

## למידה בהשגחה (Supervised Learning): מכונת וקטורים תומכים (Support Vector Machine)

מכונת וקטורים תומכים או בקיצור SVM מחפשת לסווג תצפיות על ידי גזירת נתיב (Pathway) שמחלקת את הקבוצות. במצב הפשוט ביותר, המשוואה עבור כל אחד מהצדדים של הנתיב הינה פונקציה לינארית של המאפיינים וכל התצפיות שסווגו נכונה. מצב זה מכונה הפרדה קשיחה (Hard Separation). עם זאת, הפרדה מושלמת על פי רוב אינה אפשרית וקיימת תחלופה בין רוחב הנתיב לבין ההפרות (Violations). מידת ההפרה של תצפית אשר לא "שוכבת" על הצד הנכון של הנתיב נמדדת באמצעות המרחק הקצר ביותר בין התצפית לבין המקום שבו היא צריכה לשכב על מנת להיות מסווגת נכונה.

באמצעות עבודה עם פונקציות של ערכי המאפיינים במקום עם ערכי המאפיינים עצמם, ניתן להפוך את הנתיב שמפריד את התצפיות לשתי קבוצות ללא ליניארי (Non-linear). כמוכך שניתן ליצור מאפיינים חדשים באמצעות העלאת ערכי המאפיינים בריבוע, בשלישית, בחזקת ארבע, וכו'. לחילופין ניתן ליצור ציוני דרך (Landmarks) במרחב המאפיינים (Feature Space) באמצעות מאפיינים חדשים שהינם פונקציות של מרחקי התצפיות מציוני הדרך.

גרסיית SVM משתמשת ברעיונות העומדים בבסיס סיווג SVM לצורך ניבוי ערכו של משתנה רציף (Continuous Variable). במסגרת גרסיית SVM נוצר נתיב דרך התצפיות המשמש לניבוי היעד (Target, המשתנה המוסבר). אם ערכו של היעד עבור תצפית מסוימת נופל בתוך הנתיב, או אז מונח שאין שגיאת ניבוי או טעות בחיזוי. מאידך גיסא, אם ערכו של היעד עבור תצפית מסוימת נופל מחוץ לנתיב, הרי ששגיאת הניבוי היא ההפרש שבין ערכו של היעד לבין מה שהיה צריך להיות ערכו של היעד אילו התצפית הייתה נופלת בתוך הנתיב. רוחב הנתיב (הנמדד בכיוון של ערכו של היעד) מוגדר למעשה על ידי המשתמש. נעיר כי קיימת תחלופה בין שגיאת הניבוי הממוצעת לבין כמות הרגולריזציה (Regularization), פונקציה המשמשת

לפתרון בעיית ה- Over-fitting שקיימת באלגוריתמים של למידת מכונה, כלומר, מצב שבו האלגוריתם לומד טוב מדי את נתוני סט האימון, אך לא נותן תחזיות נכונות לנתוני סט הבדיקה).

### **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 סיכונים שוק, אשראי, תפעוליים, מודל, נזילות 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 ו-SVM Regression, Artificial Neural Network, Convolutional Neural Network, 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).