

למידה בהשגחה (Supervised Learning): רגרסיה לוגיסטית

רגרסיה לוגיסטית (Regression Logistic), ממש כמו רגסיה לינארית, משמשת במחקר האמפירי מזה שנים רבות. כעת הרגרסיה הלוגיסטית הופכת לכלי סיווג חשוב עבור מדעני נתונים. באופן טבעי קיימות שתי קבוצות. האחת מכונה "חיובית"; האחרת מכונה "שלילית". פונקציית הסיגמואיד (Function Sigmoid) בצורת S משמשת להגדרת ההסתברות של תצפית מסוימת 'ליפול' בקבוצה החיובית.

נוהל של חיפוש איטרטיבי משמש למציאת הפונקציה הלינארית של ערכי המאפיינים שכאשר מכניסים אותה לתוך פונקציית הסיגמואיד עושה את העבודה הטובה ביותר בהקצאת הסתברות גבוהה לתוצאות חיוביות והסתברות נמוכה לתוצאות שליליות. מקובל לסכם את תוצאות השימוש ברגרסיה לוגיסטית על נתוני סט הבדיקה (Test Set) באמצעות שימוש במטריצת הטעות (Matrix Confusion).

משעה שסיימנו לבנות את הרגרסיה הלוגיסטית, עלינו להחליט באיזה אופן ישמשו התוצאות. כך למשל, אם הרגסיה הלוגיסטית משמשת להחלטת מתן אשראי, הרי שעל מקבל ההחלטה להגדיר ערך Z מסוים. כאשר ההסתברות לקבלת תוצאה חיובית מההלוואה נאמדת כגבוהה יותר מ-Z, או אז הבקשה להלוואה מאושרת. כאשר ההסתברות לקבלת תוצאה חיובית מההלוואה נמוכה יותר מ-Z, או אז הבקשה להלוואה נדחית.

קיים יחס תחלופה בין הצלחה לזהות הלוואות טובות לבין הצלחה לזהות הלוואות שתגענה למצב של חדלות פירעון (Default). באופן טבעי שיפור הצלחה לזהות הלוואות שתגענה למצב של חדלות פירעון מביא להרעה בהצלחה לזהות הלוואות טובות, ולהיפך. ניתן לסכם את יחס תחלופה הזה באמצעות עקומת ה-ROC (עקומה אופיינית למסווג, Receiver Operating Characteristic Curve) המקשרת בין שיעור החיובים האמיתיים (TPR), כלומר, רמת הביטחון או

רמת הסמך ופירושו האחוז מהזמן שבו תוצאה חיובית מסווגת כחיובית) לבין שיעור החיוביים הכוזבים (FPR, רמת המובהקות ופירושו האחוז מהזמן שבו תוצאה חיובית מסווגת כשלילית).

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 וחזוי 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 Clustering, k-means 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 ובמודלים 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).