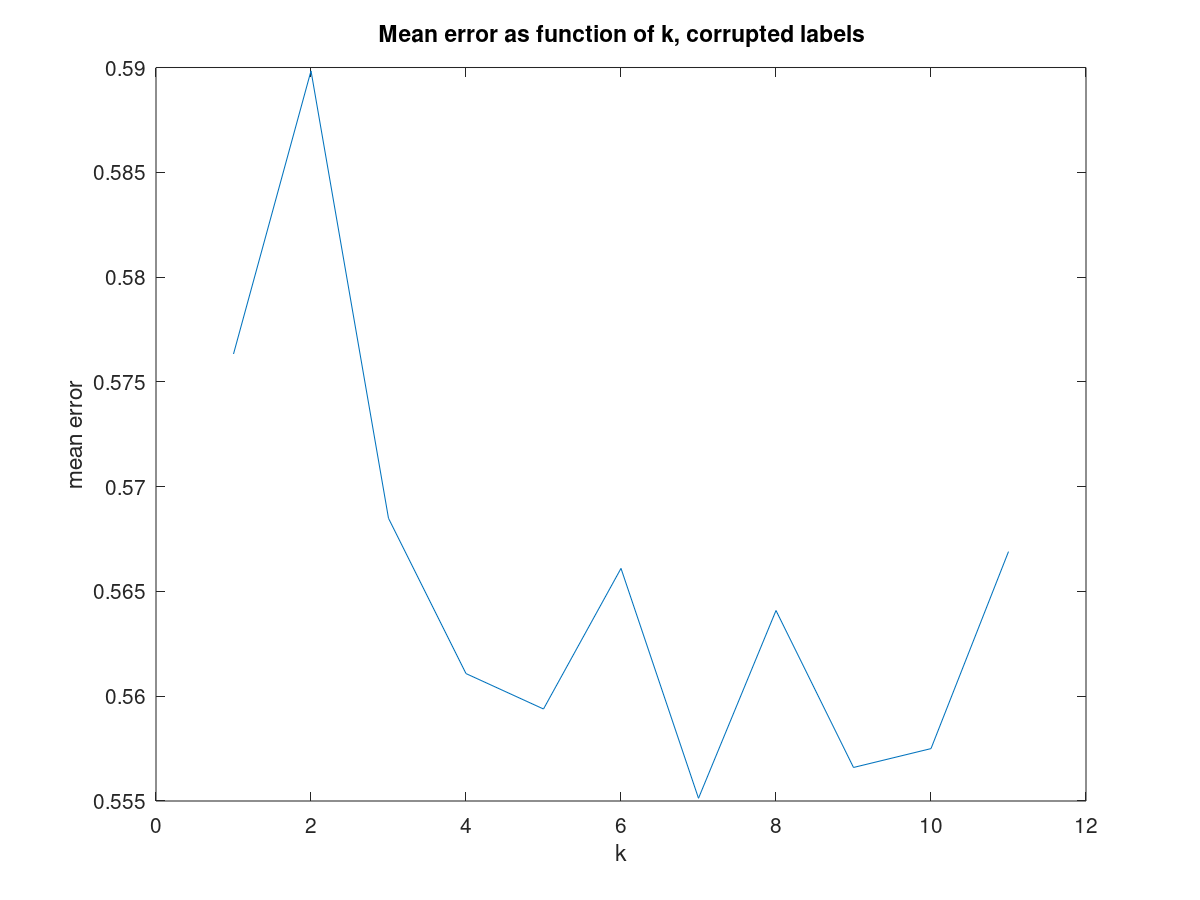
סעיף ג

כן, משום שב-gensmallm אנו יוצרים מדגם רנדומלי מה-train samples , כולכן בהרצות שונות יהיה לנו מדגמים שונים, מה שמשפיע על השגיאות.

סעיף ד

טענו בסעיף ב שככל שהמדגם גדל כך לרוב השגיאות קטנות, ולכן בפרט זה נכון עבור השגיאה המקסימלית והמינימלית. בנוסף, ניתן לראות שהשגיאה המינימלית יחסית קבועה וחסומה על ידי 0, ומשום שהשגיאה הממוצעת יורדת, השגיאה המקסימלית חייבת לקטון.

סעיף ה

סעיף ו

סעיף ז

שאלה 3

סעיף א

סעיף ב

נגדיר בה"כ ש- lettuce = 1 ו- carrot = 0.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  | 1 |
|  |  |
|  |  |

כלומר, .

סעיף ג

|  |  |  |
| --- | --- | --- |
| probability | preferred food | age |
| 0 | carrot | 7 |
| 60% | lettuce | 7 |
| 15% | carrot | 13 |
| 25% | lettuce | 13 |

סעיף ד

גם פה נגדיר בה"כ ש- lettuce = 1 ו- carrot = 0.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  | 1 |

כלומר,

סעיף ה

למדנו בהרצאה ש-

כאשר

בשאלה זו, . לכן:

עלינו לחשב את לכל .

*ולכן סה"כ נקבל:*

*לסיכום, תוחלת שגיאת אלגוריתם ה memorize כאשר הוא:*

שאלה 4

סעיף א

נסמן

*כך ש- , ולכן*

*סעיף ב*

*נשים לב ש: לפי כלל בייס, ולכן לכל מתקיים ש-*

*סעיף ג*

*נסמן*

*נמצא את כפונקציה של . נשים לב ש- כאשר . כלומר הינו סכום של משתנים מקריים שמתפלגים ברנולית עם הסתברות . לכן, , וכידוע . מכאן - .*

*סעיף ד*

*סעיף ה*

*תחילה נבחין ש-*

*סעיף ו*

*יהי ,*

*משום ש- אז מתקיים שלכל אכן .*

*אז נתבונן במקרה ש- . נגדיר . נשים לב ש- . לכן . אזי נקבל ש-*

*כנדרש.*

*שאלה 5*

*סעיף א*

*יהיו כך ש- וגם . משום ש- בהכרח מתקיים ש- התפלגות דטרמיניסטית, כלומר שלכל קיים יחיד בתמיכה של . לכן מתכונת דטרמיניסטיות של מתקיים שלכל , מקיימת ש- הוא 0 או 1. נשים לב שמשום ש- מתקיים , ולכן  
 . נתון ש- עבור היא פונקציה על פי המרחב האוקלידי ולכן בפרט עבור מתקיים , אז סה"כ לפי  
 נקבל .*

*סעיף ב*

*נסמן . צריך להוכיח ש- , כלומר צריך להראות שלכל , . יהיו כך ש- .  
נניח בשלילה ש- , כלומר השכן הקרוב ביותר של – x במדגם הוא בעל תווית , נסמנו . כלומר .*

*אבחנה: המרחק . לפי הנתון בו לכל קיים עבורו , ובפרט עבור , ולכן מכיוון ש- מכסה את כל נובע ש- , שכן קוטר הכדור הינו .*

*לפי האבחנה, , אבל ולכן, . לפי ההנחה, היא התווית של ו- היא התווית של , כך ש- . אך , בסתירה להיות פונקציה לפי סעיף א.*

*(\*) הבהרה: הטענה שהוכחנו בסעיף א מתבססת על כל מדגם , ולכן אם נוכל להוסיפו למדגם ובכך אילוץ המרחק על יתקיים.*