



רשתות עצביות (Neural Networks) – הלכה למעשה



משה אונגר

יעקב זהבי

פרופ' יעקב זהבי הוא פרופסור אמריטוס בפקולטה לניהול ע"ש קולר באוניברסיטת תל אביב. הוא אחד מפורצי הדרך בתחום כריית המידע (Data Mining) בעולם נתוני העתק, שבו הוא מעורב במספר חזיתות – מחקר, הוראה, פיתוח תוכנה ויישומים לקבלת החלטות. פרופ' זהבי החל את הקריירה המקצועית שלו בתחום מערכות מידע בתור מנתח מערכות בסקטור הציבורי. עם סיום לימודי הדוקטורט באוניברסיטת פנסילבניה הצטרף לפקולטה לניהול באוניברסיטת תל אביב ובמשך מספר שנים עסק בפיתוח וביישום של מודלים של חקר ביצועים וקבלת החלטות בתחום האנרגיה והחשמל. בסוף שנות השמונים עבר "הסבה מקצועית" לתחום של שיווק מבסיסי נתונים, וממנו הגיע לתחום של כריית מידע שבו הוא עוסק עד היום. זכה פעמיים רצופות במדליית הזהב בתחרות השנתית לגילוי ידע (Knowledge Discovery) שמאורגנת על ידי ACM (American Computation Machinery). מספר מאמרים שלו בתחום זה זכו בפרסים על מצוינות אקדמית.

ד"ר משה אונגר הוא חבר סגל בכיר בפקולטה לניהול ע"ש קולר באוניברסיטת תל אביב בתחום ניהול טכנולוגיה ומערכות מידע, וכן חוקר אורח בחברת אמזון. מחקריו כוללים בין היתר שימוש בלמידת מכונה ושיטות AI, לרבות שימוש בלמידה עמוקה, למציאת דפוסי התנהגות של משתמשים במערכות המלצה, למידה מנתוני עתק (Big Data) לקבלת החלטות עסקיות, ופיתוח שיטות מדעי הנתונים לפתרון בעיות עסקיות. סיים את לימודי הדוקטורט שלו במחלקה להנדסת מערכות תוכנה ומידע באוניברסיטת בן גוריון בשנת 2018, והשתלם בפוסט-דוקטורט במנהל עסקים באוניברסיטת ניו יורק. בעל תואר ראשון בהנדסת תוכנה ותואר שני בהנדסת מערכות תוכנה ומידע, שניהם מאוניברסיטת בן גוריון בנגב. משמש כיום חוקר אורח בחברת אמזון, ובעבר ניהל פרויקטים במרכז למחקר אבטחת הסייבר (CSRC) במעבדות דיטשה טלקום באוניברסיטת בן-גוריון ושיתף פעולה עם חברות שונות, כולל Spotify ו-Dell EMC.

תקציר

רשתות נוירונים מלאכותיות (ANN – Artificial Neural Networks) הם מודלים ממוחשבים המחקים את תהליך הלמידה של המוח האנושי. במקור, רשתות נוירונים שימשו לצורך אמידה של פונקציות. עידן נתוני העתק הביא את רשתות הנוירונים לקדמת הבמה בשל יכולתן לגלות באופן אוטומטי דפוסיים ותבניות בתוך כמויות אדירות של נתונים שגישות אנליטיות ויוריסטיות מתקשות להתמודד איתן. בשילוב עם זמינותם הגוברת של מעבדים חזקים וזולים, רשתות הנוירונים הפכו לפתרון האולטימטיבי למנעד רחב של בעיות עסקיות ומדעיות בתחומים שונים.

אבל למרות ההילה, תהליך היישום של רשתות נוירונים לפתרון בעיות עסקיות ומדעיות לא חף מבעיות, חלקן נובעות מהעובדה שאין לרשתות נוירונים תשתית תיאורטית בדומה למודלים מבוססי רגרסיה, וחלקן מהאופי והמבנה של הבעיות העסקיות. במאמר זה נסקור מקצת מן הקשיים ביישום רשתות נוירונים לפתרון בעיות בחייו אנליטי לצורך איתור קהלי יעד לשיווק מוצרים ושירותים בשיווק ישיר ("טירגוט"). כמו כן, נסקור גם את היתרונות של שימוש ברשת נוירונים לבעיות טירגוט ונשווה את ביצועי רשת הנוירונים לאלה המבוססים על רגרסיה לוגיסטית. נסיים בסקירה קצרה של נושא הלמידה העמוקה (Deep learning) שמהווה את הדור הבא של רשתות הנוירונים, ושימושיה החדשניים במגוון תחומי מחקר בתעשייה.

1. הקדמה

שרכשו את המוצר בעבר וכן דוגמאות של לקוחות שלא רכשו את המוצר בעבר, המערכת "לומדת" מתוך הדוגמאות מהם המשתנים (קניות של מוצרים אחרים בעבר, פניות שנעשו לצרכנים לרכוש מוצרים בעבר שלא נענו, משתנים דמוגרפיים כמו גיל ורמת הכנסה, מצב סוציאלי-אקונומי ועוד) שמאפיינים את הלקוחות שמניבים. בעזרת מאפיינים אלה ניתן לאתר לקוחות עם סבירות גבוהה להיענות לפנייה לקנות את המוצר הנתון. לדוגמה, אם רוב הקונים בעבר של מוצר (או שירות) כלשהו הן נשים, המערכת תיתן משקל גדול יותר למשתנה המגדר בהחלטות הטירוט. לו היה מדובר במשתנים בודדים, הבעיה הייתה יחסית פשוטה. הבעיה היא שבבעיות טירוט מספר המשתנים המסבירים נוסק לכמה מאות, אם לא יותר, חלק משפיעים יותר וחלק פחות. אז איך מתבצע תהליך החיזוי במקרה הזה? על כך נדון בהרחבה במאמר הנוכחי.

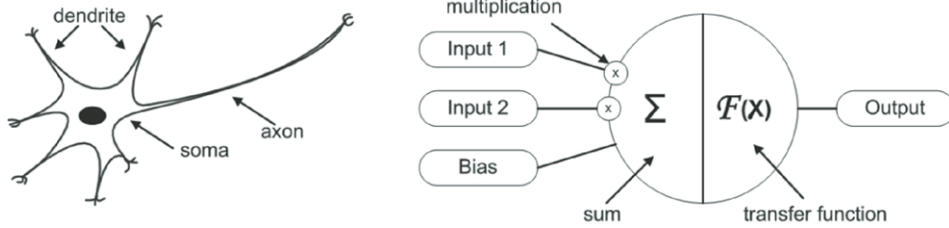
רשתות נוירונים התחילו בתרועה גדולה בתור כלי החיזוי האולטימטיבי שיכול להתמודד בהצלחה עם בעיות חיזוי מורכבות בתחומים שונים. McCulloch & Pitts (1943) היו הראשונים שהציגו את המוח האנושי כסדרה של נוירונים עם אינטראקציות. Donald Hebb (1949) העלה את ההשערה שנתבים קיימים בין הנוירונים ברשת מחזקים את הקישוריות ביניהם. Rosenblatt (1957) היה הראשון שפיתח אמצעי דמיוני נוירון שנקרא פרספטרון והשתמש בו, אומנם בהצלחה חלקית, בתור מערכת למידה. Widrow & Hoff (1960) היו בין הראשונים שפיתחו מערכות למידה מבוססות על רשתות נוירונים שנמצאות בשימוש עד היום. Minsky & Papert (1969) העלו ספק לגבי היכולת של פרספטרון דו-שכבתי לפתור אפילו בעיות פשוטות, מה שמנע מחוקרים רבים לעסוק בתחום ו"תקע" את נושא רשתות הנוירונים קרוב ל-15 שנה! Hopfield (1982) חידש את העניין ברשתות נוירונים לאחר שהראה שנישות מבוססות ANN אכן מסוגלות לפתור בעיות אופטימיזציה. Rumelhart, Hinton & Williams (1986) פיתחו את אלגוריתם ה-Back Propagation שמשמש עד היום בתור המנגנון הנפוץ ביותר לאימון של מערכת ANN. רב-שכבתית והביא את נושא רשתות הנוירונים לקדמת הבמה. ואומנם, מאז ועד היום נישות מבוססות רשתות נוירונים נהנות מפריחה מחודשת וזכות לפופולריות רבה בקרב הקהילה האקדמית והעסקית.

אלא שהיישום של נישות מבוססות רשתות נוירונים לפתרון בעיות עסקיות הוא תהליך לא פשוט, במיוחד בעולם נתוני העתק. הדבר נובע הן מהעובדה שאין לרשתות נוירונים תשתית

תופעות פיזיקליות וביולוגיות מהוות לאחרונה מקור השראה לפיתוח אלגוריתמים לפתרון בעיות מתמטיות מורכבות. בין השיטות האלה נציין את נישת SA (Simulated Annealing), (Van Laarhoven & Aarts, 1987), שמחקה את תהליך הקירור של מערכות פיזיקליות במטרה לשנות את התכונות האנרגטיות או המכניות שלהן, רשתות נוירונים (Neural Networks), (Rumelhart et al., 1986), שמחקות את תהליך הלמידה של המוח האנושי, אלגוריתמים גנטיים - GA (Genetic Algorithms), (Goldberg, 1989), המחקים תהליכים גנטיים בתחום הרבייה, ניתוח הישרדות (Survival Analysis), (Cox & Oakes, 1984), שמחקה תהליכים של הישרדות בתחום הביולוגיה והרפואה, ועוד. במאמר זה נתמקד ברשתות נוירונים ונתמקד בתפקיד שלהם לפתרון בעיות חיזוי אנליטי מורכבות.

רשתות נוירונים מלאכותיות (Artificial Neural Networks - ANN), ובקיצור רשתות נוירונים, הם תהליכים שמחקים את תהליך הלמידה של המוח האנושי. היישומים המסורתיים הראשונים של רשתות נוירונים היו לצורך אמידה של פונקציות. עידן נתוני העתק (Big Data) המכיל נתונים מסוגים שונים, תמונות וטקסט, הביא את רשתות הנוירונים לקדמת הבמה בשל היכולת שלהן לנלות דפוסים ותבניות מורכבות בתוך הררי הנתונים המגיעים כיום מכל עבר, לרוב באופן לא מובנה. אחד היישומים המובילים של רשתות נוירונים הוא תחום החיזוי האנליטי שבו נתמקד במאמר הזה. בבעיות של חיזוי אנליטי, מיזנים את רשת הנוירונים במספר גדול של תצפיות כדי למפות את הקשר בין המאפיינים של התצפיות לבין משתנה תוצאה (output or target variable). למשל, בבעיות "טירוט" בשיווק ישיר, המטרה היא למפות את הקשר בין משתנים שיכולים להסביר את החלטות הקנייה של הלקוחות לבין התגובה שלהם האם להיענות או לדחות את הצעת הרכישה. תהליך החיזוי האנליטי מתבסס על למידה מתוך דוגמאות שמדענים מאמינים שקורים במוח האנושי. למשל, אם אנחנו רוצים ללמד ילד ש-3 ועוד 4 שווה 7, אנחנו חוזרים על הפעולה הזו מספר פעמים עד שהילד לומד שאכן $3 + 4 = 7$. דוגמה נוספת מתחום התמונות: אם מראים לילד תמונות שונות של בעלי חיים מסוג מסוים, אז הילד לומד להבדיל בין בעלי החיים לבין עצמים אחרים שאינם בעלי חיים. ואיך זה מתקשר לעולם העסקי? בבעיית טירוט בשיווק, למשל, אם מיזנים למערכת רשתות הנוירונים דוגמאות של לקוחות

איור 1: נוירונים ביולוגיים ומלאכותיים



תיאור של יחידה כזו מתואר בציוד השמאלי של איור 1.

ה-Dendrites מקבלים אותות (Signals) – גירויים ואותות חשמליים מנוירונים אחרים שמוכפלים במשקלות טרם כניסתם ל-soma. ה-soma, שמהווה למעשה את יחידת העיבוד של הנוירון, מסכמת את האותות המשוקללים הנכנסים אליה, וכאשר הסכום עובר ערך סף (Threshold) מסוים היא "יורה" (Fire) אות שמועבר דרך ה-axon לנוירונים האחרים הקשורים אליה.

המקבילה ה"מלאכותית" של הנוירון הביולוגי (הנוירון ה"עסקי") מתואר בציוד הימני של איור 1. יחידת העיבוד של רשת הנוירונים העסקית ניזונה ממשתני (נתוני) הקלט. בדומה לנוירון הביולוגי, כל אחד ממשתני הקלט מוכפל במקדם (משקל). סכום נתוני הקלט המשוקללים נכנס ליחידת העיבוד המרכזית (שהיא המקבילה של ה-Soma בנוירון הביולוגי), ועובר טרנספורמציה באמצעות פונקציית הפעלה (אקטיבציה) (FX) שיוצרת את משתנה הפלט של יחידת העיבוד. משתנה הפלט מנותב לרשת הנוירונים בהתאם לטופולוגיה של הרשת. צומת ה-bias ברשת הנוירונים המלאכותית מקביל לערך הסף ברשת הנוירונים הביולוגית.

הנוירון ה"עסקי" של איור 1 מהווה את אבן הבניין (Building block) של רשת הנוירונים העסקית. הרשת הפשוטה ביותר, המכונה פרספטרון, מורכבת מנוירון אחד עם שני משתני קלט ומשתנה פלט אחד. אבל גם לרשת פשוטה זו יש כבר יכולת חישובית. בהינתן המשקלות הנכונים של הענפים ברשת, הנלמדים באופן איטרטיבי (כפי שנדון בהמשך), וערך הסף (threshold), הרשת הפשוטה הזו מסוגלת לאמוד במדויק את הפונקציות הבוליאניות AND, NOT ו-OR. לדוגמה, נסתכל על הפונקציה AND שמתארת מצב שבו משתנה התגובה (הפלט) הוא במצב "True" (=1) אם שני משתני הקלט הם במצב "True", ובמצב "False" (=0) אם לפחות

תיאורטית, והן מהמאפיינים הייחודיים של הבעיות העסקיות והכמות האדירה של נתונים המשתתפים בתהליך. במאמר זה ננסה לעמוד על חלק מהקשיים האלה, לכמה מהם נציע פתרון, ולחלק אחר גישות אפשריות לפתרון. לא נדלג גם על היתרונות הטמונים בשימוש ברשתות נוירונים בבעיות עסקיות. כדי להמחיש את הדברים, נתמקד במאמר הזה ביישום גישות מבוססות רשתות נוירונים לפתרון בעיות טירגוט בדיור ישר, שמהן ניתן יהיה להקיש גם לתחומים עסקיים נוספים. קודם לכן נתאר בקצרה את הנושא של רשתות נוירונים, תוך שימת דגש על ההקבלה בין רשתות נוירונים במוח האנושי לרשתות נוירונים בעולם העסקי. נציין שעל מנת להבדיל בין רשתות נוירונים "ביולוגיות" ו"עסקיות", יש הנוהגים לכנות את הרשתות העסקיות בתור רשתות "מלאכותיות" ומסמנים אותם ב-ANN (Artificial Neural Networks), ואילו הרשתות ה"ביולוגיות" מסומנות ב-NN (Neural Networks). כדי לפשט את הנושא אנחנו לא נעשה הבדלה כזו, ומאן והלאה נתייחס לשני סוגי הרשתות בכינוי NN.

2. נוירונים ביולוגיים ומלאכותיים

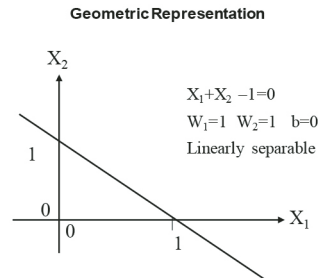
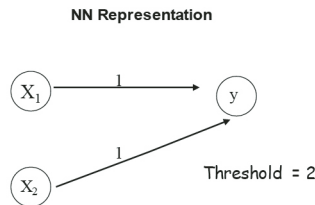
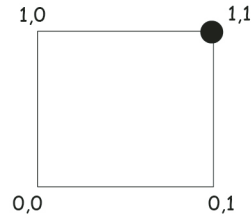
המוח האנושי מורכב מרשת המכילה כ- 10^{12} תאים, המכונים נוירונים, שכל אחד מהם מקושר בערך ל- 10^4 נוירונים. כל נוירון מכיל שלושה מרכיבים עיקריים:

- Dendrites
- Soma
- Axon

וכל יחידה כזו מהווה יחידת עיבוד חישובית (Processing element).

AND Function

X_1	X_2	y
1	1	1
1	0	0
0	1	0
0	0	0



חישוב דומה תופס גם לגבי שתי הפונקציות הבוליאניות האחרות, OR ו-AND, ולא נפרט אותן כאן. רק נציין שהמשקלות של הענפים ברשת וערכי הסף משתנות מפונקציה לפונקציה.

אחד ממשתני הקלט הוא במצב "False". איור 2 מתאר את הפונקציה הזו בצורה גרפית עם ארבעה מרכיבים: הפונקציה AND, תיאור גיאומטרי שלה, הפרסטרון שבאמצעותו ניתן לאמוד את הפונקציה AND וההצגה הגיאומטרית של הפתרון. המשקלות של כל אחד ממשתני הקלט (המיוצג על ידי הענפים בפרסטרון) הוא 1, וערך הסף עבור משתנה הפלט הוא 2. כלומר כאשר הסכום של משתני הקלט המשוקללים הוא 2 או יותר, משתנה הפלט "יורה" את הערך 1 (True), ואחרת הוא "יורה" את הערך 0 (False). החישובים לגבי הפונקציה AND מתוארים בטבלה 1.

נתבונן על הפונקציה XOR המתוארת באיור 3. פונקציה זו מקבלת את הערך "True" אם רק אחד ממשתני הקלט הוא "True". אחרת, הפונקציה מקבלת את הערך "False". כדי לתת דוגמה מעשית של העדפת מוצרים בתחום השיווק, נניח שהמשתנה X_1 מייצג את המגדר: 1 - אישה, 0 - גבר, ואילו המשתנה X_2 מייצג את קבוצת הגיל: 1 - צעיר, 0 - מבוגר. יחס XOR משמעותו שהמוצר מועדף דווקא על ידי נשים מבוגרות או גברים צעירים, אולי יחס מעט לא הגיוני אבל בהחלט לא בלתי אפשרי. כפי שנראה מתוך ההצגה הגיאומטרית של הפתרון, הפונקציה XOR היא לא ספרבילית בצורה ליניארית ולכן לא ניתן לייצג אותה באמצעות פרסטרון פשוט, אלא כפי שמתואר באיור 3 באמצעות רשת נוירונים המכילה עוד שכבת ביניים בין שכבת הקלט לשכבת הפלט. הרשת באיור 3 מכילה גם את המשקלות של הענפים ברשת ואת ערכי הסף. ראוי לציין שברשת זו יש ערך סף אחד לכל נוירון בשכבת הביניים וערך סף נוסף גם למשתנה הפלט. בהינתן המשקלות של הענפים ברשת וערכי הסף, ניתן לאמוד (בדומה לחישוב בטבלה 1) את פונקציית XOR, אם כי החישוב מורכב מקצת יותר איברים.

טבלה 1: חישוב פונקציית AND על סמך הנוירון המלאכותי

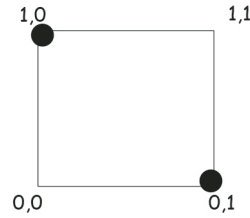
X_1	X_2	$W_1 \times X_1$	$W_2 \times X_2$	Sum	Output
1	1	1	1	2	1
1	0	1	0	1	0
0	1	0	1	1	0
0	0	0	0	0	0

כפי שנראה מתוך איור 2, הפונקציה AND היא ספרבילית עם פונקציה ליניארית שמפרידה בין ערכי ה-True וה-False. זוהי גם הסיבה שבגינה ניתן לייצג את הפונקציה הזו באמצעות נוירון פשוט אחד.

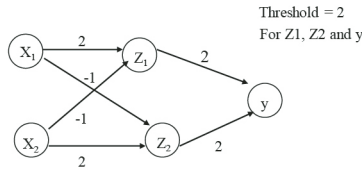
איור 3: פונקציית XOR

XOR Function

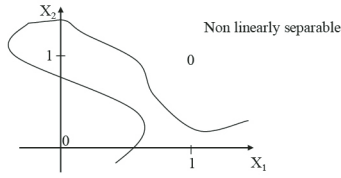
X_1	X_2	y
1	1	0
1	0	1
0	1	1
0	0	0



NN Representation



Geometric Representation

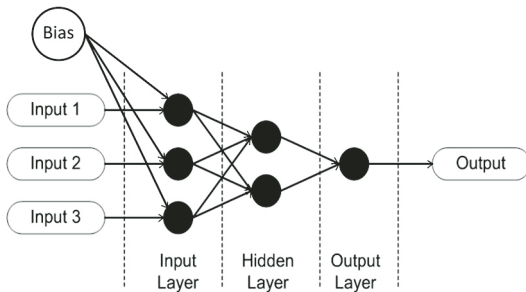


איור 4 מציג רשת נוירונים בסיסית שמורכבת משלוש שכבות של נוירונים: שכבת קלט (Input layer), שכבת ביניים המכונה שכבה נסתרת (Hidden layer), ושכבת פלט (Output layer). כל שכבה מורכבת מכמה נוירונים המכונים גם צמתים. הצמתים מחוברים בענפים עם משקלות, וככל שהמשקל גדול יותר, יש קשר חזק יותר בין הצמתים. נציין שבשכבת הקלט של הרשת באיור 4 כללנו את ערך הסף, המכונה גם Bias (בדומה לחותך, Intercept, במשוואת רגרסיה), בתור צומת נוסף בשכבת הקלט. בהתאם למבנה הבעיה, ניתן להוסיף צומתי Bias גם לצמתים בשכבת הביניים של הרשת.

תיאור נרחב וממצה יותר של השימוש ברשתות נוירונים לטיפול בפונקציות בוליאניות מופיע במאמר מקוון (Steinbach, 2002).

הפונקציות הבוליאניות שתוארו לעיל הן רק דוגמה בסיסית ליכולת של רשתות נוירונים לאמוד פונקציות מתמטיות, אם כי לא תמיד ניתן לבטא אותן באמצעות פונקציה מתמטית "סיגורה". (Leshno et al., 1993). במאמר פורץ דרך, שעם קונפינורציה נכונה של הרשת, והגדרה נכונה של הפרמטרים ופונקציית האקטיבציה, רשת נוירונים יכולה לאמוד כל פונקציה מתמטית שהיא, כולל פונקציות ליניאריות ולא ליניאריות ואפילו פונקציות מתמטיות "עמומות" (Fuzzy).

איור 4: רשת נוירונים עם הזנה קדמית



3. רשתות נוירונים

כפי שאנו למדים מהדיון בסעיף הקודם, כדי לפתור פונקציות מורכבות יותר, כמו בעיית ה-XOR לעיל, אי אפשר לאמוד את הפונקציה באמצעות רשת פרסטרון חד-שכבתית פשוטה, אלא נדרשת רשת נוירונים רב-שלבית המכילה מספר שכבות. ביישומים של חיזוי אנליטי, מרבית הבעיות העסקיות והמדעיות מתבססות על רשתות נוירונים רב-שכבתיות, למעט במקרים שבהם קיימת הפרדה ליניארית בין המאפיינים של התצפיות ומשתנה התוצאה שלהן.

רשת הניורונים באיור 4 מקיימת את הכללים הבאים:

- מדובר ברשת כיוונית (Directional) שבה הזרימה היא תמיד בכיוון אחד, מכל שכבה לשכבה הגבוהה ממנה.
- אין קשרים בין צמתים השייכים לאותה שכבה.
- ניתן לדלג על שכבות, כלומר ייתכן שצומת משכבת הקלט יקושר ישירות לשכבת הפלט.

רשת זו מכונה רשת ניורונים עם הזנה קדמית (Feed-forward neural networks).

מספר השכבות הנסותרות, וכן מספר הצמתים בכל שכבה נסותרת, תלויים בבעיה העסקית שמנסים לפתור הנקבעת מראש על ידי מתכנן הרשת. ככל שיש יותר שכבות ובכל שכבה יותר צמתים, כך המערכת "מעודנת" יותר וכיכול "טובה" יותר. מצד שני, ריבוי שכבות וצמתים מגדיל את מספר הענפים ברשת. כפי שנראה להלן, הפרמטרים שיש לאמוד ברשת הם המשקלות של הענפים, ולכן ככל שיש יותר משקלות לאמוד לא רק שהמאמץ החישובי רב יותר, אלא יש גם סכנה גדולה יותר להתאמת יתר (Overfitting). מספר הענפים ברשת הבסיסית של איור 4 הוא:

$$\{ [מספר משתני הקלט + 1] * [מספר הצמתים הנסותרים] \}$$

אם ניקח דוגמה שכיחה שבה 35 משתנים בשכבת הקלט וחמישה צמתים בשכבה הנסותרת, נקבל שיש צורך לאמוד $5*(1+35) = 180$ משקלות, וזה הרבה!

מספר הניורונים בשכבת הפלט תלוי גם הוא בבעיה העסקית ויכול להיות גדול מאחד. למשל, בבעיות סיווג (Classification) עם מספר קבוצות (Classes), שכבת הפלט תכיל ניורון אחד לכל קבוצה, מה שמגדיל גם את מספר הענפים ברשת. בנייה מתאימה של רשת ניורונים לפתרון בעיות עסקיות היא משימה מורכבת שמשפיעה גם על טיב הפתרון, ואלה אנחנו נתייחס בהמשך.

מה שקובע את התכונות של רשת הניורונים זוהי פונקציית האקטיבציה $F(x)$ שממירה את הסכום המשוקלל של הניורונים הנכנסים ליחידת העיבוד למשתנה הפלט. בדרך כלל בוחרים את פונקציית האקטיבציה על פי הבעיה שאנו רוצים לפתור. הפונקציות הפופולריות הן פונקציה ליניארית (linear function), פונקציות סף (Threshold function),

ופונקציות אקטיבציה לא ליניאריות שהנפוצות שבהן היא הפונקציה הסיגמואידלית (Sigmoidal function) והפונקציה Hyperbolic tangent.

פונקציית האקטיבציה הליניארית מתוארת באיור 5 ומאפשרת לנסח שורה של בעיות ברגרסיה ליניארית. למעשה אפשר לנסח כל רשת ניורונים עם פונקציית אקטיבציה ליניארית כבעיית רגרסיה ליניארית, אפילו אם מדובר ברשת ניורונים רב-שכבתית, ובתנאי שגם כל פונקציות האקטיבציה של השכבות הנסותרות הן ליניאריות.

בפונקציית הסף עם מדרגה אחת, המתוארת באיור 5, כבר השתמשנו לעיל על מנת לייצג את הפונקציה הבוליאנית AND באמצעות רשת ניורונים.

פונקציית סף עם 2 מדרגות מתוארת באיור 5.

פונקציית האקטיבציה הסיגמואידלית מתוארת באיור 5. במקרה של רשת עם שכבה נסותרת אחת, רשת הניורונים אקוויולנטית לבעיה ברגרסיה לוגיסטית. הפונקציה הסיגמואידלית היא בחירה מועדפת בתור פונקציית אקטיבציה לא ליניארית, שכן ערכי הפלט שלה הם בתחום 0-1, מה שמאפשר להתייחס אליהם כאל הסתברויות (או לפחות קירוב שלהן).

פונקציית האקטיבציה מסוג Hyperbolic tangent מתוארת באיור 5.

נציין שרשת ניורונים רב-שכבתית עם פונקציות אקטיבציה לא ליניאריות (סיגמואידלית, Hyperbolic tangent, ואחרות) מהווה בסיס לשורה של בעיות רגרסיה לא ליניאריות.

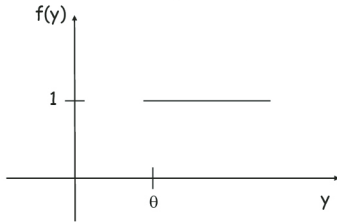
פונקציות האקטיבציה באיור 5 הן לא פונקציות אקטיבציה היחידות. קיימות גם פונקציות אקטיבציה אחרות, מעבר לפונקציות הני"ל, אבל הן ללא ספק הנפוצות ביותר.

המגוון הרחב של הקונפיגורציות האפשריות עבור רשתות הניורונים והמבחר הרב של פונקציות האקטיבציה הקיימות, מאפשרים לנסח מנעד רחב של בעיות באמצעות רשתות ניורונים, ובמאמר זה עוד לא סקרנו שורה ארוכה של רשתות ניורונים שחורגות מתחום המאמר, כגון רשתות מחזוריות (recurrent NN), LSTM, RBF – Radial basis functions.

איור 15: פונקציית מדרגה – מדרגה אחת

Threshold Function

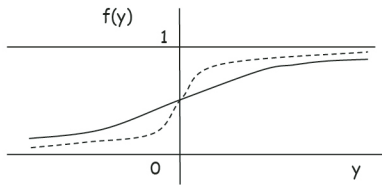
$$f(y) = \begin{cases} 1 & y \geq \theta \\ 0 & y < \theta \end{cases}$$



איור 15: פונקציית סיגמואידלית

Sigmoid Function

$$f(y) = \frac{1}{1 + \exp(-\sigma y)}$$

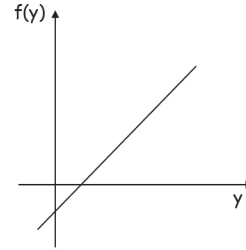


σ - scaling factor (steepness parameter)

איור 5א: פונקציית ליניארית

Linear Function

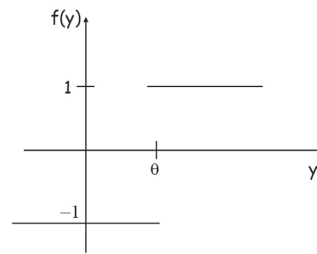
$$f(y) = a + b \cdot y$$



איור 5ג: פונקציית מדרגה – שתי מדרגות

Threshold Function

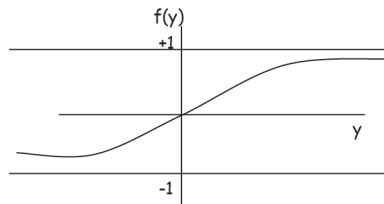
$$f(y) = \begin{cases} 1 & y \geq \theta \\ -1 & y < \theta \end{cases}$$



איור 5ה: פונקציית היפרבולית

Hyperbolic Tangent

$$f(y) = \frac{1 - \exp(-\sigma y)}{1 + \exp(-\sigma y)}$$



(Long-short term memory) models, למידה עמוקה (Deep learning), ועוד. ייתכן שזו הסיבה לכך שרשתות נוירונים צברו פופולריות רבה כשיטה האולטימטיבית לפתרון בעיות בתחומים רבים ולא רק בבעיות עסקיות. נדון בהרחבות אלו של למידה עמוקה בסעיף 10.

4. תהליך הלמידה לפתרון בעיות NN עם הזנה קדמית (Feed-forward Neural Networks)

כאמור, באמצעות רשתות נוירונים ניתן לייצג בעיות עסקיות רבות. אנו נתרכז להלן בבעיות של חיזוי אנליטי (PA - Predictive Analytics) בתחום השיווק שבה אנו מנסים לחזות את התגובה של הלקוחות למבצע שיווקי של מוצר (או שירות) מסוים בהסתמך על תגובות הצרכנים שנחשפו למבצע השיווק של המוצר הנדון בעבר. שכבת הקלט מורכבת ממשתנים המסבירים את תגובת הצרכנים, כגון רכישות קודמות, פניות בעבר שלא נענו, משתנים דמוגרפיים של הלקוח, משתני אשראי ועוד, שמספרם יכול להגיע לכמה עשרות ואף יותר. השכבה הנסתרת מכילה מספר נוירונים שמשתנה בהתאם לרמת הסיבוך של הבעיה (למשל, מספר משתני הקלט). ולבסוף, שכבת הפלט מכילה צומת (נוירון) אחד שמבטא את התגובה של הלקוח להצעה השיווקית (1 - יגיב וירכוש את המוצר/שירות, 0 - לא יגיב ולא ירכוש את המוצר/שירות). מטרת הרשת היא לאמוד את הסבירות שצרכן יענה להצעה השיווקית בסולם (אורדינלי) בין 0 ל-1, ולכן נבחר בתור פונקציית אקטיבציה את הפונקציה הסיגמוידיאלית שתחום הערכים שלה הוא גם כן בתחום (0,1).

בהינתן הקונפיגורציה של רשת הנוירונים ופונקציית האקטיבציה, הנעלמים היחידים ברשת הם המשקלות w_j של הענפים ברשת שאותם מוצאים באופן אוטומטי באמצעות תהליך למידה. כזכור, רשת נוירונים לומדת מתוך דוגמאות שבהן המשתנה התלוי Y ידוע. בבעיות שיווק ישיר תהליך הלמידה מבוסס על מדגם אימון (Training sample) שכולל דוגמאות של לקוחות שנחשפו למוצר בעבר, שמהם חלק קטן השיבו בחיוב והזמינו את המוצר, וחלק גדול (בדרך כלל מרבית הלקוחות) לא הגיבו לפנייה כלל. בהינתן המאפיינים (משתני הקלט) של הלקוחות שהגיבו ושלא הגיבו, אלגוריתם

הלמידה מעדכן את המשקלות של הענפים ברשת ומייחס משקלות גדולים יותר לענפים הקשורים עם משתני הקלט של המניבים, ומשקלות קטנים יותר למשתני הקלט של הלא מניבים. לדוגמה, אם רוב המניבים הם נשים ומיעוטם גברים, האלגוריתם ייחס משקל גדול יותר לנשים מאשר לגברים. לשם כך אנחנו מגדירים פונקציית מטרה שמשקפת את שגיאת התחזית. פונקציית שגיאה פופולרית היא סכום הסטיות הריבועיות (SSE - Sum of Squared Errors) על פני כל הלקוחות במדגם האימון בין הערכים בפועל של המשתנה התלוי Y ובין התחזית שלו על פי המודל. המטרה של תהליך הלמידה הוא למצוא את המשקלות של כל הענפים ברשת שממזערים את שגיאת התחזית.

מבלי להיכנס לפרטים, נתאר להלן את אלגוריתם הזרימה קדמית/אחורית (Forward/backward propagation), הידוע גם בשם Δ rule, למציאת המשקלות של הענפים ברשת נוירונים עם הזנה קדמית. בניגוד למודל גרסיה המטפל בכל מדגם האימון ב"מכה אחת" (כלומר סימולטנית), תהליך הלמידה כאן הוא תהליך סדרתי שבו מכניסים למערכת לקוח בודד, שלעיתים נבחר באופן מקרי, בודקים את הפלט המתקבל, שבמקרה של פונקציית אקטיבציה סיגמוידיאלית הוא מספר בין 0 ל-1, משווים אותו לפלט בפועל (שהוא 0 או 1), מעדכנים את פונקציית המטרה, מתקנים את משקלי הענפים לאור פונקציית המטרה ועוברים ללקוח הבא, וחוזר חלילה.

תהליך הלמידה הזה מורכב משני שלבים עיקריים:

שלב ראשון - זרימה קדמית (Forward propagation). בשלב זה מתקדמים משכבת הקלט לשכבת הפלט. בהינתן הערכים של משתני הקלט, הקונפיגורציה של הרשת ומשקלי הענפים, מחשבים את הערכים של הצמתים הנסתרים, הערך של צומת הפלט ואת שגיאת התחזית.

שלב שני - זרימה אחורית (Backward propagation). בשלב זה מתקדמים אחורה מצומת הפלט לצומת הקלט בתהליך רקורסיבי, מעדכנים את פונקציית המטרה על סמך שגיאת התחזית שהתקבלה בשלב ההזנה הקדמית, ומחשבים את השינויים הדרושים במשקלות הענפים על מנת להקטין (בבעיית מינימום) את הערך של פונקציית המטרה. תהליך התיקון של משקלי הענפים מבוסס על שיטת הגרדיאנט. מאחר שמדובר על בעיית מינימום, התיקון הוא בכיוון "מינוס" הנגדרת של פונקציית המטרה ביחס למשקל הענף שבו

מדובר. כדי להפעיל את שיטת הנרדיאנט יש צורך להשתמש בפונקציות אקטיבציה שהן גזירות (Differentiable), כגון הפונקציה הסיגמוידיאלית.

מסתבר שבדרך כלל מעבר אחד על כל קבוצת הצרכנים במדגם האימון לא מספיק על מנת להבטיח התכנסות "טובה" של תהליך הלמידה, וצריך לחזור על תהליך הזרימה הקדמית והאחורית עבור מדגם האימון מספר פעמים. כל מעבר כזה מהווה מחזור אימון (Training cycle). מספר מחזורי האימון יכול להגיע לעשרות ואף יותר. על מנת לייעל ולשפר את תהליך החישוב, חוזרים על תהליך החישוב של הזרימה הקדמית כדי לחשב את שגיאת התחזית אחת למספר צרכנים המוגדר מראש (המכונה epoch), ואז מעדכנים את משקלות הענפים בתהליך הזרימה האחורית על מנת להקטין (בבעיית מינימום) את שגיאת התחזית, עוברים ל-epoch הבא, וחוזר חלילה.

כדי לאתחל את תהליך הלמידה הנ"ל, יש צורך לבחור משקלות התחלתיים עבור כל הענפים ברשת. באשר למשקלות הענפים המחברים את צומתי הקלט לצמתים הנסתרים, הנישה המומלצת בספרות המקצועית היא להשתמש בערכים אקראיים. מניסיונו מצאנו שהנישה האקראית נותנת בדרך כלל תוצאות פחות טובות וגם מאריכה בהרבה את משך זמני החישובים. לכן המלצתנו היא להשתמש במקדמים של משתני הקלט ממודל רגרסיה בתור המשקלות ההתחלתיים של הענפים ברשת היוצאים ממשתני הקלט. עובדה זו מחייבת להריץ מודל רגרסיה מקדים (ליניארית או לוגיסטית) על קובץ האימון על מנת לאמוד את המקדמים של משתני הקלט. בשל אופי בעיית הטירגוט שבה המשתנה התלוי הוא בינארי 0/1 (1 - הלוקוח קנה את המוצר/שירות, 0 - לא קנה), יש עדיפות לשימוש במודל רגרסיה לוגיסטית, שהוא המודל המוביל עבור בעיית רגרסיה עם משתנה תלוי בינארי, אך גם מודל הרגרסיה הליניארית המוכר והטוב יכול להתאים לאופי הבעיה. כפי שנראה בהמשך, הרצה מקדימה של מודל רגרסיה על מדגם האימון גם מאפשרת להקטין את הממדים של רשת הניורונים וכך ולחסוך בזמני חישוב מבלי לפגוע באיכות המודל.

ובאשר למשקלות ההתחלתיים של הענפים המחברים את הצמתים הנסתרים לצומת הפלט, גם כאן ההמלצה המקובלת היא להשתמש במספרים אקראיים. אולם הספרות מציגה אפשרויות נוספות (שלא נדון בהן) לבחירת המשקלות ההתחלתיים.

כאמור, יש צורך לחזור על תהליך הזרימה הקדמית והאחורית עבור מספר מחזורי אימון עד שמתקיימים תנאי הסיום שבדרך כלל מוגדרים מראש (למשל, קביעה מראש של מספר מחזורי האימון).

לאחר בניית המודל, השלב הבא הוא יישום המודל, תהליך שנקרא scoring (בעברית - ציינון), שבו נותנים ציון (Score) עבור לקוחות חדשים, המשקף את הסבירות שהלקוח יענה להצעה השיווקית. הציונים האלה מחושבים על ידי הפעלת תהליך הזרימה הקדמית (Forward propagation), משכבת הקלט לשכבת הפלט, תוך שימוש במשקלות הסופיים של תהליך הלמידה. על סמך ציונים אלה בוחרים אילו מבין הלקוחות החדשים ישתתפו במבצע השיווק ואילו לא. הלקוחות החדשים הם לקוחות מבסיס הנתונים שלא השתתפו בבניית המודל או לקוחות שהצטרפו למערכת. דיון מפורט בתהליך הצינון מופיע בהמשך.

5. תהליך ההתכנסות

נציין שתהליך הלמידה בנישת ה-Forward/backward propagation הוא תהליך שעשוי לא להתכנס לעולם. תהליך הלמידה הוא תהליך חיפוש שאמור להסתיים כאשר השינויים במשקלי הענפים בין מחזור אימון אחד לשני הם מזעריים ונתיחים, מה שלא יקרה במקרה שלנו. נסייג ונאמר שאנו עוסקים כאן ספציפית בבעיית טירגוט בשיווק ישיר ולא מדובר כאן על מסקנה גורפת הנוגעת לכל היישומים האפשריים של רשתות ניורונים. ומדוע בבעיית טירגוט אין התכנסות? מכיוון שמדובר כאן בבעיה עם חוסר עקביות בתגובות הצרכנים, שכן קרוב לוודאי שאם נבחר מספר צרכנים עם אותם מאפייני קלט בדיוק, כלומר אותם משתנים מסבירים עם אותם ערכים, חלק מהם ירכשו את המוצר המוצע להם, והאחרים לא. לכן בעיית טירגוט היא למעשה בעיה סטוכסטית ש"מבלבלת" את תהליך הלמידה של רשת הניורונים בכך שהיא שולחת מסרים סותרים לניורונים ברשת הגורמים לכך שהרשת לא תתכנס אף פעם לפתרון יציב.

כדי לאשש את המסקנה הזו הרצנו בעיית חיזוי קטנה עם שני משתני קלט בינאריים (0/1), שני צמתים נסתרים, ומשתנה פלט שמקבל אחד משלושה ערכים אפשריים: 0.1, 0.5 או 0.9. עם זאת, דאגנו שלכל סט של משתני קלט יהיה תמיד אותו פלט (בניגוד לבעיה במציאות שבה כאמור ייתכנו שני

צרכנים עם אותו קלט אבל פלט שונה). משקלי הענפים התכנסו לאחר כ-120 מחזורי אימון עם דיוק סביר של תחזיות התגובה. כלומר השיטה "עובדת", לפחות מהבחינה המתמטית.

אולם הבעיה של חוסר ההתכנסות היא רק בעיה אחת של תהליך הלמידה. מסתבר שגם סידור הצרכנים בתהליך הלמידה משפיע מאוד על התוצאות. בדרך כלל מקובל להזין את הלקוחות במדגם האימון על פי סדר הופעתם בקובץ המקורי, ככל הנראה לפי מספר הלקוח (בהנחה שזהו סידור אקראי). למשל, אם נניח שהפרמטר של מספר מחזורי האימון הוא 10, תהליך הלמידה עובר על כל הלקוחות פעם אחת, ולאחר מכן על כולם פעם שנייה, וכך הלאה 10 פעמים.

פתרון אפשרי נוסף הוא סידור או חלוקה מחדש של הנתונים לאימון רשת הניורונים. למשל, לעבור על הלקוח הראשון 10 פעמים, אחר כך על הלקוח השני 10 פעמים, וכך הלאה. אפשרות אחרת היא לעבור תחילה פעם אחת על כל הלקוחות שהניבו וקנו את המוצר המוצע להם ("היקונים"), ואחר כך על כל הלקוחות שלא הגיבו לפנייה אליהם ("הלא קונים"), ואז שוב על הקונים, וכך הלאה 10 פעמים. לחלופין אפשר להפוך את הסדר, קודם לעבור על כל הלא קונים שבמדגם ואחר כך שוב על הקונים וכך הלאה. עוד אפשרות היא לחלק את מדגם האימון לשני חלקים, אחד שמכיל רק את הלקוחות הלא קונים והשני רק את הלקוחות הקונים, לבצע את תהליך הלמידה על כל חלק בנפרד, ולבסוף לחשב את הסבירות של כל לקוח להגיב להצעת הרכישה על פי ממוצע משוקלל של התחזיות (של משתני הפלט) שהתקבלו על ידי שני מדגמי הלמידה. שקלול סביר הוא על פי חלקם היחסי של הקונים והלא קונים במדגם האימון. זוהי רק רשימה חלקית של האפשרויות לבחור את סדר הזנת נתוני הקלט לרשת הניורונים. כאמור, הבעיה היא שכל סידור מביא לתוצאות שונות.

6. הערכת המודל

השאלה המתבקשת היא זאת: אם רשת הניורונים נותנת פתרון לא יציב, ואם סידור שונה של התצפיות המוזנות למודל נותן תוצאות שונות, מה זה אומר לגבי הישימות של מודל רשתות הניורונים לפתרון בעיות עסקיות? ובמילים אחרות, האם השיטה הזו "עובדת" גם על בעיות עסקיות "אמיתיות"? כמו בכל בעיה עסקית, הקריטריון הוא מבחן התוצאה. בבעיית הטירגוט שבה אנו עוסקים, המטרה העיקרית היא לבנות

מודל חיזוי שעושה הבחנה טובה בין הקונים ללא קונים ומצליח לסנן את הלקוחות הלא רווחים ממצע השייוק. יש לכך קריטריונים שנדונים בהרחבה בספרות על חיזוי אנליטי (PA – predictive analytics), כגון זהבי (2017). הגישה המקובלת היא באמצעות טבלת רווחים (Gains table) או ההצגה הגרפית שלה (Gains chart). בגישה זו ממיינים את התצפיות במדגם האימון בסדר יורד של תחזיות התגובה (כלומר הציון או ה-Score של התצפית) של מודל החיזוי (במקרה שלנו רשת הניורונים), מסכמים את נתוני התגובה ברמה של עשירונים (או כל אחוזון אחר), ובדקים את שיעורי התגובה (Response rate) ברמת העשירונים. מודל "טוב" הוא מודל שמצליח לרכז את רוב הקונים בעשירונים העליונים. בסעיף 9 מופיעה דוגמה לטבלת רווחים שמציגה יישום מעשי של רשתות נוירונים לבעיית טירגוט.

"מכה" אופיינית במודלים של חיזוי רב-מדמייים היא הבעיה של התאמת יתר (Overfitting), כלומר מצב שבו מודל החיזוי נותן תוצאות טובות על קובץ הנתונים המשמש לבניית המודל אבל תוצאות גרועות כאשר מיישמים את המודל על לקוחות חדשים. ברשתות נוירונים הסכנה של התאמת יתר היא גדולה במיוחד, בשל הצורך לאמוד מספר גדול מאוד של פרמטרים (משקלות הענפים), שמספרם גדל ככל שמספר הצמתים בשכבת הקלט ובשכבות הביניים גדול יותר. גם העובדה שמספר המגיבים בבעיות שיווק ישיר הוא קטן באופן משמעותי יחסית למספר הלא מגיבים, תורמת לתופעה של התאמת יתר.

כדי לבדוק מצב של התאמת יתר יש צורך לבדוק את תוצאות החיזוי של רשת הניורונים גם על קובץ נתונים שלא השתתף בבניית המודל, ולהשוות אותו לתוצאות החיזוי עבור מדגם האימון שהשתתף בבניית המודל. הדרך המומלצת ביותר היא לפצל את קובץ נתוני המבחן (Test dataset) באופן אקראי לשני קבצים בלתי תלויים – קובץ אימון (Training dataset) וקובץ תיקוף (Validation dataset), לבנות את המודל התגובה על בסיס קובץ האימון, ואז להפעיל אותו על קובץ התיקוף כדי לחזות את הערך של המשתנה התלוי (Score) עבור כל תצפית. כזכור, קובץ התיקוף מכיל גם את התגובות בפועל, מה שמאפשר להשוות את התחזיות של מספר המגיבים של המודל עם מספר המגיבים בפועל בין קובץ האימון לקובץ התיקוף. הכי נוח לעשות השוואה זו ברמה של עשירונים (או כל אחוזון אחר), כאשר העשירונים ממוינים בסדר יורד של תחזיות המודל מהעשירון הגבוה ביותר לנמוך ביותר. אם התפלגות מספר המגיבים בפועל ברמה של העשירונים בין שני

הקבצים הוא דומה, הדבר מעיד על מודל יציב ללא התאמת יתר. יש גם מבחנים סטטיסטיים מדויקים יותר המבוססים על סטטיסטי המקסימום (הידוע גם בשם *Kolmogorov Smirnov Statistics*) שנועדו לבדוק את מידת ההתאמה של נתוני המניבים בין שני הקבצים (DeGroot, 1993). אבל מבחינה מעשית אפשר להשתמש ב"כלל אצבע" שלפיו הפרשים של פלוס/מינוס 5%-10% בין תחזית מספר המניבים בקובץ התיקוף ובין מספר המניבים בפועל בקובץ האימון הם עדיין הפרשים "סבירים" שמעידים על חוסר התאמת יתר. דוגמה מספרית של תהליך בדיקת התאמת היתר מופיעה ביישום שבסוף המאמר (סעיף 9). דיון רחב יותר בנושא ניתן למצוא אצל זהבי (2017).

מכאן שעל מנת שנוכל להגדיר מודל חיזוי המבוסס על רשת נוירונים בתור מודל "טוב" שאפשר להשתמש בו לצורך קבלת החלטות, הוא צריך לענות על שני הקריטריונים הנ"ל, כלומר מודל ללא התאמת יתר שמצליח לאתר באופן נכון את מרב הקונים בעשירונים העליונים. הבעיה המרכזית במודל המבוסס על רשתות נוירונים הוא לקבוע באופן חדיערכי מי קונה פוטנציאלי ומי לא. ובמילים אחרות, כיצד מבדילים בין לקוח שהוא קונה פוטנציאלי ולכן כדאי לשתף אותו במבצע השיווק, לבין לקוח שלא סביר שיענה להצעה לרכוש את המוצר ולכן לא כדאי לכלול אותו במבצע השיווק. דיון נרחב בסוגיה זו יופיע בהמשך.

7. מה מבדיל את מודל רשתות הנוירונים ממודל רגרסיה?

המגבלה העיקרית של רשתות נוירונים היא שאין להן תשתית תיאורטית, שכן מדובר בתהליך שלא דורש כל הנחות תיאורטיות מקדימות אלא מתבסס על תהליך שהוא תוצר של למידת מכונה "נטו". זאת בניגוד למודלים מבוססי רגרסיה הנשענים על יסודות תיאורטיים מובנים ומוצקים.

ואכן, תשתית תיאורטית עוזרת בדרך כלל לבנות מודל יותר "טוב". במודל חיזוי המבוסס על רגרסיה יש משמעות למקדמים של המשתנים המסבירים. למשל, במודל רגרסיה ליניארית, המקדם של המשתנה המסביר מבטא את השינוי בערך של המשתנה התלוי אם משנים את הערך של המשתנה המסביר ביחידה אחת. רמת המובהקות של המשתנה המסביר

(*p-value*) קובעת באיזו מידה המשתנה תורם להסבר המשתנה התלוי, ככל שה-*p-value* קרוב לאפס, כך המשתנה המסביר מובהק יותר. על סמך התוצאות של מודל הרגרסיה ניתן גם להסיק על החשיבות של המשתנה המסביר. תכונות אלה ואחרות מאפשרות גם לבחור את המשתנים המסבירים למודל הרגרסיה מתוך האוסף הגדול של קבוצת המשתנים המסבירים הפוטנציאליים.

אולם גם במקרה של מודלים מבוססי רגרסיה, עדיין מדובר בתהליך מיינע שמחייב היכרות טובה עם תחום הידע (*Domain knowledge*) של הבעיה העסקית, הכרה של בסיסי הנתונים הרלוונטיים והמשתנים המסבירים של הבעיה העסקית, ידע רב בסטטיסטיקה, ועוד. כמו כן, יש צורך להגדיר מראש את הפונקציה שמתארת את הקשר הפונקציונלי בין המשתנים הבלתי תלויים למשתנה המסביר, לקיים הנחות תיאורטיות שונות, להתמודד עם בעיות של קוליניאריות בין משתנים, להגדיר טרנספורמציות לא ליניאריות של משתנים מסבירים, להגדיר מראש אינטראקציות בין משתנים מסבירים, ועוד. מאחר שבדרך כלל האינטראקציות והקשרים הלא ליניאריים בין המשתנים לא ידועים מראש, יש צורך במספר ניסיונות רב וריצות מחשב ארוכות על מנת למצוא את המודל הנכון. לא בכדי, בניית מודל רגרסיה רבי-ממדי נחשבת עדיין לאחת הבעיות המורכבות ביותר בסטטיסטיקה ומחייבת מיומנות רבה.

לעומת זאת, במודל של רשתות נוירונים המשתמש פטור מהצורך להגדיר את הנחות היסוד משום שאין כאלה. יש הסוברים שהיעדר תשתית תיאורטית דווקא מעניק יתרון מסוים למודלים מבוססי רשתות נוירונים, שכן הוא מותיר יותר דרגות חופש לחפש את הפתרון המיטבי של רשת הנוירונים. יתרה מזו, לרשתות נוירונים יש יכולת אינהרנטית, באמצעות הגדרה מתאימה של הרשת והוספת שכבות נסתרות, למצוא את הקשר הפונקציונלי בין משתנה התוצאה והמאפיינים במודל (על אף שלא תמיד ניתן להגדיר קשר זה באמצעות פונקציה מתמטית "יסגורה"), למצוא אינטראקציות בין משתנים מסבירים, לבטא קשרים לא ליניאריים בין משתנים מסבירים למשתנה התלוי, כולל פונקציות מורכבות מסוג XOR, ועוד. כל זאת נעשה על ידי הרשת באופן אוטומטי. עיקר המאמץ הוא להגדיר את רשת הנוירונים המתארת את הבעיה העסקית ה"תורנית", שהיא בעיקר תהליך יריסטי שאין מאחוריו שום שיקולים תיאורטיים אלא הוא מבוסס על ניסוי וטעייה. ככל הנראה, זוהי אחת הסיבות לפופולריות של רשתות הנוירונים לפתרון בעיות עסקיות.

8. סוגיות ביישום מודל רשתות נוירונים בבעיות טירגוט

אם כן, מודלים לחיזוי המבוססים על רשתות נוירונים הם בבחינת "קופסה שחורה" (Black box) שלא ניתנת לפרשנות (Interpretable). מדובר כאן על תהליך טכני בעיקרו, שמשמש בתהליך למידה חכם על מנת לחשב את הציונים שעליהם מתבססות ההחלטות העסקיות, אך מבלי לרדת למהות המודל ולמאפיינים שלו. בעולם נתוני העתק זה לא תמיד מספיק, שכן היום מתבקשת שיטה לניתוח פרטני הכוללת מדדים שמאפשרים להעריך את ההשפעה של המשתנים המסבירים על משתנה התוצאה (המשתנה התלוי) לא רק ברמה הגלובלית של המודל כולו אלא גם ברמת התצפית הבודדת, בדומה לאופציות הקיימות במודלים של גרסיה. מדדים כאלה נדרשים לעיתים על ידי גופים רגולטוריים, או אפילו על ידי הלקוחות עצמם, כדי להבין את תוצאות המודל, במיוחד כשמדובר על תחזיות שגויות שכרוכות בנזק רב או בעלות גבוהה. למשל, נדרש להסביר מדוע בקשתו של אדם מסוים לקבל הלוואה אושרה אך בקשתו של אדם אחר נדחתה, להסביר מדוע לקוח מואשם בהונאה, ועוד. נציין שמגבלה זו תקפה לא רק לגבי רשתות נוירונים אלא גם לגבי מודלים "סגורים" שנוצרים בצורה אוטומטית בנישות של למידת מכונה (machine learning) ואינם מאפשרים לראות מהם המשתנים והמאפיינים שהשפיעו על תהליך החיזוי או ללמוד מתוך המודל. הסיבוך של גישות בינה מלאכותית (AI) ולמידת מכונה רק מעודד את מגמת האוטומציה של בניית מודלים, שרובן ככולן מניבות מודלים סגורים. הניסיון בשטח מעיד שמודלים סגורים עשויים להתיע את המשתמשים העסקיים ומקבלי ההחלטות ליישם אותם במציאות, מפני שהמודלים המתקבלים הם לחלוטין לא ברי הבנה ולא אינטואיטיביים. וללא היגיון אנליטי אין כל ערובה שמודלים כאלה לא יטעו מדי פעם ויניבו החלטות שגויות. מכאן יש להניח שגם מודלים המבוססים על רשתות נוירונים מושפעים ממגמה זו.

מציאות זו "הולידה" לאחרונה כיווני מחקר חדשים שמכונים אינטליגנציה מלאכותית מוסברת (Explainable AI), הידועה גם בראשי תיבות XAI. המטרה היא לא רק להסביר מודלים של בינה מלאכותית ולמידת מכונה, אלא גם להסביר את כל התהליך כולו, כולל הנתונים שהביאו לבניית המודל, על מנת להעלות את האמון של מקבלי ההחלטות במודלים השונים ולעודד אותם ליישם בפועל. גישות XAI מציעות מספר מדדים לכמת את יכולת ההסבר של משתנים גם במודלים סגורים של חיזוי, והן נדונות בהרחבה בספרות. סקירה טובה שלהן ניתן למצוא במאמר המקוון של Gandhi (2019) ובמאמר המקוון של Schoenborn and Althoff (2019). דיון יותר

תהליך ההתאמה של רשת נוירונים לנתונים אינו חף מבעיות, וחלקן נובעות מהעובדה שלרשתות נוירונים אין בסיס תיאורטי מוצק. בסעיף זה נדון במספר סוגיות שיש להן השלכות רבות על הבנייה והיישום של רשתות נוירונים לפתרון בעיות עסקיות ומדעיות. סוגיות אלה מצטרפות לכשלים אחרים שנדונו במאמר קודם (זהבי, 2022), ותקפים בדרך כלל לגבי בעיות חיזוי אנליטי. שוב נציין שהדיון הזה תקף רק לבעיות טירגוט בשיווק ישיר, ולא כולן תקפות בהכרח לבעיות חיזוי בתחומים אחרים.

8.1 קופסה שחורה

היעדר בסיס תיאורטי לרשתות נוירונים הוא גם החיסרון הגדול של גישה זו. כאמור, מודלים של חיזוי המבוססים על רשתות נוירונים הם מודלים יוריסטיים שמבוססים במידה רבה על אינטואיציה ועל תהליכי ניסוי וטעייה. כל היתרונות במודלים מבוססי גרסיה לא קיימים ברשת נוירונים. זה אולי נשמע מוזר, אבל למשקלות הענפים ברשת נוירונים אין כל משמעות סטטיסטית והם לא ניתנים לפירוש (Interpretable). התפקיד העיקרי של משקלות הענפים הוא רק לחשב את הערך של משתנה הפלט של כל תצפית בתהליך הזרימה הקדמית. אך טבעי להניח שככל שמשקל הענפים על המסלול ברשת המוביל ממשתנה קלט מסוים, למשל נשים, למשתנה הפלט הוא גבוה יותר, כך הסבירות שנשים יגיבו להצעה לרכוש את המוצר המוצע להן תהיה גבוהה יותר. בפועל, לערך של משתנה הפלט אין שום משמעות הסתברותית או תיאורטית, למעט העובדה שהוא מבטא את דירוג התגובה של הצרכנים. ככל שערך משתנה הפלט גבוה יותר, סבירות התגובה גבוהה יותר. כתוצאה מכך לא ניתן לייחס משמעות תיאורטית או סטטיסטית לתוצאות המבוססות על רשתות נוירונים. אי אפשר להסביר את המשקלות של הענפים שמתקבלים בתהליך הלמידה, להעריך את החשיבות היחסית של צומתי הקלט, להסביר מדוע טופולוגיה מסוימת של הרשת עובדת טוב יותר מטופולוגיה אחרת או להעריך את האמינות של תוצאות המודל. נציין שמסקנה זו תקפה לא רק לבעיות חיזוי בתחום של שיווק ישיר, אלא גם בתחומים רבים אחרים.

ספציפי הנוגע לרשתות נזירונים מופיע במאמר המקוון של Vaughan et al. (2018).

ומכאן, אולי יותר מאשר בכל מודל חיזוי אחר, יש חשיבות רבה לאשש מודלים המבוססים על רשתות נזירונים ולתקף אותם על מנת לשכנע את מקבלי ההחלטות להשתמש בהם. השוואה של תוצאות מודל רשתות נזירונים למודל חיזוי אחר, וניסיון להסביר את תוצאות המודל תוך שימוש בגישות מהתחום של *Explainable AI*, יכולים בוודאי לעזור להעלות את האמון של מקבלי ההחלטות במודלים המבוססים על רשתות נזירונים.

8.2 מספר רב של משתנים מסבירים

בבעיות שיווק מבסיסי נתונים מספר המשתנים הפוטנציאליים שעשוי להסביר את תגובת הצרכנים הוא גדול במיוחד, ובמקרים רבים עשוי להגיע לכמה מאות, אם לא יותר. בבעיית טירגוט אופיינית המשתנים המסבירים האפשריים מורכבים מהיסטוריית הקניות של הלקוח בעבר, היסטוריית הפניות שנעשו ללקוח בעבר, היסטוריית האשראי בעבר, משתנים דמוגרפיים של הלקוח, מאפייני המוצר/שירות המוצע ועוד, שמספרם יכול להגיע בקלות לכמה מאות, וזה עוד לפני שהוספנו טרנספורמציות אפשריות של משתנים מסבירים ומשתנים המבטאים יחסים לא ליניאריים ואינטראקציות בין משתנים. למשל, כדי להזין למודל חיזוי משתנה קטגורי כגון מצב משפחתי (Marital status) נדרשים ארבעה משתנים בינריים (0/1), משתנה אחד לכל ערך של המשתנה – נשוי(אה), רווק(ה), אלמן(ה) גרוש(ה) – וכמות המשתנים הקטגוריים במודל שיווק בדרך כלל גבוהה מאוד. גם משתנים רציפים, כגון גיל הלקוח, נהוג לעיתים לייצג כמשתנה קטגורי, למשל צעיר, מבוגר, ותיק, ולייצג כל קבוצה באמצעות משתנה מסביר נפרד. כל טרנספורמציה כזו רק מגדילה את מספר משתני הקלט הנכנסים למודל.

כאמור לעיל, רשת נזירונים בבעיות טירגוט מכילה צומת קלט אחד לכל משתנה מסביר, מה שאומר שמספר הצמתים בשכבת הקלט עשוי להגיע לכמה מאות ויותר. כתוצאה מכך, מרחב המצבים של הבעיה (מספר הקונפיגורציות האפשריות של משתני הקלט) הוא כמעט אינסופי. למשל, עבור בעיה "קטנה" עם 30 משתנים מסבירים בינריים בלבד מרחב המצבים מכיל 2^{30} מצבים אפשריים (שזה מספר ענק). עם מרחב מצבים כמעט אינסופי כזה, קובץ האימון שמשמש לבניית

המודל, עד כמה שהוא גדול, מכיל רק חלק זעור של מרחב המצבים האפשרי, מה שמשבש את תהליך הלמידה שנחשף רק למספר קטן מאוד של מצבים אפשריים, לעיתים פחות משבריר של אחוז. התוצאה היא משקלות ענפים מעוותים וסיכון גבוה במיוחד להתאמת יתר.

הפתרון לבעיה חמורה זאת היא להקטין בצורה משמעותית את מספר משתני הקלט שנכנסים למודל. בשל העובדה שרשתות נזירונים חסרות תשתית תיאורטית, אין לנו מדדים כגון רמת מובהקות וערכי p-value כמו ברגרסיה ליניארית, שמאפשרים לסנן מראש משתני קלט לא רלוונטיים תוך כדי התהליך של בניית המודל. אז מה הפתרון לבעיה זו? מבחינה מעשית, הדרך המומלצת ביותר היא לבצע את בחירת משתני הקלט למודל הרשתות הנזירונים מחוץ למודל, באמצעות עיבוד מקדים (Preprocessing), למשל סינון משתנים לא רלוונטיים תוך שימוש בגישות סטטיסטיות סטנדרטיות לניתוח נתונים (EDA – Exploratory Data Analysis). ניתן גם להפעיל שיקולים אינטואיטיביים המבוססים על תחום הידע (Domain knowledge) על מנת לסנן משתני קלט לא רלוונטיים מהמודל. גישה יותר מתקדמת היא באמצעות ריצה מקדימה של מודל רגרסיה (ליניארית או לוגיסטית) על נתוני הבעיה המקורית, ואז לבחור רק את המשתנים המסבירים שנבחרו על ידי המודל המקדים בתור משתני הקלט למודל רשתות הנזירונים. נציין שלתהליך עיבוד מקדים המבוסס על מודל רגרסיה יש גם יתרונות נוספים, שכן מחד ניתן להשתמש במקדמים של מודל הרגרסיה בתור המשקלות הראשוניים של משתני הקלט במודל רשתות נזירונים, ומאידך מודל הרגרסיה מספק פתרון אלטרנטיבי שניתן להשוותו עם הפתרון של מודל רשתות הנתונים. יתרונות אלה עונים על שתי הבעיות שדנו בהן בסעיפים הקודמים לגבי מודלים מבוססי רשתות נזירונים. מצד אחד מציאת משקלות התחלתיים "מתוחכמים" יותר שיעלו את תהליך הפתרון של רשת הנזירונים, ומצד שני לקבל פתרון ייחוס (Reference solution) נוסף למודל רשתות הנזירונים שלפיו ניתן לקבוע האם הפתרון הנוכחי של רשת הנזירונים התכנס לפתרון "סביר" או שיש להמשיך את תהליך החיפוש על מנת למצוא פתרון טוב יותר.

8.3 שיעור תגובה (Response rate) נמוך

במיוחד, יש לשקול את האפשרות להעדיף מודל תגובה אחר על פני מודל המתבסס על רשתות נוירונים.

8.4 סוגיות זמן (Timing issues)

זכור, בבעיות שיווק ישיר אנו מסתמכים על דוגמאות מהעבר (לקוחות שהשתתפו במבצעי שיווק קודמים) כדי לבנות מודל שמבדיל בין המגיבים ללא מגיבים. את המודל הזה מיישמים על לקוחות חדשים או לקוחות שלא נטלו חלק במדגם האימון, על מנת לאתר את הלקוחות הרווחים שכדאי לשתף במבצע השיווק ולסנן את הלקוחות הלא רווחים. אולם בדרך כלל יש פער זמן בין המועד של בניית המודל לבין יישומו בשטח. יש לכך מספר סיבות, ואחת מהן היא הצורך להצטייד במלאי מספיק של מוצרים כדי לספק את הביקוש שיווצר על ידי מבצע השיווק הישיר. אם פער הזמן הזה ארוך במיוחד, כמה חודשים ויותר, הרכב הלקוחות בבסיס הנתונים והמאפיינים שלהם בזמן היישום עשוי להשתנות, מה שעשוי להפוך את מודל התגובה המבוסס על רשתות נוירונים למיושן מדי ולא רלוונטי. הפתרון הוא לנסות לחזות בזמן בניית המודל מה יהיה הרכב אוכלוסיית היעד בזמן היישום של המודל ולבנות מדגם אימון שמשקף את האוכלוסייה הזו, או לבנות את מודל רשתות הנוירונים מחדש סמוך למועד היישום על בסיס מדגם אימון שנבחר מתוך אוכלוסיית היעד. שתי האפשרויות מסרבלות את תהליך החיזוי וכרוכות במאמץ חישובי רב.

ואכן, לא בכדי רשתות נוירונים מתאימות יותר לבעיות היכן שהאירועים קורים באופן רציף (*Continuous*), כמו למשל בבעיות חיזוי של פשיטות רגל (*Bankruptcies*), כשלים של בנקים (*Bank failure*), בעיות נטישה (*Churning*), ובעיות חיזוי אחרות שמתרחשות על פני זמן. בבעיות אלה יש אפשרות לעדכן את מדגם האימון באופן שוטף על ידי הוספת אירועים חדשים ואולי הסרה של אירועים ישנים ויצירה של מעין "חלון נע" (*Sliding window*) שבו מתרחש תהליך המידול וקבלת ההחלטות. גישת החלון הנע מאפשרת להתאים באופן דינמי את תחזיות המודל כאשר מגיעות תצפיות חדשות. אולי זו הסיבה לשימוש הרחב יחסית של רשתות נוירונים בתחום של חיזוי סדרות עיתיות (*Time series forecasting*) (Tang et al., 1992).

כאמור, בבעיות שיווק ישיר יש פער זמן בין מועד בניית המודל ובין מועד המימוש שלו, עובדה שיש לה השלכות גם על אופן

ככל ששיעור התגובה (Response rate), המבטא את היחס בין מספר המגיבים ("היקונים") לכלל האוכלוסייה גבוה יותר, כך תהליך הלמידה מסתמך על יותר דוגמאות של לקוחות מגיבים ולכן מסוגל להבחין טוב יותר בין המאפיינים של המגיבים ללא מגיבים. אלא ששיעורי התגובה במבצעי שיווק ישירים הם בדרך כלל קטנים במיוחד. בשיווק ישיר דרך הדואר, למשל, שיעורי התגובה הם לרוב פחות מאחוז אחד (ולעיתים גם פחות מרבע אחוז), ובשיווק אינטרנטי שיעור התגובה עוד יותר נמוך מזה. שיעורי תגובה נמוכים פוגעים ביכולתו של מודל המבוסס על רשתות נוירונים לאפיין את הלקוחות המגיבים מהלא מגיבים. אם נסייף לכך גם את העובדה שמרחב המצבים האפשריים בבעיות שיווק ישיר שואף לאינסוף, ואת בעיות ההתכנסות שנידונו לעיל, זה לא מפתיע שבמקרים רבים מודל רשתות הנוירונים עשוי לסווג את כל הלקוחות החדשים בתור לא מגיבים.

הפתרון לבעיה הזו הוא לכלול במדגם האימון פרופורציה גדולה יותר של מגיבים מאשר הפרופורציה שלהם באוכלוסייה, לעיתים אף את כל המגיבים באוכלוסייה ורק מדגם קטן של לא מגיבים. בהקשר של בעיות בשיווק, מדגמים אלה נקראים מדגמים מבוססי־בחירה (*Choice based samples*), שכן הם מבוססים על הבחירה (*choice*) של הלקוח האם להיענות או לסרב להצעה לרכוש את המוצר או השירות המוצע (Ben Akiva & Lerman, 1987). אולם מודל תגובה שמתבסס על מדגמים מבוססי־בחירה אינו מייצג את כלל אוכלוסיית היעד אלא רק את אוכלוסיית המדגם ולכן עשוי לסווג גם לקוחות לא מגיבים כמגיבים, מה שעשוי להוביל להחלטות מוטעות. הפתרון הוא לעדכן את משקלות הענפים של רשת הנוירונים כך שתשקף בצורה מדויקת יותר את יחס המגיבים/לא מגיבים באוכלוסייה. במודל רגרסיה לוגיסטית התהליך פשוט יותר, שכן למעט החותך (*Intercept*) של משוואת הרגרסיה, כל יתר המקדמים במשוואת הרגרסיה לא משתנים, מה שמחייב לעדכן רק את החותך של המשוואה (Ben Akiva & Lerman, 1987). במודלים אחרים העדכון מורכב יותר, אבל כמובן שאי אפשר להתעלם מהתופעה הזו כדי להימנע ממודלים מוטעים. באשר לרשתות נוירונים, מאחר שאין להם בסיס תיאורטי שעל פיו ניתן לעדכן את משקלות הענפים בפתרון הסופי, לא מן הנמנע שבמקרים שבהם שיעורי התגובה באוכלוסייה נמוכים

בניית המודל וגם על היישום שלו. תופעה זו מאפיינת לא רק מודלים מבוססי רשתות נוירונים אלא גם מודלים אחרים של חיזוי בבעיות שבהן אין חפיפה בין מועד בניית המודל והמימוש שלו.

8.5 מתן ציון לצומת הפלט בפתרון הסופי

נזכיר שבניית מודל נוירונים אינה מטרה בפני עצמה, אלא המטרה לבנות מודל שניתן להשתמש בו לצורך קבלת החלטות. בבעיות שיווק ישיר שבהן אנו עוסקים, המטרה היא ליישם את המודל על לקוחות חדשים או על לקוחות שלא לקחו חלק במדגם האימון ששימש לבניית המודל, על מנת לאתר את הלקוחות שכדאי לשתף במבצע השיווק. החלטות אלה מתבססות על הציון (Score) שהלקוח מקבל בפתרון הסופי של מודל רשתות הנוירונים. בהינתן המשקלות הסופיים של הענפים ברשת הנוירונים, מחשבים את הציון הזה על ידי הפעלת תהליך ההזנה הקדמית על כל צרכן, ממשתני הקלט עד למשתנה הפלט. עבור פונקציית אקטיבציה סיגמואידלית, הערך החזוי של משתנה הפלט הוא מספר בין 0 ל-1. אולם בעיות שיווק ישיר הן בעיות קלסיפיקציה שבהן משתנה הפלט בפועל לכל לקוח הוא 0/1 – (כן) אם הלקוח הניב ורכש את המוצר/שירות המוצע לו, 0 (לא) אם הלקוח לא הניב לפנייה אליו או דחה את הצעת הרכישה. השאלה היא כיצד לסווג את משתנה הפלט במודל הסופי, האם כ-0 או כ-1? לסיווג הזה יש משמעות מעשית, שכן סיווג של 1 משמעו שיש לשתף את הלקוח במבצע השיווק, וסיווג של 0 אומר שאין לשתף את הלקוח במבצע השיווק.

הדעות בספרות חלוקות באשר לאופן הסיווג של משתנה הפלט בפתרון הסופי של מודל רשתות הנוירונים. Tam & Kiang (1992) הציעו את הערך 0.5 בתור נקודת החיתוך, כלומר אם הערך של צומת הפלט בפתרון של רשת הנוירונים גדול מ-0.5, צומת הפלט יקבל את הערך 1, ואחרת הוא יקבל את הערך 0. אבל מכיוון שמספר הלא מניבים בבעיות שיווק ישיר גדול באופן משמעותי מאשר מספר המניבים, לא סביר שהערך של צומת הפלט, אפילו עבור לקוחות מניבים, יהיה גדול מ-0.5, וככל הנראה יהיה קרוב ל-0. לכן שימוש בקריטריון של 0.5 עשוי לסווג את כל הלקוחות כלא מניבים וימליץ שלא לכלול אף אחד במבצע השיווק! אפשרות אחרת היא לבחור קריטריון שרירותי אחר, למשל 0.1, ולשתף כל

לקוח שעבורו הערך של צומת הפלט גדול מ-0.1 במבצע השיווק, ואחר כך לתקף את הערך הזה בדיעבד באמצעות תהליך של בדיקה לאחור (Backtesting). אבל מי ערב לכך שהערך הזה יהיה תקף גם למבצעי שיווק אחרים?

אפשרות נוספת היא להסתמך על תוצאות המודל עבור מדגם האימון, למיין את הלקוחות בסדר יורד של הציונים של משתנה הפלט, ולהשתמש בנקודת החיתוך שמשיאה את שיעור התגובה בקובץ האימון בתור הקריטריון למבצע השיווק. למשל, אם שיעור התגובה המרבי עבור מדגם האימון מתקבל באחוזון ה-62, אזי יש לשתף את כל הלקוחות החדשים הנמצאים באחוזון ה-62 ומעלה במבצע השיווק. אלטרנטיבה פשוטה יותר היא לפנות רק לעשירונים העליונים (בדרך כלל 4 או 5 עשירונים), שעל פי מדגם האימון שיעורי התגובה שלהם הם הגבוהים ביותר. נציין ששתי חלופות אלה הן לא הפתרון האידיאלי, שכן האחוזונים במדגם האימון מתבססים על מדגם האימון שעבורו בדרך כלל מתקבלת התאמה טובה של המודל לנתוני הקלט. עם זאת, בהיעדר קריטריון טוב יותר זוהי פשרה סבירה, במיוחד כי מדובר כאן על נקודת חיתוך תלוית מוצר.

נציין שמבחן התוצאה הקובע את איכות מודל רשתות הנוירונים מתבסס גם הוא על ההחלטה האם לסווג את צומת הפלט בפתרון הסופי של רשתות הנוירונים בתור 0 או 1. לכן שימוש בקריטריון הסיווג הנכון הוא קריטי גם על מנת להעריך נכון את התוצאות העסקיות והרווחיות המבוססות על מודל רשתות נוירונים.

כמובן שהדרך הנכונה ביותר היא להשתמש בשיקולים כלכליים לצורך הסיווג הסופי של הלקוחות, אבל לשם כך יש להמיר את משתנה הפלט ממודל רשתות הנוירונים להסתברויות. אנו נדון בנושא זה בהקשר של בעיות חיזוי בסעיף הבא.

8.6 חיזוי מספר ההזמנות

כאמור, ברוב המקרים יש פער זמן די ארוך בין מועד בניית המודל ובין מועד יישומו לצורך איתור לקוחות למבצע השיווק. אחת הסיבות לפער הזה הוא לאפשר לארגון השיווק להצטייד במלאי שיוכל לספק את הכמות המבוקשת של המוצר המוצע. בדרך כלל חברת שיווק לא תצא במבצע שיווק ישיר נרחב לפני שהיא מבטיחה שיש לה את היכולת לספק את כל ההזמנות למוצר. חיזוי מדויק של מספר ההזמנות הצפוי מתקבל בדרך

התיקון הוא קבוע בלי קשר לגודל הציון של הלקוח, או בצורה לא ליניארית, שהיא צורה יותר מדויקת מכיוון שמקדם התיקון משתנה בהתאם לרמת הציון של הלקוח.

ואם כבר נרמלנו את הציונים המגיעים ממודל רשתות הניורונים כך שיבטאו הסתברויות רכישה, גם אם מדובר בקרוב שלהם, ניתן להפעיל שיקולים כלכליים כדי להחליט אילו לקוחות חדשים לשתף במבצע השיווק הישיר. הקריטריון המנחה הוא לשתף צרכנים שתוחלת הרווחים מהם גבוהה מעלות הפנייה אליהם.

ובכל זאת, מאחר שמדובר כאן בקרובים, יש צורך לבדוק בכל מקרה לגופו, בדיעבד, מה מידת הדיוק של הקרובים האלה והאם התחזיות וההמלצות המתבססות על ההסתברויות המקורבות הן אכן נכונות ומביאות למבצעי שיווק רווחיים.

9. יישום של מודל מבוסס רשתות נוירונים

לבסוף נדון בסוגיות המתייחסות ליישום של מודלים מבוססי רשתות נוירונים לקבלת החלטות עסקיות, ושוב נשתמש בבעיית שיווק ישיר על מנת להדגים את הנושא. היישום שבו השתמשנו כדי להדגים את התוצאות של מודל חיזוי הוא מועדון ספרים, *BBB (BookBinders Book Club)*, שמפיץ ספרים בנישה של שיווק ישיר. מועדון הספרים מחזיק בסיס נתונים גדול של לקוחותיו ופונה אליהם בדרך כלל מדי חודש בהצעה לרכוש ספר "תורן" שיוצא לאור. אבל מאחר שרק מיעוטם של הלקוחות מגיבים להצעה לרכוש את הספר המוצע, מועדון הספרים משתמש בנישות של חיזוי אנליטי על מנת לפנות רק ללקוחות עם סבירות גבוהה לרכוש את המוצר ו"לסנן" ממבצע השיווק את הלקוחות שלא סביר שיענו להצעה. היישום המתואר להלן מתבסס על מדגם למידה המכיל 50,000 לקוחות שנוצר באופן סינטיטי ומדמה את הנתונים האמיתיים של לקוחות במועדוני ספרים. קובץ זה משמש כיום כבסיס לאירוע לימוד (*Case study*) בקורסי שיווק ומדעי הנתונים במספר אוניברסיטאות מובילות בעולם, כגון *Harvard*¹ ואחרות. תיאור נרחב של האירוע מופיע אצל Gary et al. (2007).

כלל במודלים הסתברותיים, כגון מודל רגרסיה לוגיסטית, שבהם הציונים של המודל מבטאים את הסתברות הרכישה של הלקוחות. אפשר להסתכל על הסתברות הרכישה כעל החלק (*Fraction*) של המוצר שנרכש על ידי הלקוח. למשל, חמישה לקוחות שהסתברות הרכישה של כל אחד מהם הוא 0.2 יזמינו בממוצע יחידה אחת של המוצר המוצע. לכן מספר ההזמנות הצפוי ממבצע שיווק ישיר במודל ההסתברותי מתקבל על ידי סכום הסתברויות הרכישה של כל הצרכנים הלקוחים חלק במבצע השיווק. מסקנה זו מעוגנת גם במשפט מתמטי שבמודל רגרסיה לוגיסטית סכום הסתברויות הרכישה של כל הלקוחות במדגם האימון שווה לסה"כ מספר ההזמנות בפועל (Ben Akiva & Lerman, 1987).

אלא שציוני הלקוחות במודל רשתות נוירונים לא מייצגים את הסתברויות הרכישה של הלקוחות אלא את דירוג הלקוחות (*Ranking*) המבטא את הסבירות (*Likelihood*) שלהם לרכוש את המוצר המוצע, ככל שהדירוג גבוה יותר, סבירות הרכישה גבוהה יותר. מכאן שאי אפשר להשתמש במשתנה הדירוג על מנת לחזות את מספר ההזמנות במבצעי השיווק, אפילו אם פונקציית האקטיבציה היא לוגיסטית, שכן אי אפשר להתייחס לציוני הלקוחות, על אף שהם בתחום (0,1), כאל הסתברויות רכישה. על אחת כמה וכמה אם פונקציית האקטיבציה היא *Hyperbolic tangent* ומבטאת את התחזיות של משתנה התוצאה בתחום (-1,+1).

מכאן שעל מנת שאפשר יהיה להשתמש בציונים של רשתות הניורונים גם לצורכי חיזוי, יש "להפוך" אותם להסתברויות. הרעיון הוא לנרמל את הציונים של רשתות הניורונים כך שסכום הציונים המנורמלים של רשת הניורונים (שנקרא להם *NN scores*) על פני כל הלקוחות במדגם האימון יהיה שווה למספר ההזמנות בפועל של הלקוחות האלה. כלומר:

$$\sum_{i=1}^N p_i = \sum_{i=1}^N y_i$$

כאשר:

y_i – הבחירה של צרכן i (1 – הזמין את המוצר, 0 – לא הזמין)
 p_i – *NN score* של צרכן i
 N – מספר הצרכנים בקובץ האימון

אפשר לנרמל את הציונים של מודל הרשתות הניורונים בשתי צורות: בצורה ליניארית, שהחיסרון העיקרי שלה הוא שמקדם

1 Harvard Case Study involving the BBB, <https://www.casequiz.com/bookbinders-book-club-125140/>

טבלה 2: טבלת רווחים ממודל רשתות הניירונים על קובץ התיקוף.

Cust. Rank(%)	# Cust.	% Cust.	# Resp.	% Resp.	Actual RR(%)	%Resp./%Cust.	Pred. Resp.	Pred. RR(%)
24.49	1675	10.2	622	40.7	37.13	4.1	683	40.80
11.49	1662	10.1	284	18.6	17.09	1.8	279	16.76
6.96	1664	9.9	190	12.4	11.70	1.3	146	9.00
4.71	1628	9.9	124	8.1	7.62	0.8	93	5.70
3.76	1624	9.9	89	5.8	5.48	0.6	69	4.25
3.02	1612	9.8	73	4.8	4.53	0.5	55	3.40
2.57	1604	9.7	59	3.9	3.68	0.4	45	2.78
2.27	1751	10.6	42	2.7	2.40	0.3	42	2.41
2.00	1645	10.0	31	2.0	1.88	0.2	35	2.15
1.19	1642	10.0	15	1.0	0.91	0.1	31	1.88

ליישום גם על לקוחות חדשים. במקרה שלנו ביצענו מספר ניסיונות עם קונפיגורציות שונות של רשתות ניירונים על מנת למצוא פתרון מיטבי לבעיית השייך. הפתרון שלהלן מתבסס על הפרמטרים הבאים:

- שכבה נסתרת אחת עם שלושה צמתים.
- מספר מקסימלי של חזרות בכל מחזור אימון – 10,000.
- מספר מחזורי האימון – 10, כאשר כל מחזור אימון מתחיל עם גרעין אקראי (*Random seed*) חדש.
- פונקציית אקטיבציה סיגמואידלית.

מספר משתני הקלט לאחר הטרנספורמציות טיפס לכמה עשרות. אי לכך הפעלנו תהליך שהתבסס על ניתוח נתונים (*Exploratory data analysis*) על מנת לסנן מראש משתנים מסבירים שעל פניו נראו לא חשובים, והוריד את מספר משתני הקלט שנכנסו למודל לסדר גודל של 30 משתנים.

טבלה 2 מציגה את תוצאות המודל באמצעות טבלת רווחים (*Gains table*) על קובץ התיקוף בסדר יורד של העשירונים, מהעשירון הגבוה ביותר ועד לעשירון הנמוך ביותר. על מנת לקבל את טבלת הרווחים ממיינים את הלקוחות בסדר יורד של תחזית הציונים של הלקוחות, ובמקרה של רשתות ניירונים מבטאים את דירוג הלקוחות לפי הסבירות שהם ייענו להצעה לרכוש את המוצר המוצע להם (בדוגמה שלנו הספר "היסטוריית האומנות של פירנצה"), מהציון הגבוה לנמוך, ומסכמים את הנתונים ברמה של עשירונים או כל אחוזון אחר. סידור הלקוחות בסדר יורד של משתנה הדירוג

ביישום הנוכחי, הספר המוצע (המשתנה התלוי) הוא "היסטוריית האומנות של פירנצה" (*The Art History of Florence*) והמשתנים המסבירים מכילים מספר פרטים דמוגרפיים של הלקוח בתוספת מידע על קטגוריות הספרים שהלקוח קנה דרך מועדון הספרים בעבר (למשל, מספר ספרי ילדים שקנה בעבר, מספר ספרי אומנות, מספר ספרי גיאוגרפיה, ועוד), וכן מידע על מועד ההצטרפות של הלקוח למועדון הספרים, כמה זמן עבר מאז הקנייה האחרונה, ומה סכום הכסף שהלקוח הוציא על קניות במועדון הספרים בעבר. על מנת לבחון את תגובת הצרכנים לספר החדש, המועדון הציע ל-50,000 חברי המועדון, שמהווים מדגם מקרי של מיליוני החברים בבסיס הנתונים, לרכוש את הספר "היסטוריית האומנות של פירנצה". את הקובץ הזה חילקנו באקראי לשני קבצים ביחס של 2/3 למדגם האימון ו-1/3 למדגם התיקוף. מדגם האימון שימש לבניית המודל ומנה 33,533 לקוחות שמהם 2,993 מגיבים (שיעור תגובה של 8.93%). מדגם התיקוף שימש לצורך אימות ובדיקה של המודל ומנה 16,647 לקוחות שמהם 1,529 מגיבים (שיעור תגובה של 9.29%).

מבחינה מעשית, העובדה שאין לרשתות ניירונים תשתית תיאורטית היא גם יתרון וגם חיסרון של גישת רשתות הניירונים. זה יתרון מכיוון שעובדה זו לא מחייבת הנחות מקדימות כדי לבנות את המודל. וזה חיסרון משום שבניית המודל מתבססת על תהליך חיפוש מייגע של ניסוי וטעייה על מנת למצוא פתרון מיטבי שגם מצליח להבדיל בין הלקוחות ה"טובים" וה"לא טובים" וגם מניב פתרון יציב ללא התאמת יתר שניתן

טבלה 3: מדדי ביצוע בין קובץ האימון לבין קובץ התיקוף

Decile	% Resp. TRN	% Resp. VLD	%Resp./%Cust. TRN	%Resp./%Cust. VLD	Actual RR(%) TRN	Actual RR(%) VLD	Pred. RR(%) TRN	Pred. RR(%) VLD
1	43.6	40.6	4.4	4.1	38.94	37.70	41.14	41.07
2	18.9	18.4	1.9	1.8	16.91	17.06	16.54	17.03
3	11.9	12.6	1.2	1.3	10.62	11.72	9.01	9.09
4	7.7	8.1	0.8	0.8	6.86	7.53	5.72	5.72
5	6.1	6.0	0.6	0.6	5.43	5.59	4.23	4.24
6	4.0	4.8	0.4	0.5	3.61	4.43	3.30	3.38
7	3.1	3.9	0.3	0.4	2.77	3.58	2.78	2.75
8	2.3	2.6	0.2	0.2	2.03	2.43	2.43	2.40
9	1.6	2.0	0.2	0.2	1.40	1.88	2.15	2.15
10	0.8	1.0	0.1	0.1	0.69	0.91	1.87	1.87

בפועל בעשירון העליון הוא 622 ומספר המגיבים החזוי הוא 683. בעשירון השני מספר המגיבים בפועל הוא 284, שדומה לתחזית המודל (279). וכך הלאה. אי אפשר כמובן לצפות לזהות מלאה ברמת העשירונים בין מספר המגיבים בפועל לתחזית מספר המגיבים, אבל סטייה של התחזית בסדר גודל של פלוס/מינוס 5-10% נראית סבירה.

טבלה 3 מציגה את תוצאות התיקוף של המודל הנ"ל באמצעות השוואה של תוצאות המודל בין קובץ האימון לקובץ התיקוף. בטבלה 3 אנו מציגים מספר מדדי ביצוע של המודל, זה בצד זה, כפונקציה של העשירונים, כאשר TRN מסמן את קובץ האימון (Training) ו-VLD את קובץ האימות (Validation).

נציין שמבחינה מעשית המדדים החשובים לצורך קבלת החלטות הם המדדים שמתייחסים לקובץ התיקוף, שכן קובץ זה מייצג את היילקוחות החדשים, כלומר את האוכלוסייה שלא השתתפה בבניית המודל שמתוכה אנו רוצים לבחור את הלקוחות שישתתפו במבצע השיווק המלא. כמובן שאין לצפות לשוויון מלא של מדדי הביצוע בין מדגם האימון והתיקוף, אבל ככל שההתאמה בין מדדי הביצוע טובה יותר, כך המודל תקף יותר. ואומנם, בחינה של טבלה 3 מראה שמדדי הביצוע על קובץ התיקוף תואמים פחות או יותר (בטווח הרצוי של 5%-10%) למדדי הביצוע של קובץ האימון, מה שמעיד שהמודל המבוסס על רשתות הניורונים הוא אכן מודל יציב שניתן ליישמו לקבלת החלטות.

שים את הלקוחות הטובים עם הסבירות הנבוהה לקנות את המוצר בראש הרשימה, ואת הלקוחות הפחות טובים בתחתית הרשימה. תחזיות הדירוג בפונקציית אקטיבציה סינמאידלית הן בתחום (0-1) וביטאנו אותם בטבלה 2 באחוזים באמצעות הפרמטר (Cust. Rank %). למשל, הדירוג של כל הלקוחות בעשירון העליון הוא 24.49% ומעלה, הדירוג של כל הלקוחות בעשירון השני הוא בין 11.49% ל-24.49%, וכך הלאה.

מתוך הטבלה קל לראות את טיב התחזית של המודל באמצעות מספר מדדים. לדוגמה, מספר התגובות בפועל (# Resp.) הוא פונקציה מונוטונית יורדת כשנעים על פני העשירונים מהעשירון העליון לתחתון, מה שאומר שהמודל אכן מצליח למקם את הלקוחות בעשירונים העליונים של טבלת הרווחים. בעשירון העליון מספר המגיבים בפועל הוא 622 המהווים 40.7% (Resp %) מכלל המגיבים עם שיעור תגובה בפועל (Actual RR%) של 37.13%, לעומת 15 מגיבים בלבד בעשירון התחתון המהווים רק 1% מכלל המגיבים ושיעור תגובה של 0.91%.

תופעה דומה קורית גם כשמסתכלים על תחזית מספר התגובות של המודל. כך למשל, המודל צופה 683 מגיבים (Pred. Resp.) בעשירון העליון, כלומר שיעור תגובה חזוי (Pred. Resp. %) של 40.80%, לעומת 31 מגיבים בלבד בעשירון התחתון עם שיעור תגובה חזוי של 1.88% מכלל מספר המגיבים הצפוי. מדד נוסף לאיכות החזוי של המודל הוא ההשוואה בין מספר המגיבים בפועל (# Resp.) לתחזית מספר המגיבים (Pred. Resp. %). כך למשל, מספר המגיבים

טבלה 4: טבלת רווחים ממוזל רגרסיה לוגיסטית על קובץ התיקוף

Resp. Prob.(%)	# Cust.	% Cust.	# Resp.	% Resp.	Actual RR (%)	%Resp./%Cust.	Pred. Resp.	Pred. RR (%)
21.70	1689	10.3	671	43.9	39.73	4.3	686	40.60
12.25	1653	10.0	267	17.5	16.15	1.7	267	16.13
8.18	1603	9.7	166	10.9	10.36	1.1	162	10.12
5.90	1679	10.2	137	9.0	8.16	0.9	116	6.92
4.32	1619	9.8	92	6.0	5.68	0.6	82	5.07
3.27	1611	9.8	62	4.1	3.85	0.4	61	3.79
2.35	1645	10.0	43	2.8	2.61	0.3	45	2.73
1.62	1621	9.8	46	3.0	2.84	0.3	32	1.96
0.89	1706	10.4	31	2.0	1.82	0.2	22	1.28
0.02	1641	10.0	14	0.9	0.85	0.1	8	0.51

טבלה 5: השוואה בין מודל רשתות הניורונים למודל הרגרסיה הלוגיסטית

Actual Pred.

	#Cust.	#Resp.	RR	#Pred.	RR
<i>NN</i>	6629	1220	18.40%	1201	18.12%
<i>LogReG</i>	6624	1241	18.73%	1231	18.58%

המבוסס על עשירונים, כלומר פנייה רק ללקוחות הנמצאים בארבעת העשירונים העליונים, שלפי שני המודלים מכילים את מרב הקונים. ריכזנו את התוצאות בטבלה 5.

מטבלה 5 עולה שלמודל הרגרסיה הלוגיסטית יש יתרון קל על מודל רשתות הניורונים, שכן המודל "תופס" יותר מגיבים בארבעת העשירונים העליונים. גם תחזיות המודל הלוגיסטי לגבי מספר ההזמנות הצפוי הן קצת יותר טובות. העובדה שלמודל מבוסס רגרסיה לוגיסטית יש יתרון על פני מודל מבוסס רשתות ניורונים בבעיות שיווק ישיר היא לא חדשה. מסקנה דומה התקבלה גם במחקר המקיף של Levin & Zahavi (1997). ככל הנראה הסיבה העיקרית לכך היא שהחלטות הקנייה של צרכנים בעולם השיווק הן עקביות ורציונליות. התנהגות רציונלית זו, כשהיא באה לידי ביטוי בקובצי הנתונים, גורמת לכך שקובצי הנתונים בשיווק מתנהגים היטב ללא קשרים לא לינאריים ו/או קשרים מורכבים אחרים בין המשתנים המסבירים למשתנה התלוי (כגון קשרים מסוג XOR), שעבורם יש יתרון למודל מבוסס רשתות ניורונים. שוב נדגיש שהמסקנה שלמודל רגרסיה לוגיסטית יש עדיפות על מודל מבוסס רשתות ניורונים תקפה רק לתחום השיווק, ולא בהכרח תקפה לגבי תחומים אחרים.

ולבסוף, השווינו את תוצאות המודל המבוסס על רשתות ניורונים לתוצאות של המודל המבוסס על רגרסיה לוגיסטית. לצורך ההשוואה הרצנו את מודל הרגרסיה הלוגיסטית על אותו בסיס נתונים של מודל רשתות הניורונים עם אותם פרמטרים. בשני המודלים הפעלנו את אותו התהליך של עיבוד מקדים (Preprocessing) על מנת לנפות מראש משתנים לא משמעותיים, אבל במקרה של רגרסיה לוגיסטית המשכנו אחר כך לנפות משתנים לא מובהקים, מה שלא יכולנו לעשות במודל המבוסס על רשתות ניורונים בשל היעדר תשתית תיאורטית. תוצאות המודל מוצגות בטבלה 4.

טבלה 4 זהה במבנה שלה לטבלה 2, למעט העובדה שבמודל הלוגיסטי התצפיות ממוינות בסדר יורד של תחזית הסתברות הרכישה ((Resp. Prob. (%)). מהתבוננות בטבלה 4 נראה שגם מודל הרגרסיה הלוגיסטית עושה הבחנה טובה בין הלקוחות המגיבים ללא מגיבים, שכן מרבית ה"קונים" מרוכזים בעשירונים העליונים של הטבלה ורק מיעוטם בעשירונים התחתונים.

כדי להשוות את התוצאות בין שני המודלים אנחנו זקוקים לקריטריון שמבדיל בין ה"קונים" ל"לא קונים" ובחרנו בקריטריון

טבלה 6: יישומים של למידה עמוקה בתחומי מחקר שונים בתעשייה

תחום מחקר	פעולה	שימושים בתעשייה
ניתוח נתונים	עיבוד נתונים לקבלת החלטות עסקיות	שיווק ופרסום, פיננסים, סייבר, ניהול ארגוני, מדעים מדויקים
זיהוי קול	הבנה וניתוח של הדיבור	מערכות <i>Personal Assistant</i> כגון <i>Siri</i> , מערכות <i>IoT</i> כגון <i>Google Home</i> וכי
למידה מטקסט	למידת שפה – קריאה וכתובה	תרגום אוטומטי, הבנת סנטימנט שיח ברשתות החברתיות, תמיכה בשירות לקוחות על ידי ציט בוטים
עיבוד תמונה	ראייה ממוחשבת	רשתות חברתיות, רפואה (רדילוגיה), רכב אוטונומי, רובוטיקה

10. הרחבות ללמידה עמוקה

לסיום נציין שהיכולת המרשימה של רשתות נוירונים לזהות באופן אוטומטי תבניות (*Patterns*) מורכבות בתוך כמויות אדירות של נתונים שנישואות אנליטיות ויוריסטיות אחרות מתקשות להתמודד איתן, בשילוב עם זמינותם הנוברת של מעבדים חדשים וזולים, הביאו לאחרונה לפיתוח נישת הלמידה העמוקה (*Deep learning*). גישה זו מבוססת על רשתות נוירונים רב-שכבתיות וכבר הביאה להתקדמות מרשימה במגוון תחומים בתעשייה, כמפורט בטבלה 6.

בעוד שאלגוריתמים "רגילים" של למידת מכונה מחייבים שימוש בסט של נתונים בעלי מאפיינים מוגדרים שאורגנו מראש על ידי מומחים לשם משימת למידה מוגדרת, למידה עמוקה (*Deep learning*) מאפשרת למידה עם מגוון סוגי נתונים, כולל נתונים לא מובנים ולא מאורגנים (*Unstructured*) כגון קובצי אודיו, סרטונים וכדומה. לא רק שלמידה עמוקה מאפשרת לזהות תבניות ודפוסים מתוך דוגמאות לא מזהות או לא מובנות, אלא היא גם מאפשרת לסווג ולתייג את התבניות השונות ואת הסדר שלהן, ובהדרגה גם "להבין" את הנתונים האלה ולהפוך אותם לתובנות וידע שהולך ומשתפר כל הזמן.

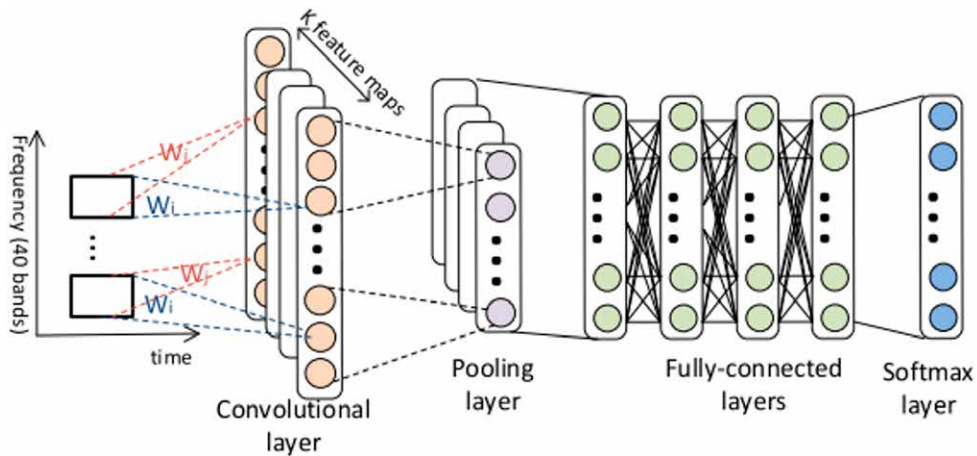
יתרה מזו, למידת מכונה מחייבת מעורבות של מומחה על מנת לאתר תכונות (*Features*) או מאפיינים מסוימים בנתונים הנולמייים שעל פיהם ניתן לאמן את המערכת הלומדת. בלמידה עמוקה, לעומת זאת, לא נדרשים בהכרח ארגון, תיוג וסידור המידע, שכן האלגוריתמים של למידה עמוקה מאפשרים לימוד באמצעות רשתות נוירונים מנתוני עתק גולמיים (כמו תמונות לזיהוי, מידע פיננסי מבוסס זמן) על מנת לאתר מאפיינים חביים, הבדלים בנתונים לאורך זמן, דפוסים, אנומליות וכדומה.

השימוש העיקרי במודלים של למידה עמוקה בתעשייה נועד לפתור בעיות עסקיות שמתבססות על כמויות גדולות של נתונים. נטפליקס וספוטיפיי, למשל, השתמשו בלמידה עמוקה במערכות המלצה על מנת ללמוד את העדפות המשתמשים (*Unger et al., 2020*). למידה עמוקה מאפשרת גם ניתוח נתונים ועיבודם לשם קבלת החלטות עסקיות בזמן אמת, כגון ניהול כספים ומסחר במניות, התראות על התקפות סייבר, וכן ראייה ממוחשבת לפענוח תמונות וקובצי וידאו ועוד. בעיות עסקיות נוספות שהן חלק בלתי נפרד מחיינו הן זיהוי קול (*Huang et al., 2015*) ועיבוד שפה טבעית (כגון תרגום משפה לשפה). טכנולוגיות אלו כבר הוטמעו בצורות שונות כמעט בכל המכשירים החכמים (*IoT*) בחיינו, וכבר מספקים מענה בתחומים מורכבים יותר כמו רפואה, תעשיית הרכבים האוטונומיים, רחפנים וכדומה.

בתחום הרפואה, חוקרים משתמשים בלמידה עמוקה על מנת לעבד נפחים עצומים של נתונים גנומיים וקליניים ממגוון רחב של אוכלוסיות במטרה לזהות קשרים לא ידועים עד כה בין גנים, תרופות וכי (*Piccialli, et al., 2021*). חלק מהיישומים בתחום זה נעזרים בלמידה עמוקה לחיזוי תאים סרטניים מתמונות רנטגן, נילוי תרופות ותמיכת בהחלטות קליניות, ניתוח וחיזוי של מצב החולה עבור מגוון רחב של מצבי חירום רפואיים, ועוד.

שימוש נרחב נוסף בלמידה עמוקה הוא הבנה של שפה אנושית. תוכנות כאלה כבר קיימות במנועי חיפוש, בחיפוש קולי ובעוזרים דיגיטליים כמו "סירי" של חברת אפל ו"קורטנה" של מיקרוסופט, או *chatGPT* שמאפשר שיח עם אלגוריתם חכם של למידה עמוקה. בעתיד יסיקו מערכות אלה גם מסקנות ויקבלו החלטות בכוחות עצמן.

איור 6: יישום של רשת נוירונים מבוססת Convolutions רב-שכבתית לעיבוד קול בחברת Microsoft



נתונים רב-שכבתיים הנקלטים במערכת לאורך זמן (כגון זו המתוארת באיור 6 ומיושמת בחברת Microsoft)². כפי שניתן לראות מאיור 6, בכל פרק זמן (Time) נקלטים במערכת אותות (קול או תמונה). אותות אלה מועברים לשכבות Convolution שמרידות את הממדיות של הנתונים באמצעות פונקציות אגרציה שונות (Pooling layer) ולאחריהן המידע מועבר לרשתות נוירונים רב-שכבתיות עם הזנה קדמית במטרה לזהות את האובייקטים בתמונה או את המילים המעובדות מקלט הקול.

מערכות ראייה ממוחשבת משמשות כיום גם לזיהוי פרצופים של חשודים בביצוע עבירות, מחבלים וחברים בארגוני טרור וכדומה. שימוש נוסף הוא זיהוי פנים לצורכי אבטחה, זיהוי החברים המצולמים ברשתות חברתיות, ומעקב אחרי פושעים וטרוריסטים במקומות ציבוריים. לעיתים קרובות מערכות אלה מופעלות יחד עם טכנולוגיות משלימות, שכן זיהוי פנים היא חלק מתחום הזיהוי הביומטרי. טכנולוגיות נוספות שמשמשות בזיהוי ביומטרי הן זיהוי טביעות אצבע, זיהוי דנ"א וזיהוי קשתית העין.

טכנולוגיית זיהוי פנים (Facial recognition), מבוססת בעיקרן על יכולת טכנולוגית מבוססת אלגוריתם למידה עמוקה, שמסוגל לזהות אדם על פי צילום דיגיטלי שלו, אם בתצלום ואם בסרט וידאו. מערכות זיהוי פנים רבות פועלות כיום באופן אוטומטי על ידי עיבוד כמויות אדירות של מידע, במהירות וביעילות,

² מקור/ <https://www.microsoft.com/en-us/research/wp-content/uploads/2015/04/Detailed-Analysis-of-Convolutional-Neural-Networks-for-Speech-Recognition-Final.pdf>

בתחום הרכב, למשל, כבר היום קיימים פיתוחים רבים מבוססי למידה עמוקה וראייה ממוחשבת, המאפשרים זיהוי עצמים במרחב וסיווגם ללא צורך בהתערבות אנושית. באמצעות חיקוי היכולת האנושית ללמוד, מערכות למידה עמוקה מצליחות גם ליצור ציורים, מוזיקה, סרטונים וטקסטים, להמליץ ללקוחות על רכישות, לייצר קריינות מלאכותית, לזהות עצמים ופנים מדויקות בתמונות וסרטונים, לאבחן מחלות, להפיק רווחים ממניות, להלביש פנים של מפורסמים על סרטי וידאו ועוד שלל יישומים רבים.

ללא ספק, אחד מהפיתוחים המרשימים של למידה עמוקה הוא התחום של ראייה ממוחשבת (Computer Vision), המבוססת על מערכות מחשב עם יכולת לזיהוי תמונות או וידאו. במערכות ראייה ממוחשבת עושים שימוש במגוון תחומים, כגון זיהוי פנים, זיהוי מבנים, מיפוי, מדידה, עיבוד תמונה ועוד. מערכות כאלה יכולות לזהות מרכיבים בתמונה ולתייג אותם (אנשים, מעברי חצייה, רמזור) על סמך מידע רב של תמונות שנלמדו על ידי המחשב תוך שימוש באלגוריתם של למידה עמוקה (Chai, et al., 2021).

מכשירים סלולריים, רובוטים ומכונות אוטונומיות נעזרים גם הם בראייה ממוחשבת כדי לראות ולהבין את סביבתם, ממש כמו בני אדם. הם מצוידים ביכולת לנתח את המרכיבים החזותיים שנקלטו במערכת בזמן אמת, ולשפר את היכולות שלהם על בסיס מה שהן למדו בפעמים קודמות, על סמך למידה ממאגרי מידע רבים, ועל פי דפוסים ותבניות שונות בחלקי התמונות. יכולות אלה מבוססות על רשתות נוירונים המותאמות לניהול

כשהם נעזרים במאגרי נתונים ענקיים המכילים צילומי פנים של אזרחים רבים (Kaur, Paramjit, et al., 2020).

סיכום

בעיות התכנסות, מרחב מצבים כמעט אינסופי, שיעורי תגובה נמוכים, מספר גדול של משתנים מסבירים, היעדר תשתית תיאורטית, ועוד סוגיות אחרות שחלקן נידונו במאמר הזה, כל אלה מקשים על מודלים המבוססים על רשתות נוירונים להתמודד עם בעיות גדולות של שיווק ישיר. במאמר זה סקרנו מקצת מהסוגיות הכרוכות ביישום מודלים של רשתות נוירונים בבעיות שיווק ישיר, אבל לא פסחנו גם על היתרונות של גישות מבוססות רשתות נוירונים לפתרון בעיות חיזוי אנליטי בכלל, ושיווק ישיר בפרט. סיימנו עם דוגמה מעשית שמשווה את הביצועים של מודל רשתות נוירונים בבעיית שיווק ישיר עם אלה של רגרסיה לוגיסטית על אותו בסיס נתונים.

מהדיון במאמר הזה עולות מספר מסקנות עיקריות לגבי היישום של מודלים מבוססי רשתות נוירונים, התקפות לא רק לבעיות טירגוט בדיור ישיר אלא גם לבעיות עסקיות אחרות:

- תהליך היישום של מודל רשתות נוירונים אינו אוטומטי, ומחייב חיפוש רחב על מנת למצוא את הטופולוגיה המיטבית של רשת הנוירונים והפרמטרים שלה.
- מודל חיזוי טוב תלוי באופן בחירת המשתנים המסבירים למודל, ככל שהמשתנים משפיעים יותר, כך מודל החיזוי מדויק יותר. אולם בהיעדר תשתית תיאורטית, מודלים מבוססי רשתות נוירונים חסרים את המנגנון לבחור משתנים מסבירים למודל, מה שמחייב שימוש בתהליך של עיבוד מקדים על מנת לבחור את המשתנים להכניס למודל.

- חוקי הסיום (*Termination rules*) עבור תהליך הלמידה במודלים מבוססי רשתות נוירונים אינם חד-משמעיים וחוקי סיום שונים יכולים להניב תוצאות שונות.
- לציונים (*Scores*) של מודל מבוסס רשתות נוירונים אין משמעות הסתברותית. כתוצאה מכך לא קיים קריטריון שמאפשר להפריד באופן חד-משמעי בין מניבים ללא מניבים.
- מודלים מבוססי רשתות נוירונים הם בבחינת "קופסה שחורה", מה שגורם לכך שמודלים אלה אינם ברי הסבר.
- ולבסוף, מודלים מבוססי רשתות נוירונים הם עתירי חישובים וצורכים משאבי מחשב רבים.

אבל בפן החיובי, מודלים מבוססי רשתות נוירונים הם ברי השוואה למודלים המבוססים על רגרסיה לוגיסטית, לפחות בנוגע לבעיות טירגוט שנידונו בהרחבה במאמר זה, שכן במבחן התוצאה הם הניבו פתרונות די קרובים לפתרונות המתקבלים באמצעות רגרסיה לוגיסטית. יתרה מזו, מניסיוננו מסתבר שרשתות נוירונים מניבות פתרונות די יציבים (*robust*) עבור מנעד רחב של קונפיגורציות רשת, מה שמקל על תהליך החיפוש של הטופולוגיה ה"אופטימלית" של הרשת. כמו כן, באמצעות מיצוע של התוצאות מכמה רשתות נוירונים ניתן למזער את הסיכונים לקבל מודל עם התאמה (*fit*) לא טובה של נתונים, הנובעת למשל מהגדרה לא טובה של הרשת.

קינחנו את המאמר בסקירה קצרה של מערכות למידה עמוקה, שמהוות את הדור הבא של רשתות הנוירונים ה"רגילות" שבהן עוסק המאמר הזה, ושל השימושים העיקריים שלהן. נציין שהנושא של למידה עמוקה חורג מהדיון במאמר הנוכחי אך נידון בהרחבה בספרות המקצועית. סקירה טובה של הנושא ניתן למצוא בספר (Kelleher 2019).

ד"ר משה אונגר mosheunger@tauex.tau.ac.il

- Ben-Akiva, M., & S.R. Lerman. 1987. *Discrete Choice Analysis*, the MIT Press, Cambridge, MA.
- Cox, D.R. & Oakes, D. (1984). *Analysis of Survival Data*, Chapman and Hall.
- Chai, J., Zeng, H., Li, A., & Ngai, E. W. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, 6, 100134.
- DeGroot, M. H. (1993). *Probability and Statistics* 3rd edition, Addison-Wesley.
- Gandhi, P. (2019). *Explainable Artificial Intelligence*, <https://www.kdnuggets.com/2019/01/explainable-ai.html>
- Gary, L., Rangaswamy, A. & De Bryun. A. (2007). *Principles of Marketing Engineering*, ISBM, Penn State University.
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wiseley, Reading, MA.
- Hebb, D. O. (1949). *The organization of behavior; a neuropsychological theory*. A Wiley Book in Clinical Psychology, 62, 78.
- Hopefield, J. (1969). *Neural networks and physical systems with emergent collective computational abilities*, *Proceedings National Academy of Science*, Vol 79(8), 2554-2558.
- Huang, J. T., Li, J., & Gong, Y. (2015). *An analysis of convolutional neural networks for speech recognition*. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4989-4993). *IEEE*.
- Kaur, P., Krishan, K., Sharma, S. K., & Kanchan, T. (2020). Facial-recognition algorithms: A literature review. *Medicine, Science and the Law*, 60(2), 131-139.
- Kelleher, J.D. (2019). *Deep Learning*. *The MIT Press Essential Knowledge Series*.
- Levin, N. & Zahavi, J. (1997). *Applying Neural Networks to Target Marketing*, *Journal of Direct Marketing*. Vol. 11(4).
- Leshno, M., Lin, V.Y., Pinkus, A. & Schoken, S. (1993). *Multilayer Feedforward Networks with a Nonpolynomial Activation Function Can Approximate Any Function*, *Neural Networks*, Vol. 6 pp. 861-867.
- Minsky, M.L. & Papert, S.A. (1969). *Perceptrons*. MIT Press, Cambridge.
- Piccialli, F., Di Somma, V., Giampaolo, F., Cuomo, S., & Fortino, G. (2021). A survey on deep learning in medicine: Why, how and when?. *Information Fusion*, 66, 111-137.
- Rosenblatt, F. (1957). *The Perceptron and Recognizing Automation*, Report 85-460-1, Cornell Aeronautical Laboratory, Buffalo, NY.
- Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986). *Learning Internal Representations by Error Propagation*, in: Rumelhart, D.E., McClelland, J.L. & the PDP Research Group, (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1: Foundations, MIT Press, Cambridge, MA, 318-362.

- Schoenborn, J. M. & Althoff, J. D. (2018). *Recent Trends in XAI: A Broad Overview on Current Approaches, Methodologies and Interactions*, Schoenborn, J. M. and Althoff, J. D. (2018). *Recent Trends in XAI*.
- Steinbach, B. (2002). *Neural Networks – A Model of Boolean Functions*, Neural Networks – A Model of Boolean Functions.
- Tang, Z., de Almeida, C., & Fishwick, P. (1991). *Times Series Forecasting Using Neural Networks vs. Bok Jenkins Methodology, Simulation*, 57.
- Tam, K.Y., & Kiang, M.Y. (1992). *Managerial Applications of Neural Networks: The Case of Bank Failure Predictions*, Management Science, 38.
- Unger, M., Tuzhilin, A., & Livne, A. (2020). *Context-aware recommendations based on deep learning frameworks. ACM Transactions on Management Information Systems (TMIS)*, 11(2), 1-15.
- Van Laarhoven, P.J.J. & Aarts, E.H.L. (1987). *Simulated Annealing Theory and Application*, Kluwer, Boston.
- Vaughan, J., Sudjianto, A., Brahim, E, Chen, J. & Nair, V.N. (2018). *Explainable Neural Networks based on Additive Index Models*, arXiv:1806.01933v1.
- זהבי, י. (2017), חיזוי אנליטי (*Predictive Analytics*) – הלכה למעשה, *חידושים בניהול*, אוניברסיטת תל אביב, 1, 69-55.
- זהבי, י. (2022), סקירה של כשלים בבנייה ויישום של מודלים בחיזוי אנליטי, *חידושים בניהול*, אוניברסיטת תל אביב, 9, 114-98.