

## פרויקט גמר – מערכות המלצה – Movielens 100k

מסמך זה מאגד את כל התוצאות אשר קיבלנו בעת הרצת מחברת ה jupyter אשר מכילה את הקוד של המימושים של הסעיפים השונים בפרויקט זה.

במהלך הפרויקט השוינו את המודלים השונים וקומבינציות של פרמטרים שונים – ההשוואה על טיב המודל נמדדת לפי ה MAE שקיבלנו מקבוצת הבדיקה – .test. לכן במידה ובמסמך רשום MAE ללא התייחסות לקבוצת אימון קבוצת בדיקה – הכוונה היא לקבוצת הבדיקה.

### תרגיל 1:

א. קוד עבור MF – matrix factorization ועבור GMF – General Matrix Factorization מצורף למחברת הקוד.

ב. התבקשנו לחשב את MAE ואת זמן הריצה של קבוצת הבדיקה כאשר אנחנו משנים כל פעם את ממד שכבת ה embedding ואת גודלו של drop-out

עבור MF – ביצענו את הבדיקות עם האפשרויות הבאות (סה"כ 4):

Embedding:

- Item – user
- User – item

Drop-out:

- 0.1
- 0.5

להלן ריכוז התוצאות שקיבלנו עבור MF עבור אימון המודל על  $MAE_{train}$  של קבוצת הבדיקה:

Embedding	Dropout	Mean Absolute Error - train	Time	Mean Absolute Error - test
Item - User	0.5	1.8017	200.0921	1.8032
Item - User	0.1	0.8012	233.1664	0.8021
User - item	0.5	1.7993	206.7634	1.8085
User - item	0.1	0.8022	231.2722	0.8035

### תוצאה הכי טובה עבור MF :

- מבחינת MAE של קבוצת הבדיקה:  
שילוב של  $\text{embedding} = \text{Item} - \text{User}$ ,  $\text{drop-out} = 0.1$
- מבחינת זמנים של אימון המודל: שילוב של  $\text{embedding} = \text{item-user}$ ,  $\text{drop-out} = 0.5$
- אנו נתייחס לתוצאה עם הMAE של קבוצת הבדיקה הכי נמוך כטובה יותר היות ומה שמעניין אותנו שהחיזוי יהיה כמה שיותר טוב.

### עבור GMF – ביצענו את הבדיקות עם האפשרויות הבאות (סה"כ 2):

Embedding:

- User – item

Drop-out:

- 0.1
- 0.5

להלן ריכוז התוצאות שקיבלנו עבור GMF עבור אימון המודל על  $\text{train}$  ו MAE של קבוצת הבדיקה:

Embedding	Dropout	Mean Absolute Error - train	Time	Mean Absolute Error - test
User - item	0.5	0.90155	338.712	0.748837
User - item	0.1	0.9359	72.9278	0.7613

### תוצאה הכי טובה עבור GMF :

- מבחינת MAE של קבוצת הבדיקה :  
שילוב של  $\text{embedding} = \text{user-item}$ ,  $\text{drop-out} = 0.5$
  - מבחינת זמנים של אימון המודל:  
שילוב של  $\text{embedding} = \text{user-item}$ ,  $\text{drop-out} = 0.1$
- אנו נתייחס לתוצאה עם הMAE הכי נמוך כטובה יותר היות ומה שמעניין אותנו שהחיזוי יהיה כמה שיותר טוב.

ג. כעת נשווה בין התוצאות השונות –

ניקח את התוצאה הטובה של MF ו-GMF ונשווה למערכת ההמלצה הלא אישית שכתבנו בעבודה 2 שהייתה מבוססת על ממוצע דירוג הסרטים.

נציין שאנו מתייחסים לתוצאה הטובה ביותר לפי המדידה עם MAE קטן ככל האפשר ולא למדידה בעלת הזמן הנמוך ביותר כיוון שככל שה-MAE קטן יותר הדבר מעיד שהמרחק תוצאות החיזוי של הבדיקה לזאת של האימון הכי קטן – החיזוי הכי טוב.

כמו כן אבחנה נוספת שנוכל לראות היא שמודל (שילוב בפרמטרים) שנתן את MAE עם התוצאה הטובה בזמן אימון המודל הוא גם זה שנתן את התוצאה הטובה של ה-MAE על קבוצת הבדיקה, אך משך זמן האימון והמהירות איננה קשורה.

להלן טבלה אשר מסכמת את התוצאות הטובות ביותר של המודלים השונים:

Model	MAE - train	Time	Mae - test
Matrix Factorization	0.8012	233.1664	0.8021
Generalized Matrix factorization User-item 0.5	0.90155	338.712	0.748837
Non-Personalized (HW2)	לא חושב - בעבודה 2	486.9826	1.0196

נדרג את המודלים לפי רמת הדיוק מהגבוה (MAE נמוך) לנמוך (MAE גבוה) לפי הMAE של קבוצת הבדיקה:  
 GMF.1  
 MF.2  
 non-Personalized.3

נדרג את המודלים לפי רמת הדיוק מהגבוה (MAE נמוך) לנמוך (MAE גבוה) לפי הMAE של קבוצת האימון:  
 MF.1  
 MFG.2  
 non-Personalized.3

נדרג את המודלים לפי משך זמן האימון של המודל (מהקצר לארוך):  
 MF.1  
 MFG.2  
 non-Personalized.3

## תרגיל 2:

א. קוד עבור CFG במחברת jupyter  
ב. חישבנו את ערכי ה MAE וזמן האימון – עבור שילוב של פרמטרים שונים:

**Size of Layer:** 8, 16, 50, 100

**Optimizer** = sgd, adam, Nadam, Adadelata

**Function Loss** = mean absolute error, mean squared error,  
mean\_absolute\_percentage\_error

**Activation Function** = relu, tanh, linear, selu

**drop-out of:** 0.1, 0.5

להלן ריכוז התוצאות השונות עבור כל אחד מהפרמטרים:

### • Size of Layer

נבחין שכל שאנו מגדילים את גודל שכבת הפרויקט ככה גדלה רמת הדיוק של המודל ולכן נריץ את השאר עם 100

Size Of layer	Time	MAE - test
8	46.7975	1.8366
16	42.6713	1.8234
50	82.5716	1.8218
100	70.3447	1.8201

### • Optimizer

נבחין כי כאשר פונקציה האופטימיזציה היא של adam רמת הדיוק של המודל גדלה

Optimizer	Time	MAE - test
sgd	133.4138	1.8201
adam	53.6670	1.818
Nadam	97.4659	1.8374
Adadelata	79.3800	1.8270

### • Function loss

נבחין כי עבור חישוב פונקציית ההפסד : **MAE** רמת הדיוק היא הגבוה ביותר ולכן לשאר הבדיקות נשתמש ב **MAE**

Function loss	Time	MAE - test
mean absolute error	47.21113	1.7840
mean squared error	156.1635	1.8211
mean_absolute_percentage_error	41.9661	1.9701

### • Activation Function

נבחין כי כאשר אנו משתמשים ב **linear** רמת הדיוק היא הגבוה ביותר

Activation Function	MAE-Train	Time	MAE - test
relu	1.7417	64.8764	1.8126
tanh	2.5359	104.3312	2.5383
linear	1.7611	41.9993	1.7660
selu	1.7610	44.3085	1.7997

### • Drop-out

נבחין כי עבור **drop-out = 0.1** רמת הדיוק היא הגבוה ביותר ולכן לשאר הבדיקות נשתמש ב **0.1**

Drop-out	MAE - train	Time	MAE - test
0.1	0.7470	77.6727	0.7967
0.5	1.7712	37.4019	1.8371

הפרמטרים שהניבו את התוצאות הכי טובות (כל אחד בתחמו) בכל אחד מהפרמטרים הם עבור תרגיל 2 מבחינת **MAE** הטוב ביותר של קבוצת הבדיקה:

**Size of Layer**100

**Optimizer** = adam

**Function Loss** = mean absolute error

**Activation Function** = linear

**drop-out of:** 0.1

נריץ פעם נוספת עם השילוב של כל הפרמטרים שהביאו את הכי טוב בתחום שלהם את התוצאה הטובה ביותר ונראה האם זה משפר את התוצאות של אחד מהם:

MAE - train	Time	MAE - test
<b>0.9445</b>	32.6114	0.9512

נשים לב שדווקא השילוב הנ"ל הביא תוצאה פחות טובה מאחת המדידות.

לכן התוצאה הכי טובה הגיע דווקא משילוב הפרמטרים:

**Size of Layer**100

**Optimizer** = nadam

**Function Loss** = mean absolute error

**Activation Function** = relu

**drop-out of:** 0.1

ולכן נמליץ להשתמש בשילוב שלהם שאכן הניב את התוצאה הטובה ביותר שכן ביצענו כאן שלב של עידון ומציאת הפרמטרים אשר יניבו את התוצאה הטובה ביותר עבור במידע עליו אנו מאמנים את המודל.

## ג. לסיכום:

כעת נשווה בין כל המודלים אשר מימשנו עד כה ונראה מי הניב את התוצאות הטובות ביותר:

Model	MAE - train	Time	MAE - test
Matrix Factorization	0.8012	233.1664	0.8021
Generalized Matrix factorization	0.90155	338.712	0.748837
Non-Personalized (HW2)	לא חושב - בעבודה 2	486.9826	1.0196
NCF	0.7470	77.6727	0.7967

נדרג את המודלים לפי רמת הדיוק מהגבוה (MAE נמוך) לנמוך (MAE גבוה) לפי הMAE של קבוצת הבדיקה:

1. GMF
2. NCF
3. MF
4. non-Personalized

נדרג את המודלים לפי רמת הדיוק מהגבוה (MAE נמוך) לנמוך (MAE גבוה) לפי הMAE של קבוצת האימון:

1. NCF
2. MF
3. GMF
4. non-Personalized

נדרג את המודלים לפי משך זמן האימון של המודל (מהקצר לארוך):

1. NCF
2. MF
3. GMF
4. non-Personalized



קיבלנו GMF הוא המודל שמניב תוצאות מדויקות יותר מכל שאר המודלים.  
נבחין כי NCF גם הוא עם תוצאות טובות מיד לאחריו.  
כמו כן יש לזכור שמבין 2 המודלים ביצענו זמן רב בעידון ומציאת  
הפרמטרים הנכונים NCF עד שהגענו לרמת הדיוק הנ"ל לכן יתכן שעם עוד  
עבודה אולי מודל ה-GMF אף יכול לקבל תוצאות מדויקות יותר.

### תרגיל 3:

המודל שנבנה כעת יהיה בעצם מודל מסוג context based – נמליץ על סרט למשתמש גם עפ"י מאפיינים של המשתמש (גיל ומין)

עבור תרגיל זה נוסיף ונשתמש במידע אשר קיים על המשתמשים. בחרתי להשתמש בגיל ובמין המשתמש היות ואלו דברים לדעתי יותר משפיעים מאשר המקצוע של האדם. גיל המשתמש יכול להוות פקטור משמעותי עבור גילאים קטנים למשל ימליץ יותר על סרטי ילדים והרפתקאות מאשר על סרטי אימה, ומין המשתמש יכול להצביע על ז'אנר מועדף.

שלב מקדים לבניית המודל עם המאפיינים החדשים יהיה עיבוד הנתונים נעשה זאת ע"י כך שאחד את המידע לכדי dataframe אחד שיכיל את המזהה של המשתמש, מזהה של הסרט אותו דירג, דירוג הסרט, מין המשתמש, גיל המשתמש על מנת שנוכל לקשר בין המידע שנתון בטבלאות השונות. עוד שלב מקדים שהיה עלינו לבצע על מנת שהמודל יוכל לקבל כקלט את מין המשתמש הוא המרתם לערך מספרי על מנת שנוכל להכניסו לחישוב במודל שנבנה.

נבחר לממש את המודל ע"י NCF היות והוא המודל שמצאנו מבין ארבעת המודלים שהגיע לתוצאות הטובות מיד אחרי GMF אך בחרנו השתמש בו.

המודל שאנו נממש הוא מודל היברידי של feature combination כפי שלמדנו בכיתה זאת כיוון שאנו משתמשים במאפיינים נוספים על המשתמש ומאחדים אותם ביחד עם המידע על דירוג הסרט.

הארכיטקטורה של המודל מבוססת על הרעיון של feature augmentation – המאפיין שמערכת ההמלצה השניה (מבוססת על המאפיינים החדשים) משתמשת הוא הפלט של מערכת ההמלצה הראשונה (user-item).

נרצה לבנות מודל NCF שיקבל מספר שכבות, תחילה יאחד שכבות embedding שהן ה user וה-item לאחר מכן ע"י שימוש בConcatenate נשרשר את הפלט של ה user-item אל הקלט של השכבה הבאה אצלנו מין הצופה או לחלופין גיל הצופה וזה יהיה המודל ההיברידי שלנו.

כמו כן נזכור לפני האימון לפצל את המידע ל80 אחוז לקבוצת האימון ול20 אחוז קבוצת הבדיקה.

ב. קוד במחברת ה jupeter המצורפת להגשה.

ג. הוספנו שכבות של מין וגיל ואימנו אותן עם ערכי אותם הפרמטרים שהגדרנו כ default -הערכים שהניבו את התוצאות הטובות ביותר עבור מודל NCF מתרגיל 2:

- **Size of Layer** 100
- **Optimizer** = Nadam
- **Function Loss** = mean absolute error
- **Activation Function** = relu
- **drop-out of**: 0.1,

בנוסף חישבנו עם פרמטרים נוספים ושילובים שלהם לראות אם גם כאן נצליח להגיע לשילוב מיטבי עבור המודל החדש:

**Size of Layer**:, 20, 100

**Optimizer** = adam, Nadam

**Function Loss** = mean absolute error, mean squared error

**Activation Function** = relu, linear, selu

**drop-out of**: 0.1, 0.5

להלן הטבלה אשר מרכזת את כל תוצאות המדידות עבור המודלים החדשים:

Additional Feature	paramter	MAE- train	TIME	MSE - test
<b>Age</b>	default	0.91272	104.7925	0.9669
<b>Gender</b>	default	0.9243	106.7996	1.0132
<b>Age</b>	Adamx	0.9138	99.2087	0.9384
<b>Gender</b>	adamx	0.9168	155.6402	0.9246
<b>Age</b>	linear	0.91791	112.5583	0.9872
<b>Gender</b>	linear	-	113.3625	0.9757
<b>Age</b>	Selu ,50, mse	0.9104	150.9089	0.9298
<b>Gender</b>	Selu,50, mse	-	104.8177	1.0013
<b>Age</b>	Mse	-	106.4828	0.9899
<b>Gender</b>	mse	-	107.7074	0.9613
<b>Age</b>	0.5	0.9039	134.1343	0.9247
<b>Gender</b>	0.5	0.92639	138.3305	0.9618

התוצאות הטובות ביותר עבור המודלים הנ"ל:

Additional Feature	paramter	MAE- train	TIME	MSE - test
<b>Age</b>	0.5	0.9039	134.1343	0.9247
<b>Gender</b>	adamx	0.9168	155.6402	0.9246

ד. כעת נשווה בין כל המודלים השונים אשר בדקנו לאורך הפרויקט הנ"ל  
ונראה מי הניב את התוצאות הטובות ביותר:

Model	MAE - train	Time	MAE - test
Matrix Factorization	0.8012	233.1664	0.8021
Generalized Matrix factorization	0.90155	338.712	0.748837
Non-Personalized (HW2)	לא חושב - בעבודה 2	486.9826	1.0196
NCF תרגיל 2	0.7470	77.6727	0.7967
NCF + age	0.9039	134.1343	0.9247
NCF + gender	0.9168	155.6402	0.9246

נדרג את המודלים לפי רמת הדיוק מהגבוה (MAE נמוך) לנמוך (MAE גבוה) לפי ה-MAE של קבוצת הבדיקה:

1. GMF
2. NCF
3. MF
4. NCF + gender
5. NCF + age
6. non-Personalized

נדרג את המודלים לפי רמת הדיוק מהגבוה (MAE נמוך) לנמוך (MAE גבוה) לפי ה-MAE של קבוצת האימון:

7. NCF
8. MF
9. GMF
10. NCF + gender
11. NCF + age
12. non-Personalized

נדרג את המודלים לפי משך זמן האימון של המודל (מהקצר לארוך):

1. NCF
2. NCF + age
3. NCF + gender
4. MF
5. GMF
6. non-Personalized

ה. לאחר מעבר על כל התוצאות והמודלים השונים – נוכל להמליץ על מודל GMF היות והוא קיבל את התוצאה של MAE הכי נמוך עבור קבוצת הבדיקה, נבחין ש NCF של תרגיל 2 קיבל את ה MAE הטוב ביותר בזמן אימון המודל וכך גם את זמן הריצה הטוב ביותר.

היות ובמערכת המלצה חשוב לנו גם השילוב של זמן החישוב וגם הדיוק נבין שאומנם GMF קיבל את ה MAE הקטן ביותר אך יחד עם זאת את הזמן הארוך ביותר מבין המודלים החדשים כך שאומנם הדיוק טוב יותר אך זה מגיע עם tradeoff של המתנה להמלצה.

בהנתן השילוב של הזמן והדיוק של קבוצת הבדיקה לדעתי כדי להמליץ על NCF של תרגיל 2 כמערכת המומלצת.

לאורך כל המודלים השונים ראינו עד כמה שימוש בפרמטרים לא נכונים יכול להשפיע משמעותי על תוצאות האימון אף ראינו מקרים בהם פרמטר אחד הכפיל את התוצאה של ה MAE גם בקבוצת הבדיקה וגם בקבוצת האימון.

לכן לדעתי כיוון שבשלב 2 וב שלב 3 ביצענו את הכי הרבה בדיקות ושילובים של פרמטרים שונים יצא שהגענו לשילוב הפרמטרים המיטבי ביותר שכן ניתן לראות בתוצאות של תרגיל 2 שיש מדידות בהן ה MAE גדול משמעותית משל המודלים העיקרים בהם השתמשנו בשילוב של הפרמטרים שהניבו את התוצאה הטובה.

כמו כן לדעתי ככל שהמידע עליו נתאמן יהיה גדול יותר הצלחת המודל ההיברידי NCF עם המאפיינים הנוספים יהיה מוצלח אפילו יותר וישתפר כיוון שלא מספיק 80000 דירוגים על מנת לגבש דעה טובה עם איזה סרט אישה תאהב, או איזה סרט להמליץ לבן 80.