

A dark blue vertical bar on the left side of the page. A blue arrow points to the right from the bar, containing the date.

10/18/2020

# דוח סיכום פרוייקט

סיווג קולות שירה של שלדג לבן החזה

Several thin, curved lines in dark blue and light grey that sweep upwards from the bottom left corner of the page.

מגישים: אסף כפתורי ורון מלמד  
מנחה: פרופסור יזהר לבנר  
נכתב על ידי: רון מלמד

## תוכן עניינים

3	פרק 1: מבוא	➤
3	1.1 סווג ואשכול של אותות קול	•
3	1.2 רקע על השלדג לבן החזה	•
4	1.3 נושא המחקר	•
4	1.4 מטרות הפרויקט	•
4	1.5 מאפייני האות	•
5	1.6 העבודה שלנו	•
6	פרק 2 : שיטות ואלגוריתמים	➤
6	2.1 MFCC – Mel frequency coefficients	•
6	2.1.1 הסבר כללי	❖
7	2.1.2 פונקציית המרה	❖
7	2.1.3 MFB - Mel Frequency bank	❖
8	2.1.4 חישוב המסננים	❖
9	2.1.5 Bird filter bank	❖
10	2.2 K-Means	•
10	2.2.1 הסבר כללי	❖
11	2.2.2 סימונים והגדרות	❖
11	2.2.3 האלגוריתם	❖
12	2.3 K-Medoids	•
12	2.3.1 הסבר כללי	❖
12	2.3.2 סימונים והגדרות	❖
13	2.3.3 תיאור האלגוריתם	❖
14	2.4 Rand index	•
14	2.4.1 הסבר כללי	❖
14	2.4.2 חישוב המדד	❖
15	2.5 Silhouette - מדד סילואט	•
15	2.5.1 ערך סילואט	❖
15	2.5.2 סימונים והגדרות	❖
16	2.5.3 האלגוריתם	❖
16	2.6 T-distributed Stochastic Neighbor Embedding (t-SNE)	•
17	2.7 Artificial Neural Network	•

17	.....	2.7.1 הקדמה	❖
18	.....	2.7.2 מבנה	❖
19	.....	2.7.3 שימוש	❖
19	.....	2.7.4 פונקצית מחיר	❖
20	.....	Back propagation 2.7.5	❖
21	.....	2.8 רשת קוהונן – Self Organizing Map (Som)	•
21	.....	2.8.1 הקדמה	❖
21	.....	2.8.2 הרעיון הכללי	❖
21	.....	2.8.3 סימונים והגדרות	❖
22	.....	2.8.4 האלגוריתם	❖
24	.....	פרק 3: תוצאות	➤
24	.....	3.1 תוצאות – Clustering	•
26	.....	3.2 תוצאות T-SNE	•
29	.....	3.3 תוצאות – SOM	•
30	.....	3.4 תוצאות רשת נוירונים	•
30	.....	3.5 סיכום	•
31	.....	ביבליוגרפיה	➤

# פרק 1: מבוא

## 1.1 סווג ואשכול של אותות קול

בעידן המידע הדיגיטלי, אותות קול הפכו להיות חלק משמעותי במספר רב של יישומים מודרניים. החל מניתוח גלים סיסמיים המשמש להבנת המבנה הפנימי של כדור הארץ, ועד לניתוח אותות ביו-אקוסטיים שמספקים לחוקרים פרספקטיבה חדשה על עולם החי. ליישומים אלה השפעה רבה על חיי היום יום, ועל ההתקדמות הטכנולוגית של המין האנושי במישרין או בעקיפין. סווג של אותות קול מאפשר לאתר ולזהות דפוסים שונים באוספי תצפיות שנעשו, ולהסיק מסקנות על אוספים אלה. למשל יישומים המשמשים לזיהוי קולי או זיהוי דפוסים חוזרים באותות קול המופקים על ידי בעלי חיים, למטרת הסקת מסקנות הנוגעות למציאות בטבע, שבפרויקט זה תוצג דוגמא לכך. תחום זה הולך ומתפתח באופן מתמיד, ועושה בין השאר שימוש נרחב בכלים מעולם למידת המכונה. מאחר ולא תמיד יש בדינו, או שלא נהיה מעוניינים לעשות שימוש בסיווגים ידניים שנעשו על ידי גורם אנושי, אנו נעזרים בין השאר באלגוריתמי אשכול שמטרתם היא לסווג ערכים דומים לקבוצות מתאימות בצורה אובייקטיבית ואוטומטית מקצה לקצה.

## 1.2 רקע על השלדג לבן החזה

שלדג לבן חזה הוא מין עוף הנמנה עם משפחת השלדגים שמסדרת הכחלאים הוא נפוץ מאוד באיזורים שבין איזור הבלקן והמזרח התיכון במערב, עד לפיליפינים בשבמזרח. בניגוד למיני שלדג אחרים, ניתן למצוא את לבן החזה גם באיזורים שאינם קרובים למקורות מים. בישראל לבן-חזה נפוץ ביותר בצפון הארץ ובמרכז, והחל מסוף שנות ה-80 נצפה גם באיזורים חקלאיים בנגב. (ויקיפדיה). לבן-חזה, זכה בשמו המדעי *Halcyon smyrnensis* על שם העיר סמירנה שהוא שמה היווני של איזמיר. תזונת לבן-חזה מתחלקת לשתי קבוצות. הראשונה כאשר הוא מצוי ליד מקורות מים מורכבת בעיקר מדגים, ראשנים של דוחיים ודוחיים בוגרים כמו אילניות וצפרדעים. כאשר לבן-חזה אינו מצוי בקרבת מקור מים, תזונתו מורכבת בעיקר מערצבים, זוחלים שונים (זיקיות, חומטים, נחשים), וכן סרטנים, חגבים וחיפושיות (yardbirds.org).

שירת השלדג הוא תחום שכמעט לא נחקר, אחת הסיבות לכך היא שאזורי התפוצה של שלדג לבן החזה הם ברובם במדינות שאינן מתמחות בנושאי מחקר זה לדוגמת הודו, המזרח התיכון ודרום מזרח אסיה. הפרויקט המוצע בעקבות פרויקט מחקרי משותף של דנה קליין בהנחיית יוני וורטמן ויזהר לבנר, מחקר אקולוגי התנהגותי וביו-אקוסטי על שלדגים לבני חזה, ובפרט בקשר בין השירה להתנהגות.

### 1.3 נושא המחקר

סיווג קולות שירתו של השלדג.

דו"ח סיכום זה מהווה חלק ממחקר מקיף שעוסק בשלדג לבן החזה ובשירתו בפרט, אשר במסגרתו חולקו קולות השירה של השלדג לכ- 13 מחלקות על-ידי מומחה אנושי באמצעות האזנה, צפיה בהתנהגות וניתוח ספקטרוגרמות, בחרנו לעסוק ב- 10 מהן:

- Alert – נשמע בדרך כלל בעת מנוסה או בהברחה/ תקיפה של בעל חיים אחר
- Call – נשמע בין היתר בחלק מהתקשורת עם הפרחונים
- Rattle – נשמע בזמן הפריחה מהקן
- Snore – נשמע לעיתים כתוספת לשירה ה-long trill
- Thin trill – נשמע לרוב בעת מפגש בין שני שלדגים. Short trill – נשמע לרוב בזמן תעופה
- Long trill – נמצא בקורלציה עם שלב החיזור, בניית הקן והקינון
- Trembling
- Tut
- Twittering

יש לצין שהסיווג שבוצע התבסס על תצפיות רבות ומומחיות של צוות המחקר, בתהליך זה מיצו החוקרים מדדים אמפיריים שונים כדוגמת משך ההברה, רוחב הפס תדירות מינימלית ומקסימלית בשירה, משך השירה ועוד. מטרת הפרוייקט היא הפעלה של כלים חישוביים כדי לתקף את החלוקה שהתבצעה על ידי צוות החוקרים. על-ידי השוואה בין החלוקה לקבוצות של האלגוריתמים האובייקטיביים של האות האקוסטי לבין החלוקה (שעשויה להיות סובייקטיבית) של המומחה האנושי, נוכל לבחון את מידת ההתאמה בין שתי החלוקות.

### 1.4 מטרת הפרויקט

על-ידי סיווג של קולות שירת השלדג באמצעות אלגוריתמי למידה מודרכת (supervised) ולא מודרכת (unsupervised) רצינו להשיג שתי מטרות:

1. לענות על השאלה - האם אלגוריתם ניתוח אשכולות (Clustering) אוטומטי ואובייקטיבי תומך בחלוקת הקולות השונים שבוצעה על-ידי מומחה אנושי?
2. שימוש בדרכים שונות להצגת מידע רב-מימדי שהופק בתהליכי הניתוח והסיווג במימד נמוך כדי שנוכל להציג באופן ויזואלי את הקשרים והקירבה בין המחלקות השונות המייצגות את קולות השירה של השלדג.

### 1.5 מאפייני האות

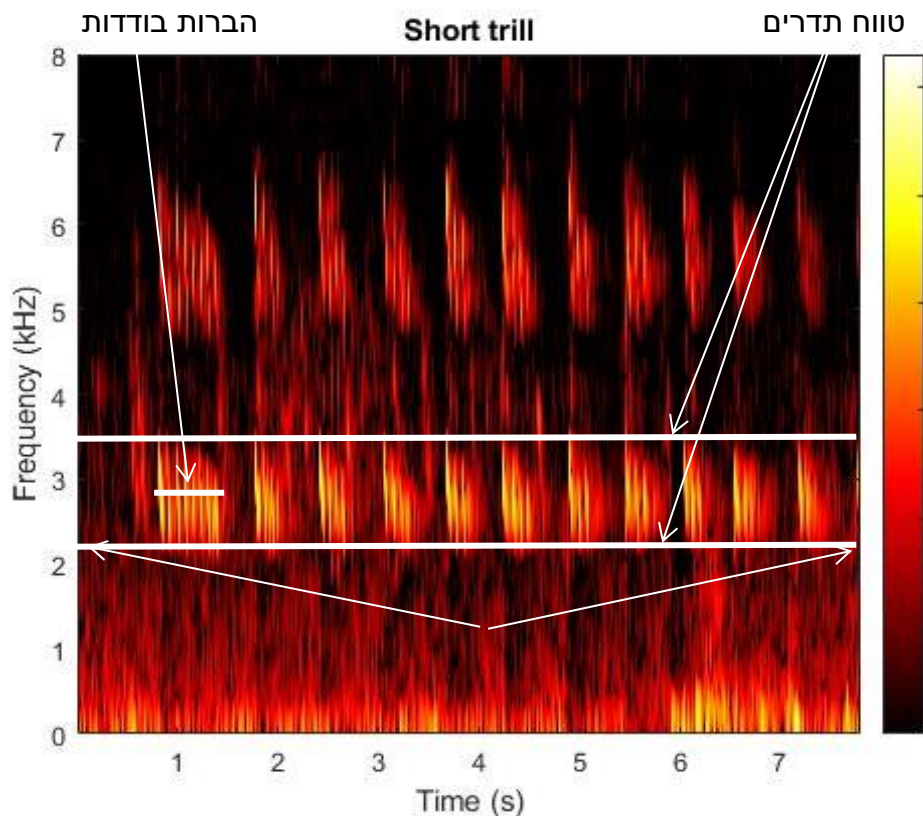
1. רוב קריאות השלדג בנויות ממספר הברות עוקבות שמרכיבות מילה אחת. ישנן גם מילים שמורכבות

מהברה אחת מתמשכת בעלות מאפיינים תדריים ייחודיים למילה.

2. נמצא שרוב האנרגיה בקריאות השלדג השונות נמצאת בטווח התדרים 2-4 kHz.

3. מילים אלה נבדלות אחת מהשניה בעיקר בקצב ההברות, מספר ההברות בכל קריאה, משך ההברה ותדירותה.

אחד מהכלים שבהם נעשה שימוש לבחינת אותות הקול נקרא ספקטרוגרמה. (איור 1) הספקטרוגרמה מייצגת את השתנות האנרגיה בתדרים השונים של האות לאורך זמן. באיור שלפנינו אפשר לראות את הייצוג של קריאת Short-trill על ידי הספקטרוגרמה. נשים לב למילה שמורכבת מהברות המופיעות (באופן יחסי) בקצב, טווח תדירויות ובמשך הברה אחדים.



איור 1: ספקטרוגרמה של קריאת Short-trill.

## 1.6 העבודה שלנו

תהליך הסיווג וסיווג אותות קול בפרט כולל שני שלבים:

- שלב 1** – עיבוד מקדים לאות, הכולל לרוב סינון מרעשים, ומיצוי נתוני התדר מהאות בעזרת שיטות יעודיות לקבלת ייצוג עם מימד נמוך יותר מהמקור תוך שמירה על מרכיביו החיוניים.
- שלב 2** – הפעלה של אלגוריתמי למידה על ייצוגי האותות למטרת סיווג והפרדה למחלקות, ולאחר מכן בחינת התוצאות כאשר השאלה המרכזית הנשאלת היא "האם יש התאמה לסיווג שניתן על ידי צוות המחקר הביולוגי?".

**בשלב הראשון** השתמשנו ב MFCC – Mel Frequency Cepstral Coefficient

מטרת הליך זה היא מיצוי מאפייני התדר של אות קול נתון תוך הפחתת מספר המימדים של הייצוג הספקטראלי. התכונה המרכזית של MFCC היא שהמאפיינים התדריים מחושבים באופן דומה לייצוגם במערכת השמיעה האנושית. הליך זה נפוץ מאוד במטלות של זיהוי וסיווג אותות קול. הוצע להשתמש בהליך דמוי MFCC, ובנוסף נבחנו גם גישה לשיפור השיטה על-ידי שינוי הרזולוציה של סולם התדירויות כך שהרזולוציה המירבית תתקבל בתדירויות שאנו משערים שבהן לשלדג איכות השמיעה היא אופטימלית. הגישה מכונה Bird filter bank ונפרט עליה בפרק 2.1.5.

בנוסף לוקטור המאפיינים שהתקבל יש ברשותנו מאפיינים נוספים שנמדדו באופן ידני על-ידי החוקרים. בחנו את כל אחד ממאגרי המאפיינים בנפרד וביחד, כלומר סה"כ שלושה מאגרי מאפיינים שונים:

1. מאגר מאפיינים המבוסס על מקדמי ה- MFCC
  2. מאגר הכולל את המאפיינים שחושבו על-פי הספקטרוגרמה בעזרת שימוש בתוכנת Avisoft בעבודת המחקר.
  3. שילוב של שני המאגרים הראשונים.
- מאגר המאפיינים המשותף הוא אוסף של וקטורים, שבו כל וקטור מייצג מיצוי מאפיינים של מילה אחת. וקטור זה נוצר משירשור של המאפיינים האמפירים אשר חושבו עבור אותה המילה, עם ממוצע מאפייני ה-MFCC אשר חושבו עבור כל הברה ששיכת למילה.

**בשלב השני** לאחר ביצוע העיבוד המקדים לקולות שירת השלדג, הופעלו כלים של למידה מודרכת ולא מודרכת לסיווג הנתונים והצגתם. נעשה שימוש באלגוריתמים של למידה למידה לא מודרכת (Clustering) כמו K-Means ו-K-Medoids, וכן במפות אירגון עצמי (Self-Organizing Map) להצגה ויזואלית של הנתונים על המישור. כמו-כן השתמשנו ברשת עצבית פשוטה לסווג סוגי הקריאות באופן אוטומטי, שעבורה נתוני האימון התקבלו מהסווג שנעשה במחקר הביולוגי.

## פרק 2 : שיטות ואלגוריתמים

### MFCC – Mel frequency coefficients 2.1

#### 2.1.1 הסבר כללי

ייצוג אות קולי באמצעות מקדמי קפסטרום התדירות לפי סולם מלל MFCC- Mel Frequency Cepstrum coefficients היא אחת הדרכים המקובלות ביותר לניתוח מאפייני אותות קול. אחת הסיבות לכך היא מיצוי הספקטרום של האות על-ידי מסננים המרווחים בדומה לדרך העיבוד של מערכת השמיעה האנושית. הניתוח מתבצע באמצעות סידור מסננים בסקלה לוגריתמית המכונה סולם מל על ספקטרום התדירויות של האות (איור 2), העברת האות דרך מסננים אלה ושמירת יציאות המסננים כוקטור פרמטרים.

### 2.1.2 פונקציית המרה

ראשית נגדיר את פונקציית ההמרה שמתארת את המעבר מתדירות  $l$  בסקלה ליניארית, לתדירות בסקלת Mel:

$$freq2mel(l) = \frac{F}{\log(1 + \frac{F}{700})} \log(1 + \frac{abs(l)}{700}) sign(l)$$

כאשר הפרמטר  $F$  מייצג את הנקודה שבה התדירות לפי סולם מל שווה לתדירות לפי הסקלה הליניארית הרגילה,  $Mel = Frequency$ . כלומר שמתדירות  $F$  והלאה, הטווח שכל מסנן יכסה יגדל באופן לוגריתמי ויביא למשקל הולך ופוחת לתדירויות הגדולות מ- $F$ .

כעת נגדיר את פונקציית ההמרה המתארת את המעבר מתדירות  $m$  בסקלת Mel לתדירות בסקלה ליניארית.

$$mel2freq(m) = 700 * sign(m) * \left( \exp\left( abs(m) \frac{\log(1 + \frac{F}{700})}{F} \right) - 1 \right)$$

כאשר  $F$  הוא קבוע שערכו נקבע כ-2500 Hz.

### 2.1.3 MFB - Mel Frequency bank

באופן כללי, נרצה לחלק את טווח התדרים הנתון לחלונות באורכים הגדלים באופן לוגריתמי לפי הגדרת MFCC המותאם. בין החלונות תהיה חפיפה שתוסבר בהמשך. כעת נבחר את מספר המסננים הרצוי  $K$ . עבור אות כניסה  $x(n)$  נבצע את התמרת פורייה הבדידה.

$$X[k] = |\sum_{n=0}^{M-1} x[n] * w[n] * e^{-i\frac{2\pi}{N}nk}| \quad k = 0, 1, \dots, N-1$$

על מנת לקבל  $K$  חלונות, נזדקק ל  $K+2$  נקודות שיהיו את נקודות ההתחלה והסיום של כל חלון:

$b_1, b_2, \dots, b_K$  מייצגים את החלונות עצמם.

$z_1, z_2, \dots, z_{K+2}$  מייצגים את מיקומי החלונות.

מיקומו של כל חלון יקבע לפי טווח התדרים שעליו לכסות.

בשביל להשיג מיקומים אלו ב  $X[k]$  נבצע התאמה של תדירות רצויה, למיקום על האות הבדיד  $X[k]$ .

כאשר  $fs$  היא תדירות הדגימה של האות המקורי,  $\frac{fs}{2}$  היא התדירות המירבית. היחס בין תדירות כלשהי  $f$  לבין

התדירות המירבית, הגדולה או שווה ל- $f$ , יהווה את המיקום היחסי של התדירות  $f$  באות שנדגם בתדירות  $fs$ . כעת

אם נכפול יחס זה באורך הדגימה  $X[k]$ , נקבל את המיקום שבו תדירות  $f$  מיוצגת באות  $X[k]$ .

$f_l$  מייצגת את הטווח התחתון של התדירויות שהמשמש מעוניין לנתח, ו- $f_h$  את הטווח העליון.

$N(X[k])$  מייצג את אורך הדגימה  $X[k]$ .



נמיר את ערכי תדירויות הקצה לסקלת mel.

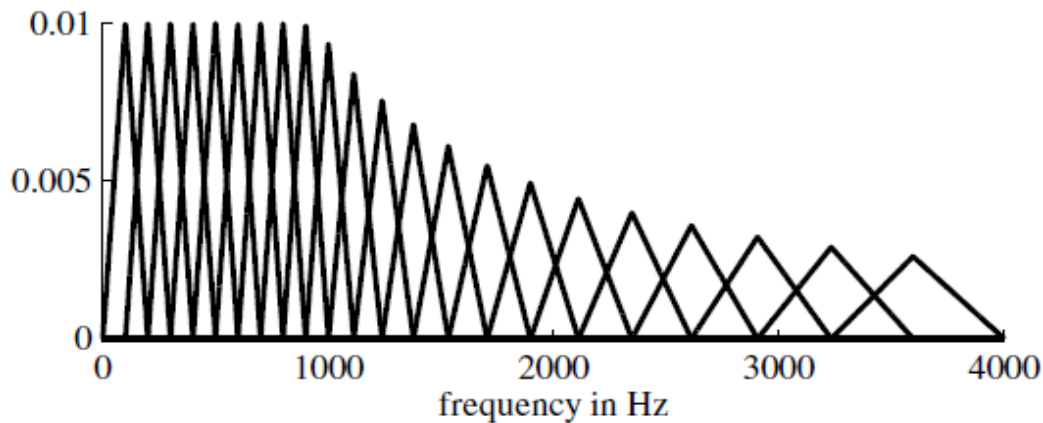
$$mf_h = \text{freq2mel}(f_h) \quad mf_l = \text{freq2mel}(f_l)$$

כעת נחלק את תחום הערכים החדש למספר הפילטרים הרצוי, ונקבל "צעד קפיצה"  $p$  המגדיר טווח תדירויות לכל חלון ביחס לתדירות ההתחלה והסיום באופן הבא:

$$p = \frac{mf_h - mf_l}{K}$$

כאשר המרווח בין כל חלון הוא ערך קבוע בסקלת Mel, מעבר בחזרה לסקלה ליניארית יביא למרווחים הלוגריתמים הנדרשים. ערכי mel יקבעו לפי:

$$mf_i = mf_l + p(i - 1), \quad i = 1, 2, \dots, K + 2$$



איור 2: מסנני mel ממושקלים עבור הקבוע  $F = 1000$

#### 2.1.4 חישוב המסננים

כעת, נחשב את המיקומים  $z_1, z_2, \dots, z_{K+2}$  של כל החלונות מתוך הערכים  $mf_i, i = 1, 2, \dots, K + 2$  שהתקבלו באופן הבא:

$$z_i = \frac{2 \text{mel2freq}(mf_i)}{f_s} * N(X(H)), \quad i = 1, 2, \dots, K + 2$$

כל חלון הוא בעצם טווח מסוים מתוך  $X(H)$ , ומוגדר באופן הבא:

$$b_k = [X(z_k), X(z_k + 1), \dots, X(z_{k+2})]$$

אורכו של  $b_k$  הוא ההפרש בין מיקום התחלתו למיקום סיומו.  $\text{length}(b_k) = z_{k+2} - z_k$

נגדיר מסנן  $S_k$  עבור כל חלון  $b_k$ .

אורכו של מסנן  $S_k$  יהיה כאורכו של  $b_k$ . על מנת לקבל משולש שווה שוקיים שהערך בקודקוד הראשי יהיה 1, נגדיר ערך קפיצה  $step(b_k)$  שיחושב כך:

$$step(b_k) = \frac{2}{length(b_k)}$$

ערכי  $S_k$  יעלו בצורה הדרגתית החל מ-  $step(b_k)$  במיקום הראשון עד ל-1 במיקום האמצעי, ומשם ירדו בכל שלב ב- $step(b_k)$ , עד לערך של  $step(b_k)$  במיקום האחרון.

$$S_k = \left[ 1 * step(b_k), 2 * step(b_k), \dots, \left( \frac{length(b_k)}{2} \right) * step(b_k), \left( \frac{length(b_k)}{2} - 1 \right) * step(b_k), \dots, 1 * step(b_k) \right]$$

צורתו המשולשת של החלון תביא לכך שרוב המשקל הוא באמצע החלון.

פעולת הסינון תבצע על ידי מכפלה פנימית ולאחר מכן מיצוע בין חלון האות לבין המסנן המתאים לו.

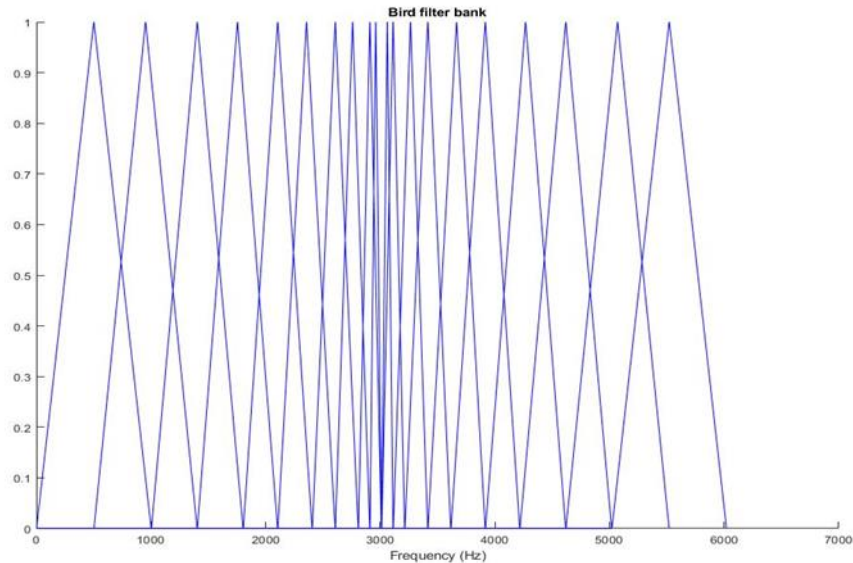
הערך המתקבל, המסומן ב-  $R(k)$ , מהווה ערך משוקלל למידת האנרגיה בתחום התדרים של החלון ה- $k$  האלגוריתם מחזיר ערך זה עבור כל חלון, ועל ערכים אלו יתבצע הסיווג.

$$R(k) = \frac{1}{length(bk)} \sum_{i=1}^{length(b_k)} X(z_k + i) S_k(i)$$

בתהליך ה MFCC הסטנדרטי יש שלב נוסף discrete cosine transform לביצוע דה-קורלציה שנוצרה עקב חפיפת המסננים.

### Bird filter bank 2.1.5

מכיוון שרוחב הפס של קריאות השלדג נמצא בטווח 2-4 Khz הוצע לשנות את ה-MFB כך שעיקר הרזולוציה של מאגר המסננים יהיה גם כן סביב טווח זה (איור 3). המסננים הולכים וגדלים בצורה לוגריתמית בכיוונים מנוגדים, החל מהתדירות הרצויה, במקרה זה התדירות היא 3000. תהליך בניית המסננים ואופי חפיפתם דומה לזה שתואר בפרקים 2.1.3, 2.1.4.



איור 3: בנק המסננים לפי סולם "Bird" (Bird filter bank).

## K-Means 2.2

### 2.2.1 הסבר כללי

בהינתן קבוצה  $S$  המכילה  $m$  תצפיות מטרתו של האלגוריתם היא חלוקה של הקבוצה ל- $K$  תתי קבוצות על ידי מציאת  $k$  מרכזי כובד ושיוך של כל תצפית לאחד מהם לפי קרבתם. בחירה נכונה של מרכזי הכובד עשויה לאפשר חלוקה המפרידה את התצפיות בצורה טובה. לעיתים נדרשות תצפיות רבות על מנת להשתמש במודל בצורה אופטימלית, והוספה של תצפית חדשה מחייבת חישוב חוזר. מדובר באלגוריתם היוריסטי כלומר ניתן להראות שסכום המרחקים בין כול התצפיות באותה הקבוצה למרכז הכובד המיצג אכן יתכנס למינימום אך לא מובטח שהמינימום הוא מוחלט, כלומר יתכן שקיים פתרון או פתרונות אחרים טובים יותר. סכום המרחקים נקרא גם "שגיאה" והוא משערך עד כמה התצפיות קרובות למרכזי הכובד שלהם בזמן נתון. מאחר והאלגוריתם רגיש לתנאי התחלה (נבחרים באופן רנדומלי), כדי להגיע לתוצאה אופטימאלית יש להריץ את האלגוריתם מספר פעמים ולבחור את ההרצה שהשיגה את התוצאה הטובה ביותר. זהו מודל סטטיסטי שאינו מתבסס על ידע מוקדם על הנתונים אלא מסתמך אך ורק על התצפיות בפועל. המטרה היא מציאה של  $k$  מרכזי כובד ושיוך של כול תצפית לאחד מהם, תוך כדי צמצום של סכום המרחקים בין כל מרכז כובד לתצפיות המשויכות אליו. ניתן להשתמש במטריקות מרחק שונות, הנהוגה ביותר היא המרחק האוקלידי:

$$d(p, q) = d(q, p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

תחילה יש לקבוע מרכזי כובד ראשוניים, ישנן כמה דרכים לעשות זאת אנחנו בחרנו בשיטת E.W Forgy הבוחרת באופן  $k$  נקודות אקראיות במרחב להיות מרכזי הכובד. ניתן לשייך באופן רנדומלי את כול אחת מהתצפיות לקטגוריה מסויימת אך במקרה זה השיוך הראשוני לכל תצפית, הוא מרכז הכובד הקרוב ביותר אליה שמיקומו נבחר באופן אקראי. לאחר מכן האלגוריתם הוא איטרטיבי ופועל במספר שלבים

## 2.2.2 סימונים והגדרות

-  $V(x_i)$  - מייצג את מרכז הכובד של הנקודה  $x_i$ .

-  $N(\Phi_k^t)$  - מייצג את מספר הנקודות המשויכות למרכז הכובד  $\Phi_k$  בזמן  $t$ .

## 2.2.3 האלגוריתם

נגדיר חסם עליון למספר האיטרציות המירבי  $T$ .

2. נבחר  $K$  מרכזי כובד  $\Phi_k^t$ ,  $k = 1, 2, \dots, K$  במיקום אקראי ממימד  $n$ .

3. נגדיר פונקציית עלות  $J(X)$  - סכום ההפרשים הממוצע בין מיקומי מרכזי כובד בזמן  $t$  לבין לבין מיקומם בזמן  $t + 1$ .

$$J(X)^{t+1} = \sum_{k=1}^K \|\Phi_k^{t+1} - \Phi_k^t\|^2$$

4. כל עוד מספר האיטרציות  $t$  קטן מ-  $T$  נבצע:

4.1 נשייך כל נקודה למרכז הכובד הקרוב ביותר אליה.

$$V(x_i) = \{\Phi_z | d(x_i, \Phi_z) \leq d(x_i, \Phi_k) \text{ , } k = 1, 2, \dots, K \} \text{ , } i = 1, 2, \dots, m$$

4.1 עבור כל מרכז כובד  $\Phi_k^t$ :

בשלב זה נחשב את מיקום מרכז הכובד החדש. מיקום זה יהיה ממוצע ערכי כל התצפיות המשויכות למרכז הכובד  $\Phi_k^t$ . בצורה וקטורית:

$$\Phi_k^{t+1} = \frac{1}{N(\Phi_k^t)} \sum_{x_j \in \Phi_k^t} x_j$$

4.2 אם פונקציית העלות קטנה מערך השגיאה המקסימלי  $\varepsilon$ , כלומר מרכזי הכובד החדשים קרובים במידה רבה למרכזי הכובד באיטרציה הקודמת.

$$J(X)^{t+1} = \sum_{k=1}^K \|\Phi_k^{t+1} - \Phi_k^t\|^2 < \varepsilon$$

האלגוריתם יעצור.

אחרת, הוא ימשיך לאיטרציה הבאה.

## K-Medoids 2.3

### 2.3.1 הסבר כללי

לאלגוריתם k-medoids יש מספר קווי דמיון לאלגוריתם k-means. שניהם אלגוריתמי אשכול שנועדו לחלק קבוצת נתונים S ל-K מחלקות זרות, ושניהם עושים זאת באמצעות מזעור המרחק בין תצפיות לבין מרכזי כובד המתאימים להם. מרחק זה נקרא "השגיאה". גם כאן, ניתן לחשב את המרחקים בשיטות שונות, אך יש לשמור על עקביות. בפרויקט זה נעשה שימוש במרחק אוקלידי. האלגוריתם בוחר את מרכזי הכובד מהתצפיות הנתונות בלבד, בניגוד ל k-means שבו אין הגבלה על מיקום מרכזי הכובד במרחב. מהסיבה הזאת k-medoids עמיד יותר למחריגים ופחות רגיש לתנאי התחלה מאשר k-means. בשלב הראשון מרכזי הכובד האלו נבחרים באופן אקראי מתוך הנתונים. בדומה ל k-means, האלגוריתם רגיש (במידה פחותה) לתנאי ההתחלה ומבטיח למזער את השגיאה לכדי מינימום מקומי.

### 2.3.2 סימונים והגדרות

- נחליט על פונקציית מרחק שבה נרצה להשתמש. במקרה זה נבחר במרחק אוקלידי ונסמן

$$d(x_i, x_j) \text{ כך:}$$

-  $V(x_i)$  - מייצג את המדואיד של הנקודה  $x_i$ .

-  $N(\Phi_k)$  - מייצג את מספר הנקודות המשויות למדואיד  $\Phi_k$ .

-  $M(\Phi_k)$  - מייצג את נקודת מרכז הכובד הנוכחית של מדואיד  $\Phi_k$ .

-  $\delta_k^t$  - מייצג האם נקודת מרכז הכובד של המדואיד  $\Phi_k$  השתנתה בזמן  $t$ . תחילה

יאותחל ב-1 עבור כל k.

$d(x_z^k, \Phi_k)$  - מייצג את המרחק הממוצע של נקודה  $x_z^k$  מכלל הנקודות שמשויכות ל  $\Phi_k$ ,

והוא מחושב באופן הבא:

$$d(x_z^k, \Phi_k) = \frac{1}{N(\Phi_k)} \sum_{i=1}^{N(\Phi_k)} d(x_z^k, x_i^k)$$

### 2.3.3 תיאור האלגוריתם

1. נגדיר חסם עליון למספר האיטרציות הרצויות T.
2. מתוך אוסף נתונים  $X = x_1, x_2, \dots, x_m$  בעל m נקודות ממימד n, נבחר באקראי K נקודות כמדואידים התחלתיים  $\Phi_k, k=1,2,\dots,K$ .
3. נגדיר פונקציית עלות  $J(X)$  שמטרתה לספק אינדיקציה לשינויים שנעשו בנקודות מרכזי הכובד בזמן כלשהו t. אם בזמן t לא נבחרו מדואידים חדשים, משמע שהנקודות שמאפיינות אותם ממורכזות בצורה אופטימלית, ולכן ניתן לסיים את הריצה.

$$J(X) = \sum_{k=1}^K \{\delta_k^t\}$$

4. כל עוד מספר האיטרציות t קטן מ-T נבצע:  
4.1 נשייך כל נקודה למדואיד הקרוב ביותר אליה.

$$V(x_i) = \{\Phi_z | d(x_i, \Phi_z) < d(x_i, \Phi_k), k = 1, 2, \dots, K \wedge i = 1, 2, \dots, m \wedge k \neq z\}$$

4.2 עבור כל מדואיד  $\Phi_k$ :

קיימת  $x_p^k$  שהיא הנקודה הממוכזת ביותר מבין כל הנקודות שמשויכות ל  $\Phi_k$ . נקודה זו תחליף את מקומה של הנקודה הנוכחית שמאפיינת את המדואיד, במידה והן שונות.

$$x_p^k = \arg \min \{d(x_z^k, \Phi_k) | x_z^k \in \Phi_k\}$$

כאשר  $d(x_z^k, \Phi_k)$  מייצג את המרחק הממוצע של נקודה  $x_z^k$  מכלל הנקודות שמשויכות ל  $\Phi_k$ , והוא מחושב באופן הבא:

$$d(x_z^k, \Phi_k) = \frac{1}{N(\Phi_k)} \sum_{i=1}^{N(\Phi_k)} d(x_z^k, x_i^k)$$

אם הנקודה  $x_p^k$  שונה מנקודת מרכז הכובד שהייתה עד עכשיו, נעדכן את  $\delta_k^t$  להיות 1, אחרת נעדכן ל-0.

$$\delta_k^t = \begin{cases} 1, & x_p^k \neq M(\Phi_k) \\ 0, & \text{otherwise} \end{cases}$$

5. לאחר מעבר על כל המדואידים  $\Phi_k, k=1,2,\dots,K$  נרצה לברר האם המדואידים לא קיבלו מרכזי כובד חדשים כלל בזמן t, כלומר האם פונקציית המחיר שווה ל-0.

$$J(X) = \sum_{k=1}^K \{\delta_k^x\}$$

במידה וכן, נסיים. אחרת, נבצע את תהליך זה מחדש.

## Rand index 2.4

### 2.4.1 הסבר כללי

אלגוריתם זה מספק מדד הדמיון בין שתי חלוקות שונות של אותה קבוצת נתונים. משמש בעיקר לצורך השוואה בין שני סיווגים שהגיעו מאלגוריתמי סיווג בלתי מודרכים למשל K-means או K-medoids.

מכיוון שאלגוריתמי סיווג בלתי מודרכים מבטאים שייכות למחלקה מסוימת באמצעות מספר מחלקה, ומספר זה נקבע על פי ערכים שנקבעים שרירותית בזמן ריצת האלגוריתם, יתכן מצב שעבור זוג סיווגים זהה, מספר המחלקה שיותאם בפועל לכל מחלקה הוא שונה. במקרה כזה אפשר לחשוב בטעות שהסיווגים שונים כשלמעשה הם זהים. האלגוריתם מחזיר ציון בצורת סקלר שמבטא את רמת הדמיון בין שני סיווגים של אותו סט נתונים.

### 2.4.2 חישוב המדד

בהינתן קבוצה  $S = \{\sigma_1, \sigma_2, \dots, \sigma_n\}$  ובהינתן שתי חלוקות ל- $S$  נסמן:  $X = \{X_1, X_2, \dots, X_r\}$  חלוקה של  $S$  ל- $r$  תתי מחלקות זרות. וכן:  $Y = \{Y_1, Y_2, \dots, Y_t\}$  חלוקה של  $S$  ל- $t$  תתי מחלקות זרות.

עבור כל הזוגות האפשריים של איברים מ- $S$ , נבחן את הקטלוג משתי החלוקות  $X, Y$ .

כל זוג יהיה משויך לאחת מ-4 הקבוצות הבאות:

a – זוגות הנמצאים באותה מחלקה גם ב- $X$  וגם ב- $Y$ .

b – זוגות הנמצאים במחלקות שונות גם ב- $X$  וגם ב- $Y$ .

c – זוגות הנמצאים במחלקות זהות ב- $X$  אך שונות ב- $Y$ .

d – זוגות הנמצאים במחלקות שונות ב- $X$  אך זהות ב- $Y$ .

$$R = \frac{a + b}{a + b + c + d} = \frac{a + b}{\binom{n}{2}}$$

מדד זה הוא למעשה יחס בין מספר הזוגות המקוטלגים זהה או שונה באופן אחיד בשני הסיווגים, לבין סה"כ הזוגות האפשריים.

יתרונות:

- יכול לתת מענה לבעיה שכיתה למדי שעולה בעת שימוש באלגוריתמי סיווג בלתי מודרכים.
- מתבצע בזמן ריצה נמוך יחסית.

חסרונות:

- הנסיון שלנו עד כה המדד לא נתן תוצאות אמינות כיוון שהרצתו על סיווג רנדומלי הובילה לקבלת ערכים גבוהים ביחס לאלה שהתקבלו מהרצתו על חלוקות אחרות שביצענו.

## 2.5 מדד סילואט - Silhouette

מדד סילואט הוא מדד המאפשר לבחון עד כמה חלוקת נתונים במאגר מידע למחלקות שונות שביצענו נכונה ומתאימה. מדד זה מאפשר לתאר בצורה גרפית תמציתית עד כמה דומה נתון מסוים באשכול נתונים לכלל הנתונים באותו אשכול. ערך סילואט נע בין 1 ל -1. (איור 4)

### 2.5.1 ערך סילואט

משמעות ערך סילואט של נקודה  $s(i)$ :

1. כאשר  $s(i)$  קרוב ל-1 - ניתן לומר כי הנתון ממורכז באשכול שלו - לכידות גבוהה.
2. כאשר  $s(i)$  קרוב ל-0 - ניתן לומר כי הנתון נמצא קרוב מאוד לגבול בין שני אשכולות שכנים - נמצא בגבול בין שני אשכולות.
3. כאשר  $s(i)$  קרוב ל-(-1) ניתן לומר כי הנתון נמצא באשכול לא מתאים - לכידות נמוכה.

על מנת לחשב ערך סילואט  $s(i)$  נדרשים שני פרמטרים:

1.  $a(i)$  - מסמל את המרחק הממוצע של נקודה  $i$  מכלל הנקודות באשכול שלה.
2.  $b(i)$  - מסמל את המרחק הממוצע של נקודה  $i$  מכלל הנקודות באשכול הקרוב ביותר אליה.

ערך סילואט מייצג את היחס בין  $a(i)$  ל  $b(i)$  באופן הבא:

$$s(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))}$$

### 2.5.2 סימונים והגדרות

- $x_i$  - מייצג את מספר האשכול של הנקודה  $V(x_i)$ .
- $N(\Phi_k)$  - מספר התצפיות שמשייכות לאשכול  $\Phi_k$ .
- $d_k^i$  - מרחק בין נקודה  $x_i$  לבין אשכול  $\Phi_k$  כממוצע סכום המרחקים בין  $x_i$  לבין שאר הנקודות ששייכות לאשכול  $\Phi_k$ . כלומר

$$d(x_i, \Phi_k) = \left\{ \frac{(\sum d(x_i, x_j))}{N(\Phi_k)} \mid x_j \text{ in } \Phi_k \right\}$$



### 2.5.3 האלגוריתם

-יהי אוסף נתונים  $X = x_1, x_2, \dots, x_m$  בעל  $m$  תצפיות ממימד  $n$ , שחולק בטכניקה כלשהי ל- $K$  אשכולות. כלומר לכל תצפית יש תווית שניתנה מראש המייצגת את מספר האשכול שבה היא נמצאת.

1. עבור כל נקודה  $x_i$ .

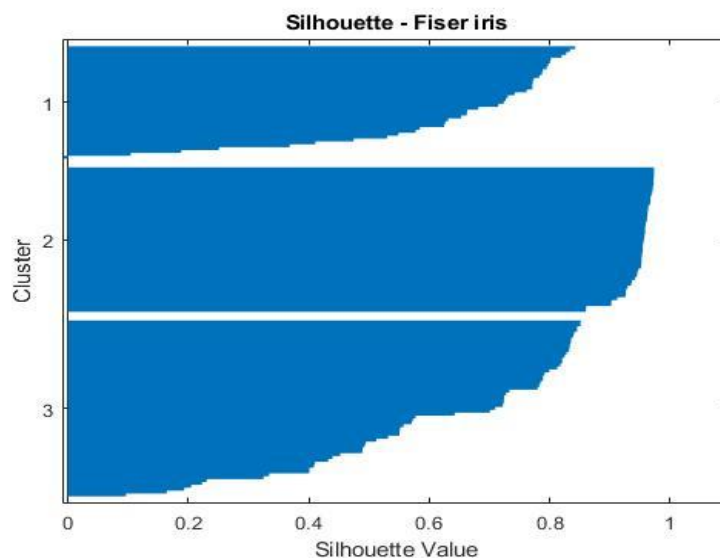
a. עבור כל אשכול  $\Phi_k$  מצא את המרחק ממנו אל הנקודה  $x_i$ .  $d(x_i, \Phi_k)$  שד

b. נמצא את האשכול השכן  $\Phi_c$ , שהוא אשכול שונה מהאשכול של  $x_i$ , בעל ערך  $d(x_i, \Phi_k)$  מינימלי ביחס לשאר האשכולות.

$$\Phi_c = \arg \min \{d(x_i, \Phi_r) \mid r = 1, 2, \dots, K, r \neq k\}$$

c. בשלב זה, נוכל לחשב את ערך הסילואט לנקודה  $x_i$ :

$$a(i) = d(x_i, V(x_i)) \quad b(i) = d(x_i, \Phi_c) \quad s(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))}$$



איור 4: ערכי סילואט עבור כל הדגימות ב Fisher iris dataset מחולקים לפי חלוקת אלגוריתם K-means.

### 2.6 T-distributed Stochastic Neighbor Embedding (t-SNE)

אלגוריתם t-SNE הוא אלגוריתם למידה חישובית המשמש להורדת מימדים וויזואליזציה. האלגוריתם פותח על ידי לורנס ואן דר מאטן וג'פרי הינטון (van der Maaten and Hinton). זאת שיטה שמתאימה במיוחד להורדת מימד של מרחבים ממימד גבוה למרחבים מממד 2 או 3 לכדי מפות. האלגוריתם מציג כל נקודה מהמרחב הרב-ממדי בעזרת נקודה דו ממדית או תלת ממדית כך שאובייקטים דומים יוצגו בסמוך אחד לשני, ואובייקטים שונים יוצגו רחוק יותר.

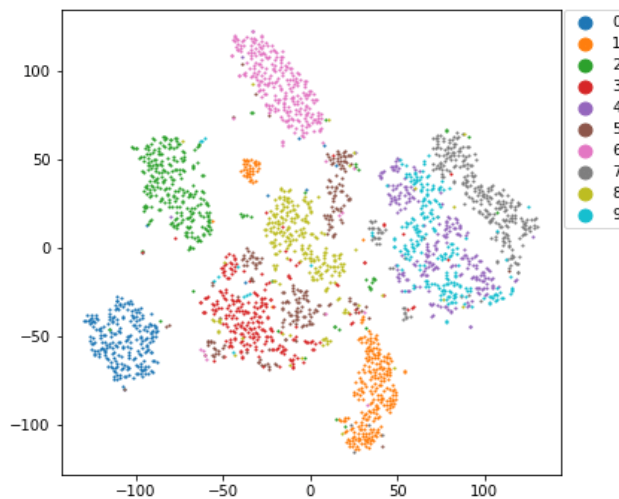
האלגוריתם פועל בשני שלבים עיקריים.

בשלב הראשון מחשבים דמיון בין כל זוג נקודות שיוצג באמצעות ערך הסתברותי גבוה לנקודות קרובות, וערך הסתברותי נמוך לנקודות רחוקות. הדמיון בין נקודה לעצמה מוגדר להיות 0, מכיוון שדמיון זה לא תורם לחלוקה הרצויה.

בשלב השני נגדיר נקודה ממימד נמוך במיקום אקראי עבור כל נקודה נתונה ממימד גבוה, ונחשב באופן דומה את הדמיון בין כל זוג נקודות במימד הנמוך.

לאחר מכן האלגוריתם מנסה למזער את המרחק בין ההתפלגויות השונות באמצעות Gradient decent האלגוריתם עושה שימוש במרחק אוקלידי לחישוב המרחקים.

אלגוריתם זה לא מומש על ידינו בפרויקט זה אלא נעשה שימוש באלגוריתם המובנה של Matlab.



איור 5: תוצאת הרצת tSNE על אוסף ספרות בכתב יד, הצבעים מתאימים למחלקות המקוריות. מקור: (tsne)

## Artificial Neural Network 2.7

### 2.7.1 הקדמה

רשת עצבית מלאכותית היא מודל מתמטי חישובי שפותח בהשראת תהליכים טבעיים המתרחשים ברשת העצבית של המוח האנושי.

נוירון הוא תא עצב המשמש כיחידת התפקוד הבסיסית של המוח. במוח האנושי רשת מסועפת של תאי עצב המתקשרים זה עם זה באמצעות אותות אלקטרו כימיים. כל תא עצב מסוגל ליצור דחפים עצביים בתגובה לגירויים מסוימים ולתקשר באופן זה עם תאים אחרים.

כאשר המוח האנושי לומד לבצע פעולה כלשהי, קשרי העצב או הסינפסות הקשורים לאותה פעולה מתחזקים. ככל שהפעולה תתבצע יותר, כך יתהדק הקשר בין הנוירונים בקשרים אלו.

נשים לב ל-3 מרכיבים עיקריים שניתן למצוא בתא העצב האנושי:

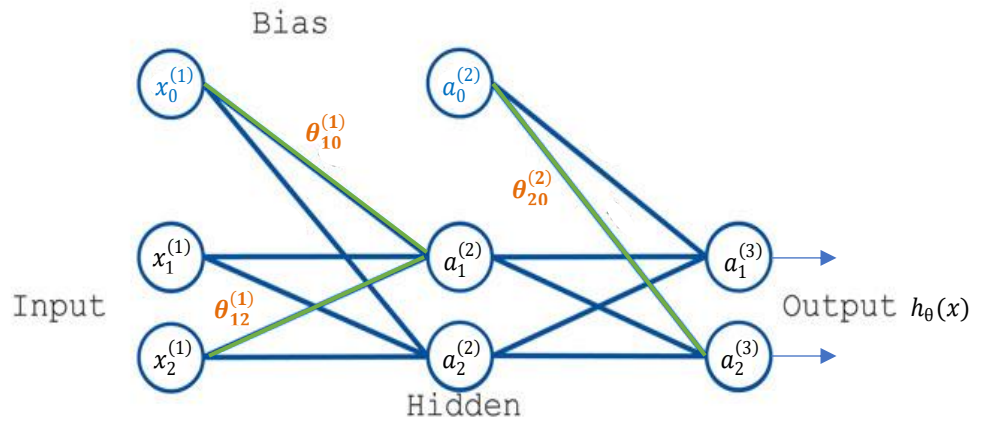
1. דנדריטים Dendrites – ערוצי הכניסה של נוירון מסוים מנוירונים שכנים, מידע חשמלי מתקבל ממספר נוירונים מקושרים בו זמנית.
  2. גוף התא Soma - כאשר הזרם שמתקבל מהדנדריטים חוצה ערך סף מסוים, תא העצב "יורה" אות חשמלי, ואז מאתחל את עצמו בציפייה לאירוע חציית סף נוסף.
  3. אקסונים Axons – דרכם תא העצב מעביר מידע חשמלי לנוירונים הבאים.
- נוירון מלאכותי מדמה את אופן פעולת הנוירון האנושי בכך שהוא מקבל קלט יחיד שמקורו בנוירון אחד או מספר נוירונים המחוברים אליו מכיוון אחד, מפעיל עליו פונקציית אקטיבציה כלשהי (בהקבלה לחציית ערך הסף), ומעביר ערכים הלאה לנוירונים המחוברים אליו בכיוון השני באופן שיוסבר בהמשך. באמצעות מודל זה ניתן ליצור משטחי הפרדה לא ליניארים מורכבים ביותר.

## 2.7.2 מבנה

רשת עצבית מלאכותית מכילה שכבות עוקבות של נוירונים מלאכותיים (ראה איור):  
שכבת כניסה – מורכבת ממספר נוירונים השווה למימד הקלט  $n$  ועוד כניסת bias.

$$x_0^{(1)}, x_1^{(1)}, \dots, x_n^{(1)}$$

1. שכבה חבוייה ראשונה – כל נוירון בשכבת הכניסה מחובר לכל נוירון בשכבה החבוייה הראשונה. (איור 5)  
חיבורים אלו נקראים גם "משקלות" והם מיוצגים באמצעות מטריצה  $\theta^{(1)}$ .  
למשל המשקולת בין הנוירון בשכבה הראשונה  $x_i^{(1)}$  לבין הנוירון בשכבת הקלט השניה  $a_j^{(2)}$  יסומן ב  $\theta_{ji}^{(1)}$ .  
המשקלות מבטאות את "חוזק" הקשר בין נוירונים, מכאן שניתן להשפיע על קשרים אלו באופן השואף לדמות את תהליך הלמידה הטבעי שקורה במוח.  
באופן כללי אם לרשת  $m$  יחידות בשכבה ה- $i$ , ו- $n$  יחידות בשכבה ה- $i + 1$ , מימדי מטריצת המשקלות יהיו  $n \times m + 1$
2. ניתן להוסיף שכבות חבוייות נוספות כאשר החיבור בין כל שכבה  $i$  לשכבה  $i + 1$  ייעשה באותו אופן שתואר ב-2.
3. שכבת היציאה – מורכבת ממספר נוירונים כמספר התוצאות האפשריות (ללא bias).



איור 5: דוגמה לרשת עצבית פשוטה

### 2.7.3 שימוש

הרשת פועלת במספר שלבים עיקריים.

1. קביעת מספר שכבות ומספר נוירונים בכל שכבה, אתחול וקטור המשקלות לערכים אקראיים ובחירת פונקציית אקטיבציה רצויה. במקרה שלנו נבחר בפונקציית אקטיבציה סיגמואידית.

$$g(z) = \frac{1}{1 + e^{-z}}$$

2. הזנת וקטור קלט והזרמתו ברשת. נקרא גם Feed Forward.

למשל הנוירון  $a_1^{(2)}$ , מקבל את ערכו באופן הבא:

$$a_1^{(2)} = g(\theta_{10}^{(1)} x_0^{(1)} + \theta_{11}^{(1)} x_1^{(1)} + x_1^{(1)} \theta_{12}^{(1)}) = g(z_1^{(2)})$$

פעולה זו מתבצעת עבור כל נוירון בכל שכבה עד השכבה האחרונה לפי סדר ההזרמה.

3. הערך הגדול ביותר מבין הנוירונים בשכבת הפלט, מייצג את ה"החלטה" של הרשת על הקלט.

### 2.7.4 פונקציית מחיר

פונקציית המחיר מוכללת כולל רגולריזציה עבור רשת עצבית זו היא:

$$J(\Theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \sum_{k=1}^K y_k^{(i)} \log(h_{\Theta}(x^{(i)}))_k + (1 - y_k^{(i)}) \log(1 - (h_{\Theta}(x^{(i)}))_k) \right] + \frac{\lambda}{2m} \sum_{l=1}^{L-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (\Theta_{ji}^{(l)})^2$$

$m$  – מספר דוגמאות הקלט.

$K$  – מספר הנוירונים בשכבת הפלט.

$y_k$  – וקטור תיוג המשתמש למחלקה  $k$  באמצעותו הרשת יודעת אם הניחוש היה נכון או לא.

מזער  $J(\Theta)$  לפי הנגזרת:

$$\frac{\partial}{\partial \theta_{ij}^l} J(\Theta) , \quad \theta_{ij}^l \in R$$

כדי למזער את  $J(\Theta)$  על-ידי שינוי הפרמטרים, נחשב עבור כל יחידה את הנגזרת החלקית הבאה:

$$\frac{\partial}{\partial \theta_{ij}^{l-1}}$$

$J(\Theta)$  מושפע עלי ידי  $\theta_{ij}^{l-1}$  דרך המשתנה  $z_i^{(l)}$  שהוא סכום הקלטים ליחידה ה- $i$ :

$$z_i^{(l)} = \sum_{j=0}^{S_{l-1}} \theta_{ij}^{l-1} a_j^{l-1}$$

ולכן לפי כלל השרשרת:

$$\frac{\partial}{\partial \theta_{ij}^{l-1}} J(\theta) = \frac{\partial J(\theta)}{\partial z_i^{(l)}} * \frac{\partial z_i^{(l)}}{\partial \theta_{ij}^{l-1}}$$

כעת נגדיר:

$$\delta_i^{(l)} = \frac{\partial J(\theta)}{\partial z_i^{(l)}}$$

הנגזרת החלקית אם כן תחושב כך:

$$\frac{\partial}{\partial \theta_{ij}^{l-1}} J(\theta) = \delta_i^{(l)} * a_j^{(l-1)}$$

בדרך כלל ה  $\delta$ -ות נקראות גם השגיאות. חישוב ה  $\delta$  – ות של שכבת הפלט:

$$\delta_i^{(L)} = \frac{\partial J(\theta)}{\partial z_i^{(L)}} = \dots = a_i^{(L)} - y_i$$

לאחר פיתוח, במקרה הכללי יתקבל:

$$\delta_i^{(l)} = g'(z_i^{(l)}) \sum_{k=1}^{S_{l+1}} \theta_{ki}^l \delta_k^{(l+1)}, \quad g'(z_i^{(l)}) = a_i^{(l)} * (1 - a_i^{(l)})$$

כדי לקבל את ה  $\delta$  – ות של השכבות החבויות, נצטרך להזרים אחורנית את ה  $\delta$  – ות החל משכבת הפלט. הערה: לשכבת הקלט אין ערכי  $\delta$ .

### Back propagation 2.7.5

נעשה שימוש בעקרון ה Gradient decent על מנת למזער את  $J(\theta)$ .

חישוב הנגזרות החלקיות שתואר לעיל מאפשר לדעת את כיוון הירידה התלולה ביותר כלפי מטה של  $J(\theta)$ . לכן בכל שלב נעדכן את הפרמטרים בצעדים הדרגתיים בעזרת מקדם הלמידה  $\alpha$  לכיוון שהתקבל, אל עבר מינימום מקומי של  $J(\theta)$ .

ערך העדכון שמתווסף לכל פרמטר  $\theta_{ij}^{(l)}$  נקרא הגרדיאנט ומוגדר כך:

$$\Delta \theta_{ij}^{(l)} = -\alpha \frac{\partial J(\theta)}{\partial \theta_{ij}^{(l)}}$$

לכן, כלל העדכון לכל פרמטר בכל איטרציה יהיה מהצורה:

$$\theta_{ij}^{(l)} = \theta_{ij}^{(l)} + \Delta \theta_{ij}^{(l)}$$

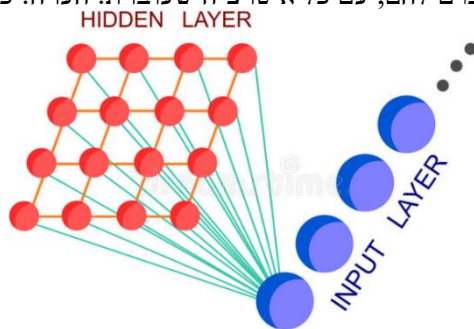
## 2.8 רשת קוהונן – Self Organizing Map (SOM)

### 2.8.1 הקדמה

רשת קוהונן הידועה גם כ  $SOM$  (Self Organizing Graph), היא מודל של רשת עצבית מלאכותית המסתמך על למידה לא מודרכת להצגה ויזואלית של אוסף נתונים רציפים ממימד גבוה במימד נמוך ובדיד – 2 או 3. ניתן להשתמש במודל זה גם לביצוע קלאסטינג כאשר כל קבוצה נפרדת במפת התוצאה מייצגת קלאסטר (ראה איור 8). את המודל ניסח טאוכו קוהונן (להוסיף reference).

### 2.8.2 הרעיון הכללי

עבור אוסף נתונים בעל  $m$  נקודות ממימד  $n$ , מפת התוצאה מורכבת מסריג ( $grid$ ), או "מפה", בעלת מימדים  $M \times N$ , שבה כל יחידה מותאמת לנוירון ממימד  $n$ . שכבת הכניסה היא ממימד  $n$  וכל יחידה בה מחוברת לכל נוירון במפה על ידי פרמטר משקל. תחילה, כל המשקלות, כלומר הערכים המגדירים את הנוירונים מאותחלים לערכים אקראיים. בכל איטרציה תחילה מוזנת נקודה אקראית מאוסף הנתונים לרשת. לאחר מכן, נמצא הנוירון הכי קרוב לנקודה זו ומשקלותיו מותאמות בכיוון נקודת הקלט. לרוב נעשה שימוש במרחק אוקלידי לביצוע חישוב זה, ואכן בזה השתמשנו. כעת במפת התוצאה נצבע את האיבר שמתאים לנוירון שהתקבל בצבע מוגדר מראש שמתאים למחלקת נקודת הקלט. לאחר מכן, מתבצע עדכון של משקלות הנוירונים השכנים של הנוירון הנבחר גם כן לכיוון נקודת הקלט, אך בשיעור פחות, הנגזר מקרבת היחידות במפת ההצגה המאפיינות נוירונים אלה. דבר זה יוביל לכך שעם אימון הרשת, יוצרו איזורים או "שכונות" על מפת ההצגה (איור 8) כך שלנקודות דומות מאוסף הנתונים יהיה סיכוי הולך וגדל להיות ממופים לאיזורים המתאימים להם, עם כל איטרציה שעוברת. הערה: ככל שהרשת רצה, עוצמת עדכון המשקלות תדעך.



באיור הבא ניתן לראות את שכבת הכניסה, שבנויה מ  $n$  נוירונים וגריד הנוירונים עצמו ( $Hidden layer$ ). באיור זה מוצגים הקשרים בין הנוירון הראשון בלבד, אך בפועל כל נוירון בשכבת הכניסה מחובר באותו אופן לכל הנוירונים בגריד.

איור 6: המחשה לקשר בין שכבת הכניסה לגריד הנוירונים

### 2.8.3 סימונים והגדרות

- נגדיר תחילה מספר פרמטרים.
- $T$  - מספר מקסימלי של איטרציות
- $\alpha_0$  - קצב הלמידה ההתחלתי
- $\alpha_t$  - קצב הלמידה בזמן  $t$
- $\sigma_0$  - פרמטר התחלתי להגדרת פונקציית השכנות ההתחלתי
- $\sigma_t$  - פרמטר להגדרת פונקציית השכנות בזמן  $t$
- $\mu_\alpha$  - מגדיר את קצב דעיכת  $\alpha$

$\mu_\sigma$  - מגדיר את קצב דעיכת  $\sigma$   
 $G(\Phi_m)$  - מסמל את היחידה מהמפה שמתאימה לנוירון  $\Phi_m$   
 בנוסף, לכל מחלקה ממחלקות הקלט נגדיר צבע ייחודי

#### 2.8.4 האלגוריתם

כל עוד מספר האיטרציות  $t$  קטן ממספר המקסימלי של האיטרציות נבצע:

1. נבחר נקודה אקראית  $x_i$  מאוסף הנתונים. כעת יש למצוא נוירון  $\Phi_m$  שמרחקו מ-  $x_i$  מינימלי ביחס לשאר הנוירונים ברשת.  

$$\{\Phi_m^t | d(\Phi_m^t, x_i^t) \leq d(\Phi_j^t, x_i^t) \forall j = 1, 2, \dots, MN\}$$
2.  $\Phi_m^t$  מכונה בתור "הנוירון הנבחר" כינוי מקביל ל- *Best Matching Unit*, או *BMU* בקיצור. אפשר להגיד שהנוירון הנבחר הוא הכי "דומה" לנקודת הקלט. מכאן שאם הנוירון הנבחר יעודכן להיות מעט דומה יותר לנקודת הקלט (או קרוב יותר), הסיכוי שנקודות שדומות לנקודת הקלט ימופו לנוירון זה גדל. מאותה סיבה מעדכנים גם את שכניו, רק בשיעור פחות - לפי פונקציית השכנות. במפת התוצאה, נצבע את היחידה שמייצגת את  $\Phi_m^t$  בצבע שהוגדר למחלקה שלו, ונציגה. (איור 7)
3. לכל נוירון  $\Phi_j^t, j = 1, 2, \dots, MN$  יותאם ערך שכנות  $N(\Phi_j^t)$  ביחס לכמה שהיחידה  $G(\Phi_j^t)$  קרובה ליחידה  $G(\Phi_m^t)$  שמייצגת את הנוירון הנבחר. זוהי למעשה פונקציית השכנות.

בהצגה וקטורית:

$$N(\Phi^t) = e^{\frac{-d(G(\Phi_m^t), G(\Phi^t))^2}{2\sigma^2}}, \quad \Phi^t = (\Phi_1^t, \Phi_2^t, \dots, \Phi_j^t)$$

4. וקטור ההפרשים בין נקודת הקלט  $x_i$  לכלל הנוירונים:

$$d(x_i, \Phi) = (d(x_i^t, \Phi_1^t), d(x_i^t, \Phi_1^t), \dots, d(x_i^t, \Phi_j^t)) \quad .i$$

5. כעת יתבצע כלל העדכון מהצורה:

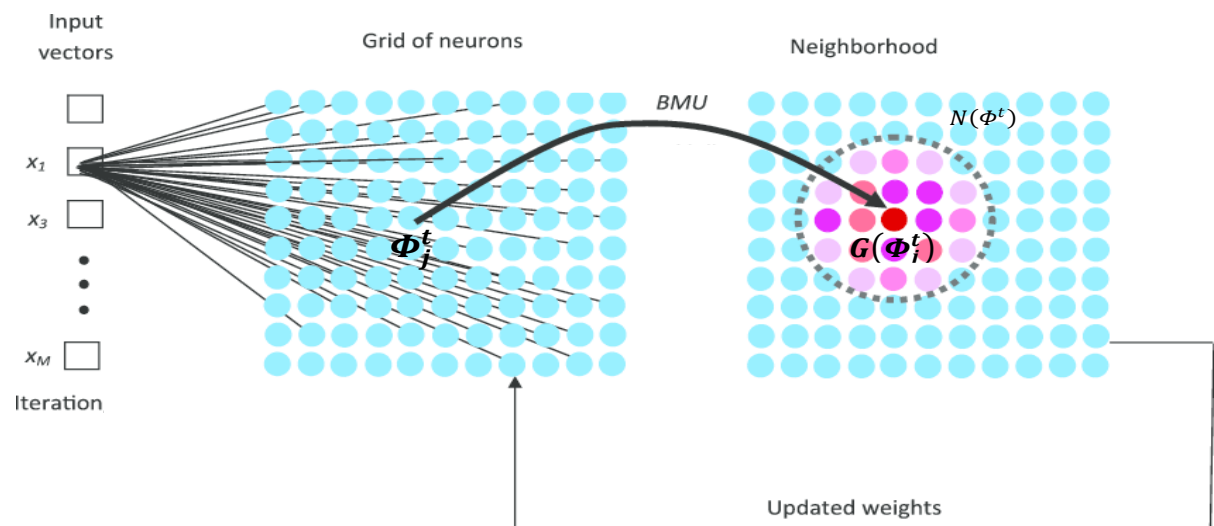
$$\Phi^{t+1} = \Phi^t + N(\Phi^t) \alpha_t d(x_i, \Phi)$$

6. מקדם הלמידה דועך בכל איטרציה באופן הבא:

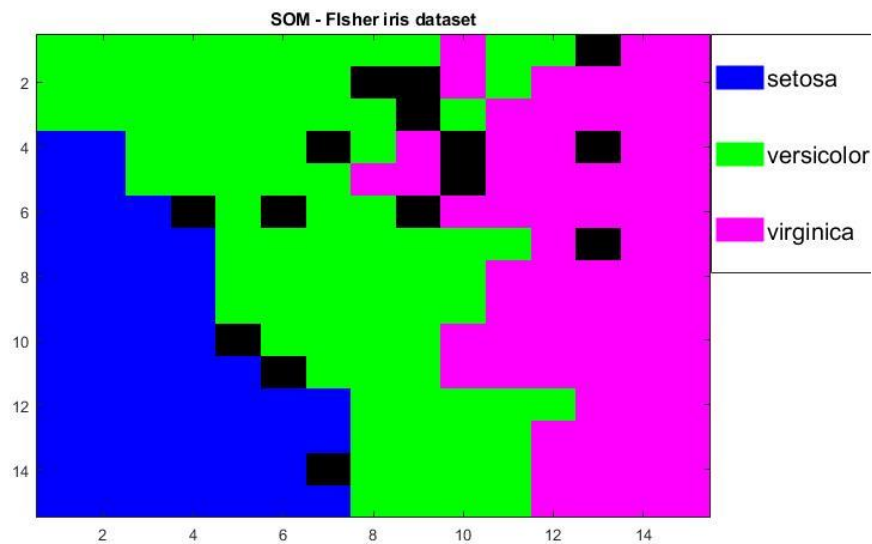
$$\alpha_{t+1} = \alpha_0 e^{-\left(\frac{t}{\mu_\alpha}\right)} \quad .a$$

7. פרמטר להגדרת פונקציית השכנות דועך בכל איטרציה באופן הבא:

$$\sigma^{t+1} = \sigma_0 e^{-\left(\frac{t}{\mu_\sigma}\right)}$$



איור 7: (משמאל לימין) המחשת שכבת הכניסה, גריד הנוירונים והגריד שממנו נגזרת פונקציית השכנות. (Som)



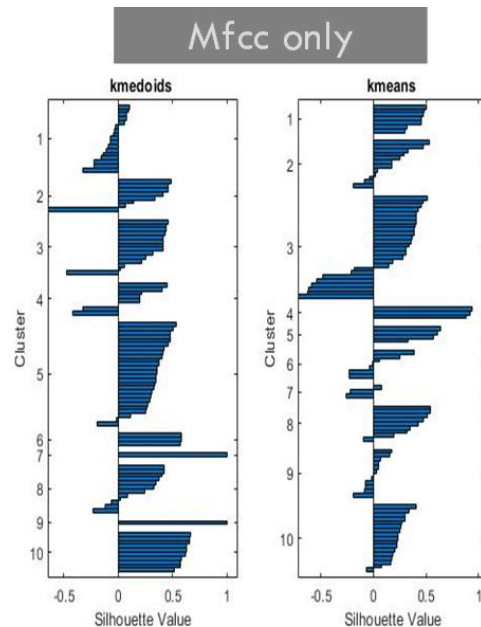
איור 8: תוצאת הרצת SOM על סט הנתונים fisheriri



## פרק 3: תוצאות

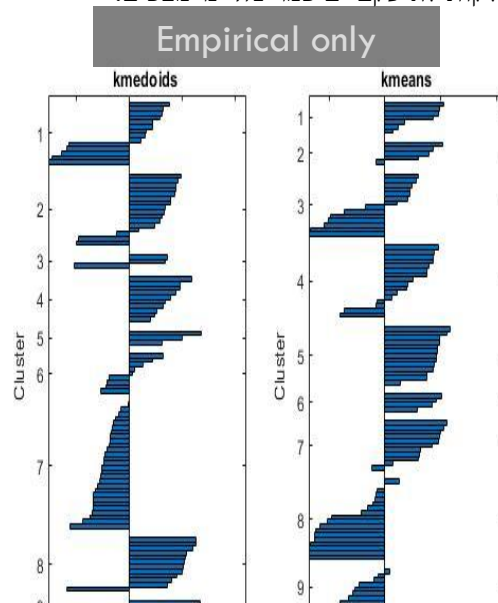
### 3.1 תוצאות – Clustering

1. באיור 9 מופיעים ערכי הסילואט עבור תוצאות הרצת האלגוריתמים *K-Medoids & K-means* על סט הנתונים האמפיריים בלבד. ניתן לראות שקיימים ערכים נמוכים יחסית, וגדלי המחלקות לא אחידים כמו שהיינו מצפים.



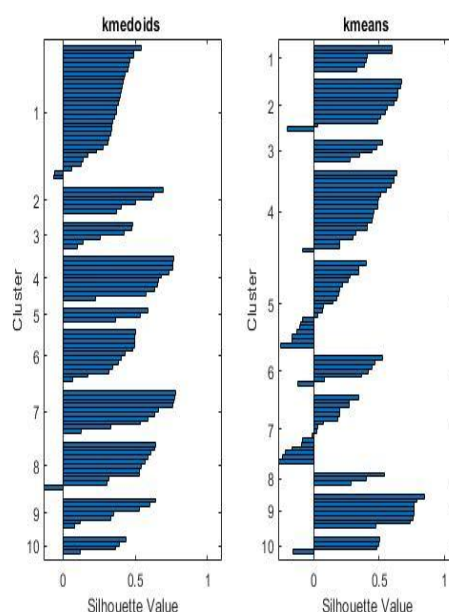
איור 9: תוצאות ערכי סילואט עבור קלאסטרינג של הנתונים התדריים בלבד.

2. באיור 10, מופיעים ערכי סילואט עבור הנתונים האמפיריים שנמדדו על ידי הביולוגים. גם כאן ישנם ערכים שליליים וגדלי המחלקות לא עקביים כמו שהיינו מצפים.



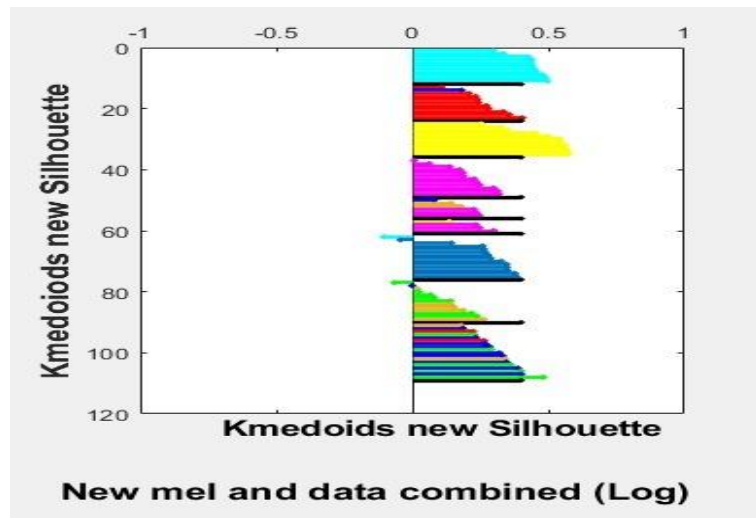
איור 10: תוצאות ערכי סילואט עבור קלאסטרिंग של הנתונים האמפיריים בלבד.  
 3. באיור 11 מופיעים ערכי סילואט עבור שילוב הנתונים האמפיריים עם הנתונים התדריים. ניתן לראות שיפור ניכר. ערכי הסילואט ככלל גבוהים יותר ואף גדלי המחלקות השונות עקביים יותר כמו שאנו מצפים.

### Mfcc & empirical



איור 11: תוצאות ערכי סילואט עבור קלאסטרिंग של שילוב הנתונים האמפיריים עם הנתונים התדריים.

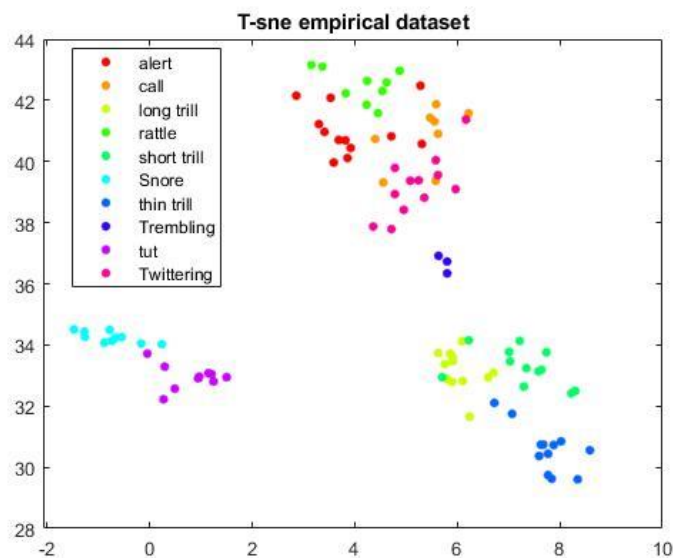
4. כזכור, אחת מהמטרות המרכזיות של הפרויקט היא לנסות ולתקף את החלוקה שהצוות הביולוגי ביצע, באמצעות אלגוריתמים אובייקטיביים. באיור 12, ניתן לראות את ערכי הסילואט עבור כל וקטור תכונות שמיצינו. כל מקטע צבוע בצבע מסוים שמייצג שייכות למחלקה כלשהי לפי סיווג הביולוגים. הקווים השחורים מייצגים את הגבול בין סיווגי המחלקות של אלגוריתם הקלאסטרिंग. במקרה זה *k-medoids*. ניתן לראות שבכ-5 המחלקות הבאות לפי סדר הופעת הצבעים באיור 12 מלמעלה למטה: *Long-trill, Rattle, Snore, Short-trill, Thin-trill* ישנה התאמה בין חלוקת הביולוגים לבין חלוקת הקלאסטרिंग, כלומר שבמידה מסוימת הושגה מטרה זו!



איור 12: תוצאת הקלאסטרינג (מחלקות מופרדות בפס שחור) בהשוואה לחלוקת הביולוגים (מיוצגת בצבעים).

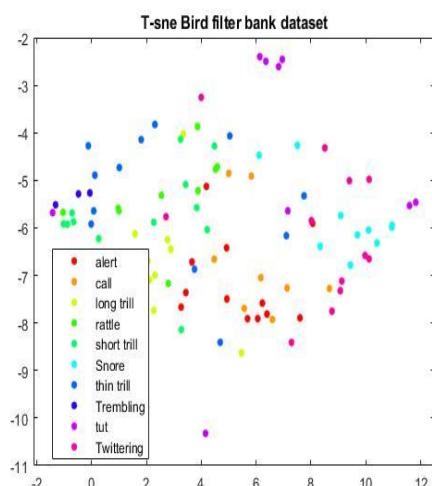
### 3.2 תוצאות T-SNE

1. באיור 13 מוצגת תוצאת הרצת t-SNE על הנתונים האמפיריים בלבד.
    - המחלקות Short-trill, Thin-trill, long-trill מופרדות יחסית אחת מהשניה ומשאר המחלקות.
    - המחלקות Rattle, Alert מופרדות במידה מסוימת.
    - המחלקות Call, Twittering מעט מעורבבות אחת עם השניה.
    - המחלקות tut, Snore מופרדות במידה טובה בין עצמן ובין שאר המחלקות.
- ניתן לשים לב שקיימים מקבצים של מחלקות, כך שניתן להסיק שיש מחלקות דומות יותר ודומות פחות. דוגמה מובהקת שגם מתיישבת עם המציאות היא ששלושת קריאות ה trill ממוקמות במקבץ נפרד.

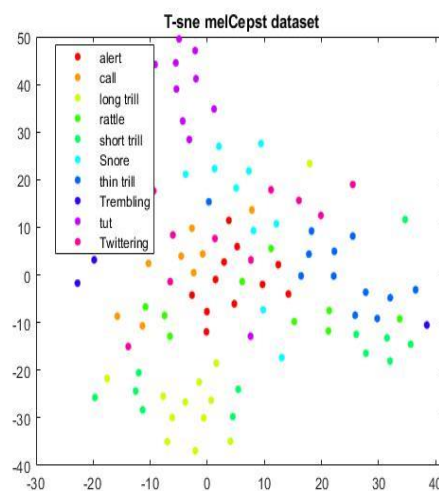


### איור 13: תוצאת $t$ -SNE

Bir



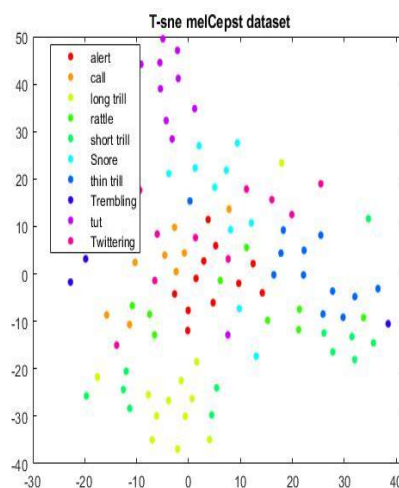
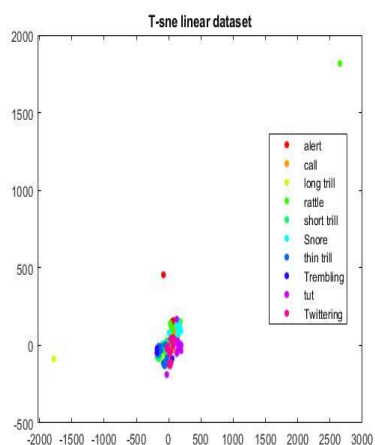
$t$ -SNE על התכונות  
מפרדה משמעותית



איור 15:  $t$ -SNE על תכונות תדירות (Bird)

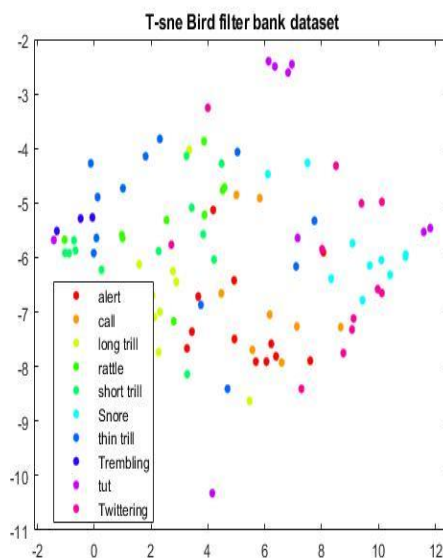
איור 14:  $t$ -SNE על תכונות תדירות (MelCepst)

3. באיורים 16, 17, 18 מוצגות תוצאות הרצת  $t$ -SNE מוצאות הרצת  $t$ -SNE על הנתונים שהתקבלו משילוב הנתונים התדריים בלבד, עבור הסקלות: MFCC, Linear, Bird filter bank



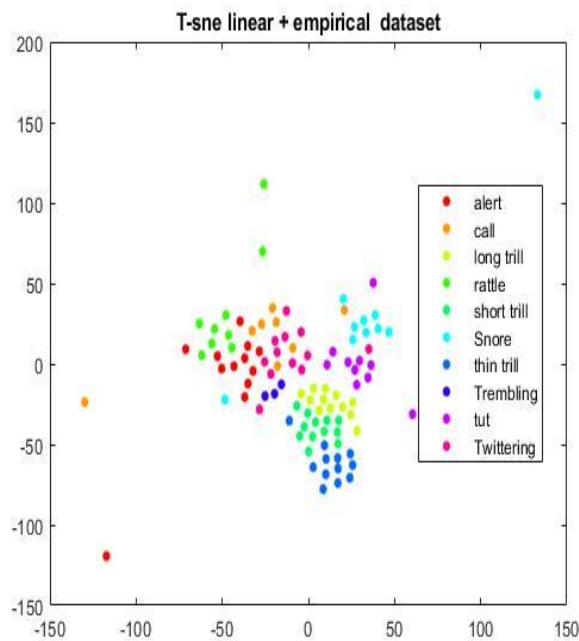
איור 17: סקלה ליניארית

איור 16: סקלת melCepst

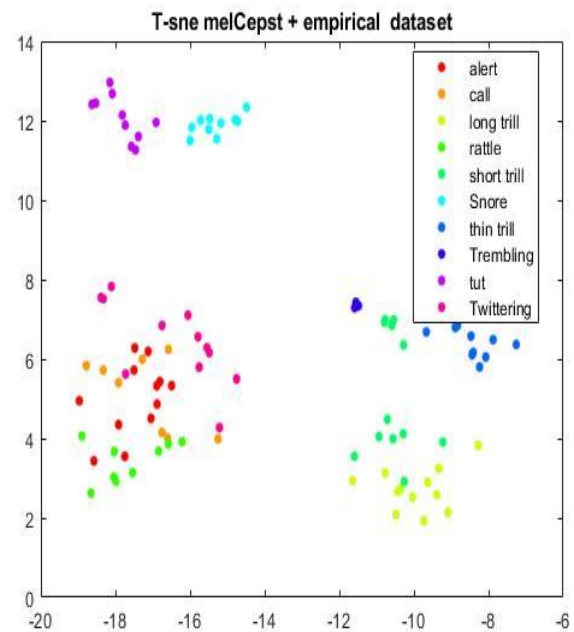


איור 18: סקלת Bird

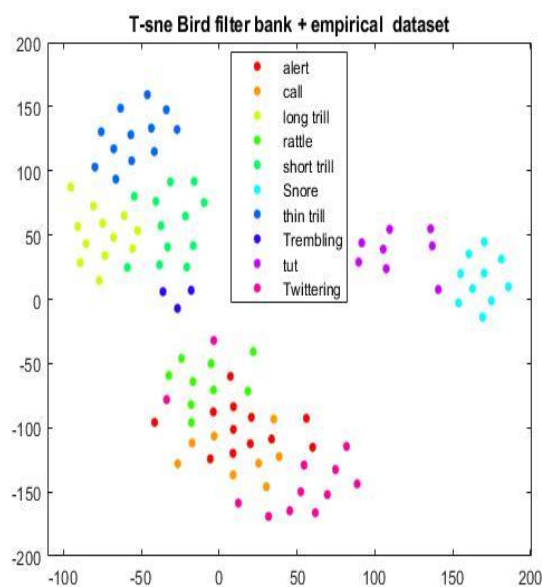
4. באיורים 19,20,21 מוצגות תוצאות הרצת t-SNE תוצאות הרצת t-SNE על הנתונים שהתקבלו משילוב MFCC , Linear, Bird filter bank: עבור הסקלות: MFCC , Linear, Bird filter bank



איור 20: סקלה ליניארית ונתונים אמפיריים



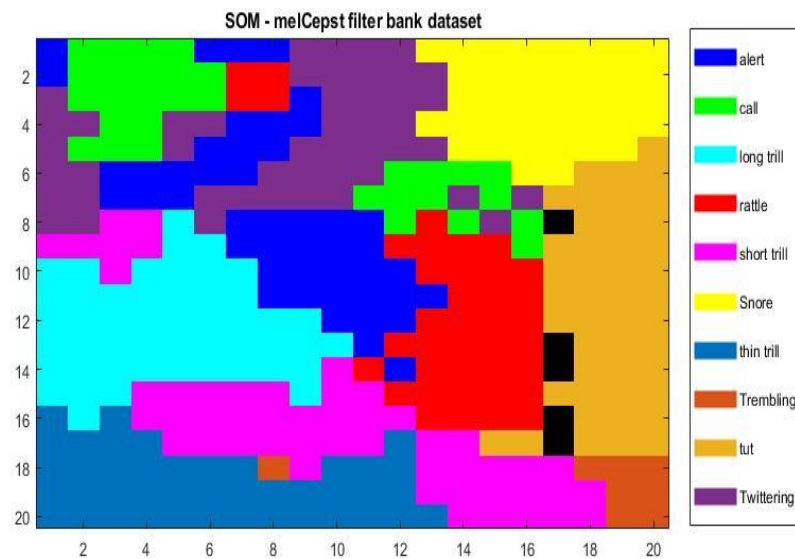
איור 19: סקלת melCepst ונתונים אמפיריים



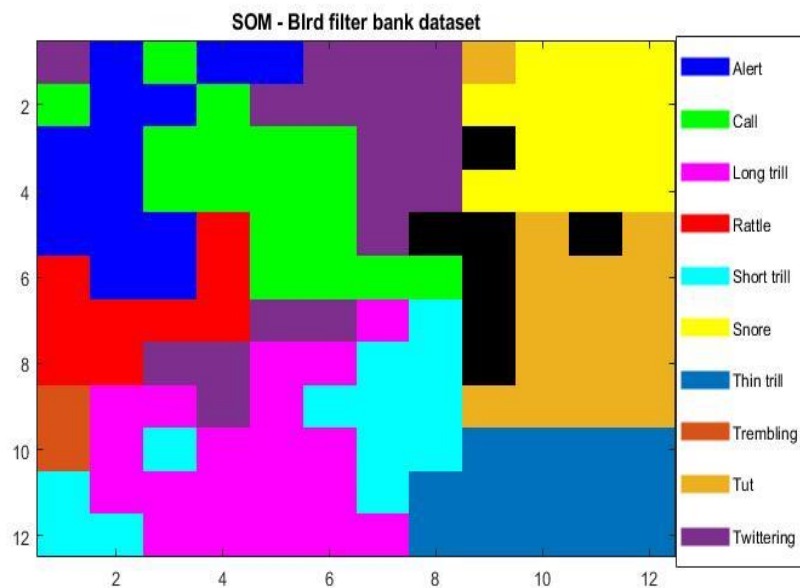
איור 21: סקלת Bird ונתונים אמפיריים

### 3.3 תוצאות – SOM

1. באיורים 22 ו 23 ניתן לראות את תוצאות הפעלת SOM על סט התכונות התדרייות לפי סקאלת Bird ו-Mel בהתאמה. בשני האיורים המשטח מחולק בצורה טובה לאיזורים בעלי צבעים אחידים. מכאן שיש הבדלים ברי מדידה ומשמעותיים בין נתוני המחלקות השונות שהתקבלו לאחר שלב מיצוי הפרמטרים.



איור 22: הרצת Som על שילוב הנתונים האמפיריים על הנתונים התדריים לפי סקלת melCepst



איור 23: הרצת Som על שילוב הנתונים האמפיריים על הנתונים התדריים לפי סקלת Bird

### 3.4 תוצאות רשת נוירונים

מיצוי התוצאות מרשת הנוירונים נעשה בצורה של *Cross-Validation*, מאחר וסט הנתונים הבסיסי אינו מאוד גדול. באיור 24 אפשר לראות את ה- *Confusion-Matrix* שמייצגת את ביצועי הרשת, והצלחת ביצועי הרשת עבור הקריאות השונות. לאחר מספר הרצות אימון עם פרמוטציות שונות של נתוני אימון ונתוני בדיקה שנגזרו מאוסף הנתונים הכולל, ממוצע אחוזי ההצלחה עומד על 85.93% וסטיית התקן עומדת על 4.71%.

	Alert	Call	Long trill	Rattle	Short trill	Snore	Thin trill	Trembling	Tut	Twittering
Alert	0.72	0.194	0	0.05	0	0	0	0	0	0.02
Call	0.2	0.53	0	0	0	0	0	0	0	0.26
Long thrill	0	0	1	0	0	0	0	0	0	0
Rattle	0.12	0	0	0.82	0	0	0	0	0	0.05
Short thrill	0	0	0.02	0	0.85	0	0	0.12	0	0
Snore	0	0	0	0	0	0.91	0	0	0.09	0
Thin thrill	0	0	0	0	0	0	1	0	0	0
Trembling	0	0	0	0	0	0	0	1	0	0
Tut	0	0	0	0	0	0	0	0	1	0
Twittering	0.15	0.11	0	0	0	0	0	0	0	0.73

איור 24: *Confusion Matrix* עבור רשת הנוירונים.

### 3.5 סיכום

1. מימוש אלגוריתמי הלמידה, הסיווג והוויזואליזציה התבצע בהצלחה.

האלגוריתמים שמומשו עבור פרויקט זה הם:

a. *Linear, Logarithmic, Bird – MFB*

b. *K-Means*

c. *K-Medoids*

d. *Rand-Index*

e. *Silhouette*

f. *ANN*

g. *SOM*

2. האלגוריתמים שנלקחו מספריות:

a. *MATLAB – FFT*

b. *VOICEBOX – MFCC*

3. תיקוף חלוקת הביולוגים התבצע בהצלחה יחסית בעזרת קלאסטרינג עבור המחלקות:

*Long-trill .a*

*Short-trill .b*

*Thin-trill .c*

*Rattle .d*

*Snore .e*

4. רשת עצבית מאומנת עם נתוני אימון שהגיעו מחלוקת הביולוגים, הצליחה לזהות בכ-85.93% בממוצע את הקולות השונים (איור 24).

5. הצלחה בהצגת המחלקות והקשרים ביניהם בצורה ויזואלית וברורה לעין בעזרת SOM וכן בעזרת אלגוריתם סילואט מותאם, שהוסיף סימון צבע לכל ערך סילואט שמייצג את חלוקת הביולוגים, ומאפשר לראות את הקשר בין חלוקת הביולוגים לחלוקת הקלאסטרינג (איור 12).

## ביבליוגרפיה

Hinton, L. v. (2008). Visualizing High-Dimensional Data Using t-SNE. 9(Nov):. *Journal of Machine Learning Research* , 2579-2605, .

SOM. (n.d.). *Som*. [https://www.researchgate.net/figure/Self-organizing-map-learning-process-The-2-dimensional-grid-of-neurons-is-characterized\\_fig2\\_257550049](https://www.researchgate.net/figure/Self-organizing-map-learning-process-The-2-dimensional-grid-of-neurons-is-characterized_fig2_257550049).

tsne. (n.d.). *T-Sne image*. [https://www.researchgate.net/figure/tSNE-representation-of-MNIST-dataset-28x28-grayscale-handwritten-digits-preprocessed\\_fig17\\_319151468](https://www.researchgate.net/figure/tSNE-representation-of-MNIST-dataset-28x28-grayscale-handwritten-digits-preprocessed_fig17_319151468).

van der Maaten and Hinton, 2. (n.d.). *van der Maaten and Hinton, 2008*.

yardbirds.org. (n.d.). [https://www.yardbirds.org.il/show\\_bird/77](https://www.yardbirds.org.il/show_bird/77).

שלדג. (n.d.). ויקיפדיה.