

### סוגי מודלים של למידת מכונה

קיימות 4 קטגוריות של מודלים של למידת מכונה: (1) למידה בהשגחה; (2) למידה ללא השגחה; (3) למידה בהשגחה למחצה; ו- (4) למידה בחיזוקים.

למידה בהשגחה (Supervised Learning) עוסקת בשימוש בנתונים לצורך ניבוי (קרי, ביצוע תחזיות). אחת הדוגמאות הפשוטות ללמידה בהשגחה היא שימוש במודל רגרסיה פשוטה על מנת לנבא משכורות. דוגמא אחרת היא שימוש במודל מורכב לניבוי מחירי דירות. נאמר מראש מקובל להבחין בין מודלים של למידה בהשגחה המשמשים לניבוי משתנה מסוים שיכול לקבל ערכים רציפים (כמו למשל משכורות של פרט בודד או מחירה של דירה) לבין מודלים המשמשים לסיווג (כמו למשל קביעת דירוג אשראי). מודלים של סיווג נפוצים מאוד בלמידת מכונה. לדוגמא, אחד היישומים של למידת מכונה הוא סיווג לווים פוטנציאליים כסיכון אשראי שאנו מוכנים לקבל או כסיכון אשראי שאנו לא מוכנים לקבל.

למידה ללא השגחה (Unsupervised Learning) עוסקת בזיהוי דפוסים בתוך נתונים. המטרה העיקרית איננה לנבא משתנה פרטני כלשהו, כי אם להבין טוב יותר את הסביבה המיוצגת על ידי הנתונים. ניקח למשל חברה שמשווקת מנעד של מוצרים ללקוחות. למעשה נתונים אודות רכישותיהם של הלקוחות יכולים לשמש לקביעת המאפיינים של לקוחות אשר רכשו מוצרים שונים, אשר בתורם יכולים להשפיע על הדרך שבה אותם מוצרים יפורסמו מכאן ואילך. כך למשל, ניתוח אשכולות (Clustering, קיבוץ) הינו כלי עיקרי המשמש בלמידה ללא השגחה.

הנתונים הנדרשים ללמידה בהשגחה כוללים מאפיינים (Features, משתנים מסבירים) ותוויות (Labels). התוויות הן הערכים של היעד (Target, המשתנה המוסבר) שאותו רוצים לנבא. המאפיינים הינם המשתנים שבאמצעותם מבוצעות התחזיות. לדוגמא, כאשר רוצים לנבא את מחירה של דירה, אז המאפיינים יכולים להיות מ"ר, מספר חדרים, מספר מרפסות, האם יש או אין מעלית, וכך הלאה. התווית תהיה מחיר הדירה. הנתונים הנדרשים ללמידה ללא השגחה

כוללים מאפיינים אך אינם כוללים תוויות מאחר ומדלים של למידה ללא השגחה משמשים לזיהוי דפוסים ולא לניבוי. כך למשל, ניתן להשתמש במודל ללא ההשגחה על מנת להבין את הדירות הקיימות בשכונה מסוימת מבלי לנסות לנבא את מחיריהם. אנו עשויים למצוא שקיים אשכול אחד של דירות של 40-70 מ"ר, 3 חדרים ומרפסת ואשכול אחר של דירות של 120-210 מ"ר, 6 חדרים ושתי מרפסות.

למידה בהשגחה למחצה (Semi-Supervised Learning) נופלת איך לא בין למידה בהשגחה ללמידה ללא השגחה. למידה בהשגחה למחצה מתעוררת כאשר אנו מנסים לנבא משהו מסוים ולחלק מהנתונים שבידינו יש תוויות (קרי, ערכים עבור היעד שאותו אנו רוצים לנבא) בעוד שליתר הנתונים (בדרך כלל לרב המוחץ של הנתונים) אין תוויות. ניתן לחשוב שנתונים ללא תוויות הם חסרי תועלת, אך אין זה בהכרח נכון. ניתן להשתמש בנתונים ללא תוויות בשילוב עם נתונים עם תוויות על מנת לייצר אשכולות המסייעים לתחזיות. לדוגמא, נניח שאנו מעוניינים לנבא האם לקוח מסוים ירכוש מוצר מסוים מתוך מאפיינים כמו למשל גיל, רמת הכנסה וכך הלאה. עוד נניח שיש לנו כמות קטנה של נתונים עם תוויות (קרי, נתונים המצביעים על המאפיינים של לקוחות כמו גם האם הם קנו או לא קנו את המוצר) וכמות הרבה יותר גדולה של נתונים ללא תוויות (קרי, נתונים המצביעים על המאפיינים של לקוחות פוטנציאליים, אך נתונים אלו לא מספרים לנו אותם לקוחות קנו או לא קנו את המוצר). למעשה, ניתן ליישם למידה ללא השגחה ולהשתמש במאפיינים הללו במטרה לבצע ניתוח אשכולות ללקוחות פוטנציאליים.

שווה בדמיון את הסיטואציה הבאה, כאשר קיימים שני אשכולות, א' ו- ב' בסט הנתונים המלא וכי הלקוחות שידוע מתוך הנתונים עם התוויות שהם רכשו את המוצר מתאימים כולם לנקודות באשכול א', בעוד שהלקוחות שידוע מתוך הנתונים עם התוויות שהם לא רכשו את



המוצר מתאימים כולם לנקודות באשכול ב'. כפועל יוצא מכך ניתן לסווג את כל הפרטים באשכול א' כקונים ואת כל הפרטים באשכול ב' כחנטרישים.

אנו בני אדם עושים שימוש רב בלמידה בהשגחה למחצה. שווה בדמיוןך שאתה לא מכיר את השמות "חתול" ו- "כלב", אבל שמת לב לשני אשכולות נפרדים של חיות מחמד בשכונה שלך. כעת בא מישהו ובאדיבותו הרבה מצביע על שתי חיות ספציפיות ומגלה לך שאחת מהן היא חתול והשניה היא כלב. מאותו רגע ואילך אין לך בעיה להשתמש בלמידה בהשגחה למחצה כדי ליישם את התוויות הללו על כל החיות האחרות שראית בעבר ושתראה בעתיד. ולכן, ם משתמשים בלמידה בהשגחה למחצה בצורה שכזו, אז אין זה פלא שמכונות יכולות לעשות אותו דבר. מרבית האלגוריתמים של למידת מכונה מבוססים על למידה של הדרכים השונות שבהן המוח שלנו מעבד נתונים.

הקטגוריה האחרונה של למידת מכונה, למידה בחיזוקים (Reinforcement Learning), עוסקת במצבים שבהם אלגוריתם של למידת מכונה מקיים אינטראקציה עם הסביבה ומקבל סדרת החלטות. מכונות ללא נהג משתמשות באלגוריתמים של למידה בחיזוקים. אלגוריתמים של למידה בחיזוקים עומדים מאחורי תוכנות שפותחו על מנת לשחק משחקים כמו שח וגו (משחק סיני עתיק). חלק ממכונות המסחר האוטומטיות (אלגוריתמי) בשוק ההון עושות שימוש באלגוריתמים של למידה בחיזוקים.

## פרטים אודות כותב המאמר: האקטואר רועי פולניצר, FRM

רועי בעל תואר שני במימון (התמחות בניהול סיכונים ואקטואריה) ותואר ראשון בכלכלה (התמחות במימון), שניהם מאוניברסיטת בן-גוריון בנגב, בעל דיפלומה בניהול סיכונים פיננסיים (FRM®) מאוניברסיטת אריאל בשומרון ולמד בתוכנית ללימודי תעודה באקטואריה באוניברסיטת חיפה. כמו כן, רועי אקטואר מלא



(Fellow) בלשכת מעריכי השווי והאקטוארים הפיננסיים בישראל (F.I.L.A.V.F.A.), מוסמך כמעריך שווי מימון תאגידי (CFV) מטעם לשכת מעריכי השווי והאקטוארים הפיננסיים בישראל (IAVFA), מוסמך כמנהל סיכונים פיננסיים (FRM) מטעם האיגוד העולמי למומחי סיכונים (GARP) ומוסמך כמומחה לניהול סיכונים (CRM) מטעם האיגוד הישראלי למנהלי סיכונים (IARM).

לרועי ניסיון של מעל ל-15 שנה בביצוע ניתוחים כמותיים במכשירים פיננסיים, בהערכת שווי תאגידים ונכסים בלתי מוחשיים, באמידה וכימות סיכונים כמו תמותה, אריכות ימים, תחלואה, ביטולים והחלמה מנכות, ובמידול ומדידת סיכוני שוק, אשראי, תפעוליים, מודל, מזילות והשקעות לצורכי יישום הוראות רגולטוריות ותקינה חשבונאית, פיתוח, יישום ותיקוף מודלים בתחומים של הערכות שווי, ניהול סיכונים, אקטואריה והנדסה פיננסית, קביעת תעריפי ביטוח חיים, הערכת פרמיות סיכון והערכת עתודות ביטוח, קביעת עלות תנאי פנסיות (צוברות ותקציביות) והכנת מאזנים אקטואריים לקרנות פנסיה, ניתוח וחיזוי מצבים פיננסיים מורכבים וכן העברת סמינרי הדרכה והשתלמויות בתחומי התמחותו: מימון, אקטואריה, הערכות שווי, בנקאות, ניהול סיכונים, אופציות והנדסה פיננסית.



ניסיונו של רועי בתחום ה-Data Analysis, כולל: עבודה עם מאגרי מידע גדולים Big Data תוך שימוש ב- Statistical Learning (כגון: סטטיסטיקה תיאורית, הסתברות, הסקה סטטיסטית, סטטיסטיקה א-פרמטרית, חלוקת נתונים, נרמול נתונים, Fitting ו- Bayes Theorem) ובאלגוריתמים מסוג Unsupervised Learning (כגון: Hierarchical k-means Clustering, Clustering, Density-based Clustering, Distribution-based Clustering ו- Principle Components Analysis) למציאת דפוסים וזיהוי מגמות ואנומליות בעולמות ניהול הסיכונים, ההשקעות, האקטואריה, הביטוח והפנסיה, פיתוח תשתית לצורך ניתוח נתונים, שילוב והטמעת כלים לצורך גישה ושליפה עצמאית של נתונים ממאגרי מידע, פיתוח דוחות, ממשקים ומסכים באמצעות כלי ויזואליזציה.

ניסיונו של רועי בתחום ה-Data Science, כולל: עבודה עם מסדי נתונים גדולים Big Data תוך שימוש באלגוריתמים מסוג Supervised Learning (כגון: Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, Logistic Regression, Maximum Likelihood Estimation, k-Nearest Neighbors, Decision Tree, Random Forest, Ensemble, Bagging, Boosting, Naïve Bayes Classifier, Linear Separation, Support Vector Machine, Non-Linear Separation, SVM Regression, Artificial Neural Network, Convolutional Neural Network ו- Recurrent Neural Network) לניבוי וסיווג בעולמות ניהול הסיכונים, ההשקעות, האקטואריה, הביטוח והפנסיה ובמדלים מסוג Reinforcement Learning (כגון: Q-learning, Monte Carlo Simulation, Temporal Difference Learning ו- n-Step Bootstrapping) לקבלת החלטות מרובות שלבים בעולמות ניהול הסיכונים, ההשקעות, האקטואריה, הביטוח והפנסיה, זיהוי אתגרים עסקיים שבהם DATA יכול להוות גורם מכריע בשיפור קבלת החלטות, איתור ואיסוף מקורות מידע, הגדרה ואיפיון של שימושי המידע, בניית מסד המידע, אפיון והגדרת הצגת המידע



ותוצריו, פיתוח כלים, מודלים, תהליכים ומערכות בתחום האנליזה, תוך שימוש בכלי אנליזה מתקדמים (EXCEL, VBA ו-R).