

חלק א:

Word Pairs and Distances:

1. fast - rapid : 0.61314803
2. fast - slow : 0.87440497
3. tale - story : 0.8427329
4. real - actual : 0.6286853
5. short - long : 0.8872582
6. silent - quiet : 0.6172162
7. help - support : 0.7755802
8. reply - answer : 0.6652572
9. help - support : 0.7755802
10. pear - apple : 0.5434469

עבור כל צמד המילים ציפיתי לראות דמיון גבוה יחסית. אמנם כאשר מדובר במילים 'מנוגדות' כמו fast-slow אשר מהוות שמות תואר המעידים שמשהו הוא או מהר או איטי וכו' ניתן לצפות כי ישנו דמיון רב. אך אני מניח שקיבלתי דמיון לא כזה גבוה ולא אחוז יותר גבוה כי וקטורי המילים של fast יכולו מילים (שמות תואר בעיקר) שיתארו מקרים מהירים, ובאותו האופן וקטורי המילים של slow יכולו מילים המייצגות ומתארות אירועים איטיים, כלומר, יש הבדלים בקונטציות של המילים שגורמים לכך שהדמיון לא גבוה יותר.

אזי כאשר מדובר במילים מנוגדות כן נצפה לדמיון רב כמו הדוגמאות שהראיתי אך זה מאוד תלוי בקונטקסט של וקטורי המילים. במקרה של fast-slow קיבלנו דמיון רב.

ההפתעה היותר גדולה הייתה בדמיון של מילים נרדפות. ניתן לראות אחוזים יחסית נמוכים (60 אחוז בין silent-quiet) ועוד.. על פניו אנחנו מצפים שמילים אלו יקבלו את האחוזים הכי גבוהים אך נשים לב שבקונטקסט של כל מילה במיוחד בשפה האנגלית כל מילה יכולה להיכלל למשמעויות שונות בתחליט ולכן נקבל רק 60 אחוז כי וקטורי המילים לעיתים קרובות מתייחסים למשמעויות שונות.

ההסברים הללו מייצגים את כלל המילים שבחרתי בין אם במילים מנוגדות או במילים נרדפות ועוד...

חלק ב:

```
told baby, told baby, told baby
What you done? (What you done?)
Where 'll at? (Where 'll at?)
Oh, know got plans? (know got plans?)
Don't believe that (shut your trap)
I'm sipped wine (sip, sip) in a robe (drip, drip)
I look takeo good (look takeo good) take be alone (woo-woo)
My house putting (house putting), my pool warm (pool warm)
Just trimmed (smooth like a newborn)
We must be dancing, romancing
In the east wings and the west wings
Of it mansion, what's happening?
I ain't played no games
Every word that I say is coming straight in the heart
So if you tryna lay in these weapons
I'ma stay the door open
(I'ma leave the doors open)
I'ma leave the door set, girl
(I'ma leave which door open, hoping)
That you feeling the way I feeling
And you want me like I want you tonight, babies
(Tell i that you're coming through)
Ooh, you're so sweet (so sweet), so sideline (so sideline)
I won't bites (ah-ah), unless you like (unless you like)
If you smoke (what you smoke?) I got the smog (Purple smog)
And if you're hungry, girl, I getting filets (woo-woo)
Ooh, baby, don't keep i (waiting)
There's so much love sure could be making (shamone)
I'm talking hugging, cuddling
Rose sepals in the bathtub
Girl, let's jumps in, it's bubbling
I ain't playing no game
Every word that I say is coming tied from the heart
So (if you) if you (tryna) tryna lay in these weapons
I'ma leave the doors open
(I'ma stay the door open)
I'ma stay the door open, girl
```

```
(I'ma leave the doors open, hoping)
That you feeling the way I feeling
And you want i like I want you tonight, baby
(Tell me that you're coming through) go on, girl
La-la-la-la, la-la-la-la (I need you, babies)
La-la-la-la, la-la-la-la (I gotta come you, baby)
La-la-la-la, la-la-la-la (girl, I'm tryna take you this)
nee-nee-nee-nee, nee-nee
Hey-hey, I'ma leave my doors open, baby (I'ma leave the doors open)
I'ma stay, I'ma stay my door open, girl
(I'ma leave the door set, hoping) and I'm hoping, hoping
That you feel the so I feel
And you want me like I want you night, baby
(Tell me that you're come through) woo!
La-la-la-la, la-la-la-la (tell i)
(know me that you're coming through)
whahng-whahng-whahng-whahng, whahng-whahng-whahng, whahng-whahng-whahng
whahng-whahng-whahng-whahng, whahng-whahng-whahng, whahng-whahng
de-de-de, de-de-de-de (de-de-de, de-de)
(Tell me that you're come through)
Girl, I'm here just wait for you (oh!)
Come on while, I'll adore you (I gotta know!)
La-la-la-la, la-la-la-la (I'm wait, wait, wait)
(know me that you're coming through) for you
Girl, I'm here just waiting making you
Come on over, I'll adore 'll
de-de-de, de-de-de-de (de-de, de-de)
```

ניתן לראות כי מבחינה תחבירית רוב המילים הוחלפו בצורה מוצלחת אך כמו שראינו בתרגיל בית 2, המילים אכן טובות מבחינה תחבירית אך לא מבחינה של שירים ושל מוסיקה. נסתכל ונראה שהחרוזים בשיר החדש פחות מוצלחים, בנוסף הקשר בין שורה אחד לשניה הוא פחות קוהרנטי ביחס לשיר המקורי.

בכל זאת ברוב השיר המילים אשר הוכנסו לרוב היו קוהרנטיות אך נשים לב כי המילים הפחות קוהרנטיות הן בסדר מבחינה תחבירית אך פחות מבחינה סמנטית.

חלק 2:

ראשית כל הגרפים מוצגים בעמוד הבא. (קלאסים = קטגוריות)

וקטור המשקלים שבניתי בהתחשב בכמות המילים של כל קלאסס. נבחין כי בכמות דאטה קטנה כזו של כל קלאסס יהיה קשה להבדיל בין כל קלאסס רק ע"י ספירת המילים בו ולכן עשיתי סוג של אגמנטציה לכל קלאסס כך שהדאטה בו יותר מייצג וכך אפשר באמת לתת משקל לכל מילה אשר מופיע בכל קלאסס ע"י ספירת כמות המילים בכל קלאסס ואיחודם למילון אחד אשר מייצג כל מילה ומשקלה.

בנוסף אם הייתי מתייחס רק ככה אזי לא הייתי מקבל תוצאות טובות כי מילים כמו the חוזרות הרבה פעמים ולא קשורות לאף קטגוריה ולכן השתמשתי גם בחשיבות של כל מילה, מילה אשר הופיעה בהרבה קטגוריות קיבלה חשיבות מופחתת שכאמור בהתחשב בכמות הפעמים שמילה מופיעה בקורפוס מסוים והאם היא חשובה או לא. במידה והמילה חסרת משמעות כמו the אזי קיבלה משקל קטן ואם המילה בעלת חשיבות גדולה אזי קיבלה משקל יותר גדול.

1. באילו גרפים (כלומר, ע"פ איזו פונקציית משקל) ישנה הפרדה מובהקת בין הקטגוריות, אם בכלל?

כמובן שניתן לראות שישנה הפרדה מובהקת בגרף האחרון אשר אני ייצרתי בעצמי את וקטורי המשקלים. למה? כי נתתי משקל יותר גדול לכל מילה הקשורה לקלאסס מסוים וכך יצרתי הפרדה יחסית מובהקת בין כל קטגוריה.

כמובן שאפשר לצפות כי אם ניתן משקל רנדומלי לכל הקטגוריות או משקל זהה אז לא יהיה אפשר להפריד בין הקטגוריות השונות.

2. אילו ציורים הופרדו הכי טוב באופן קונסיסטנטי? נסו לתת השערה מדוע, על סמך הסתכלות בציורים.

המילים שהופרדו הכי טוב הן המילים אשר לא חזרו על עצמם בין הקלאססים השונים או אשר קיבלו מובהקות יותר גדולה. כך כמובן היה קל לזהות את הקטגוריה הרצויה אשר המילים מופיעות אך ורק בקטגוריה זאת.

ולכן קיבלתי מובהקות גדולה בין הקטגוריות בוקטור משקלים שאני יצרתי.

3. מה ניתן להסיק מכך על היכולת לבנות מסווג ע"פ וקטורי מילים?

היכולת לבנות מסווג ע"פ וקטורי מילים היא יכולת חזקה מאוד שאנחנו יכולים להשתמש בה כי המחשב יודע לזהות קטגוריה לפי קבוצה של מילים אשר יותר קשורות אל קטגוריה ספציפית לעיתים יהיה יותר קשה לזהות בין קטגוריות אשר קרובות אחת לשניה אך אם נעבוד עם מספיק דאטה אולי גם משימה זאת תצליח בקלות.

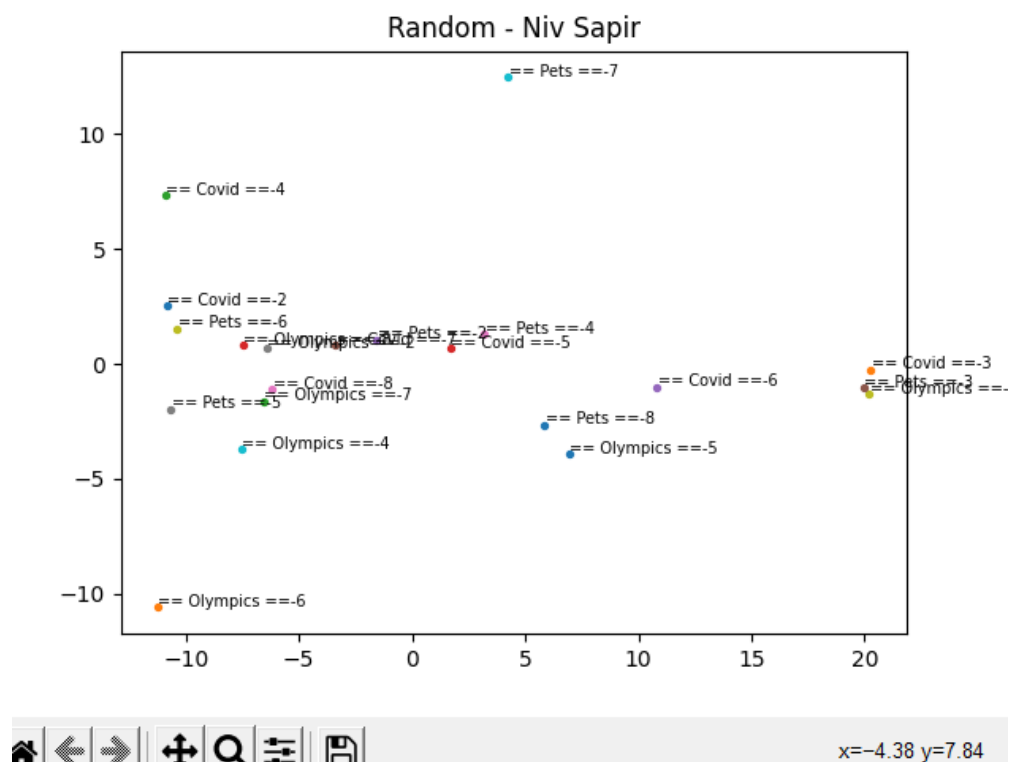
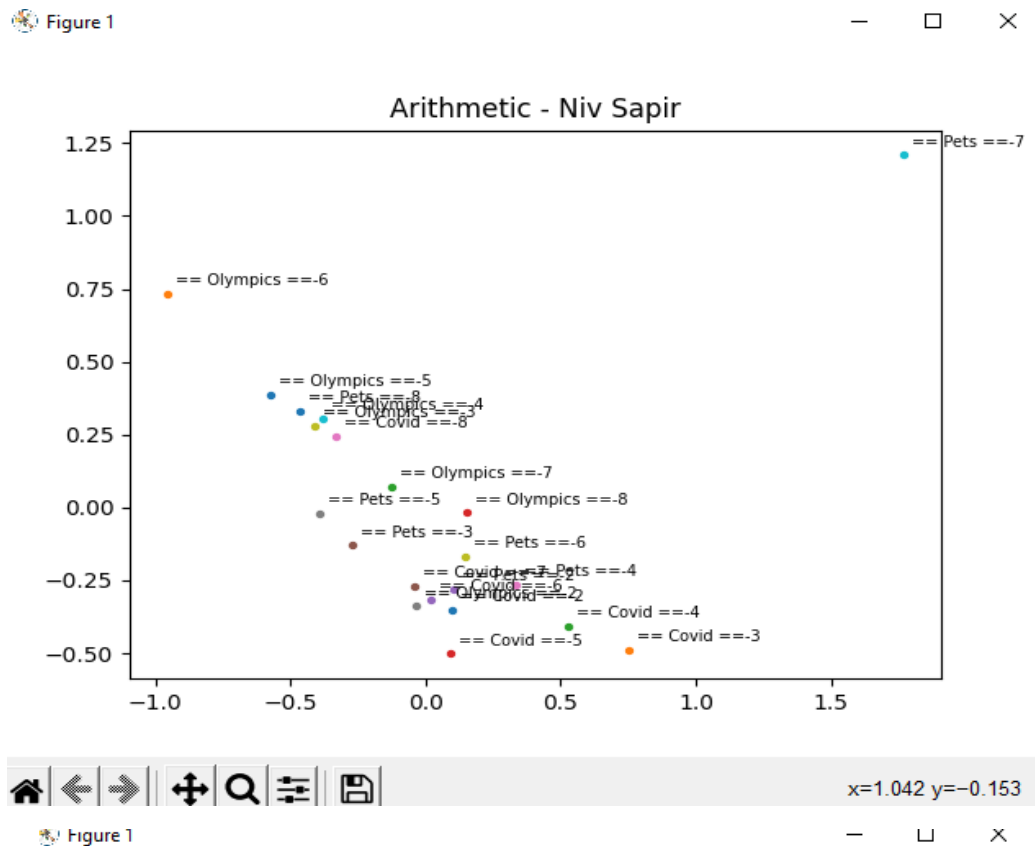
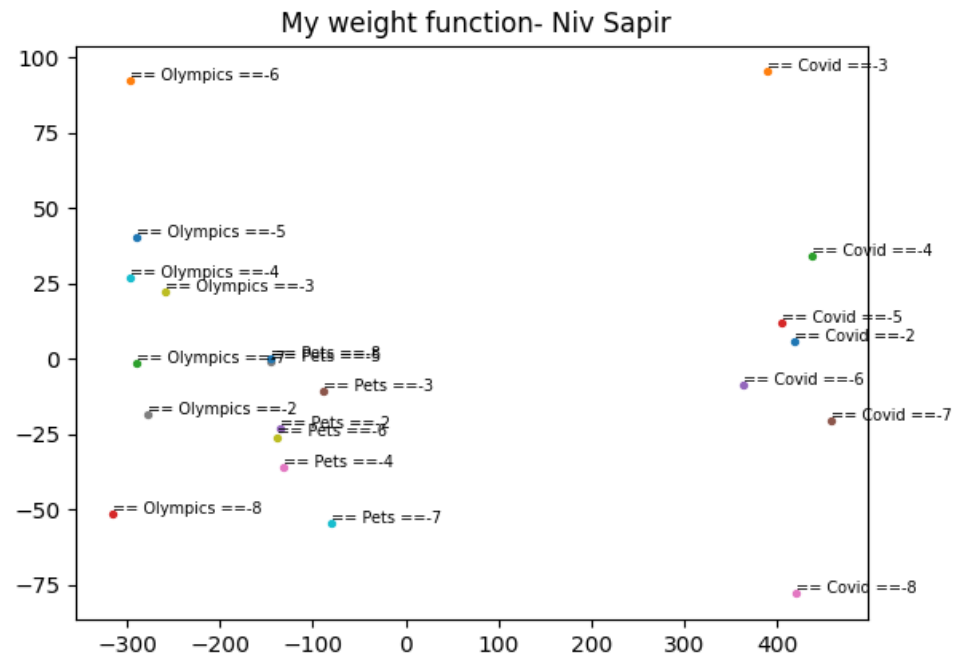


Figure 1

— □ ×



x=-219. y=58.9