

מורד השיפוע הסטוכסטי

מורד השיפוע הסטוכסטי (SGD- Stochastic Gradient Descent) הוא אלגוריתם של למידת מכונה המשמש למזעור פונקציית העלות (cost function, המכונה גם פונקציית ההפסד) על ידי תהליך איטרטיבי של עדכון המשקולות על סמך השיפועים. במאמר זה, אספק מבוא לאלגוריתם מורד השיפוע הסטוכסטי בלמידת מכונה ואציג את יישומו באמצעות שפת פייתון (Python).

מורד השיפוע הסטוכסטי

אלגוריתם ה-SGD "מכבב" בכמה פונקציות הפסד. במילים פשוטות, האלגוריתם משמש למזעור פונקציית עלות על ידי תהליך איטרטיבי של עדכון המשקולות על סמך השיפועים. במקום להסתכל על מערך הנתונים המלא, עדכון המשקולות מופעל על בציים (batches, אצוות) שחולצו ממנו באופן מקרי, ולכן האלגוריתם מכונה גם מורד השיפוע מסוג אצווה-קטנה (mini-batch gradient descent).

להלן תיאור התהליך של אלגוריתם מורד השיפוע הסטוכסטי:

1. האלגוריתם מתחיל בנקודה אקראית על ידי אתחול המשקולות באמצעות ערכים אקראיים.
2. לאחר מכן האלגוריתם מחשב את השיפועים (gradients) באותה נקודה אקראית זו
3. לאחר מכן האלגוריתם נע בכיוון ההפוך לשיפוע (עם השיפוע הוא שלילי אז האלגוריתם יזוז ימינה לכיוון החיובי ואם השיפוע הוא חיובי אז האלגוריתם יזוז שמאלה לכיוון השלילי)
4. התהליך ממשיך באופן רפטיטיבי עד שהאלגוריתם מוצא את נקודת ההפסד המינימלית

מורד השיפוע הסטוכסטי באמצעות שפת פייתון

לאחר שהסברתי מהו אלגוריתם ה-SGD בלמידת מכונה, כעת אציג את היישום שלו באמצעות שפת פייתון:

```
from sklearn.datasets import make_classification
from sklearn.linear_model import SGDClassifier
```

```
samples = 500
x, y = make_classification(n_samples=samples,
                          n_features=2,
                          n_informative=2,
                          n_redundant=0,
                          n_clusters_per_class=1)
```

```
SGD_classifier = SGDClassifier(loss="perceptron",
                              learning_rate="optimal",
                              n_iter_no_change=10)
```

```
from sklearn.model_selection import cross_val_score
print(cross_val_score(SGD_classifier, x, y, scoring="accuracy", cv=10).mean())
```

0.866

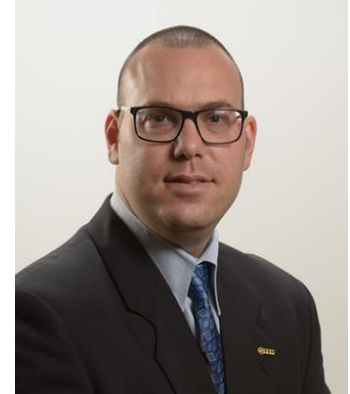
סיכום

במאמר לעיל הצגתי כיצד ניתן ליישם את אלגוריתם ה-SGD הסיווג בלמידת מכונה באמצעות שפת התכנות של פייתון. אלגוריתם ה-SGD "מכבב" כאמור במספר פונקציות הפסד. במילים פשוטות, האלגוריתם משמש למזעור פונקציית עלות על ידי תהליך איטרטיבי של עדכון המשקולות על סמך השיפועים.

פרטי השכלתו של מדען הנתונים האחראי מטעם פרדיקציות יועצים: מר רועי פולניצר

■ מדען נתונים מוסמך בישראל, בעל תואר M.B.A. במנהל עסקים (עם התמחות בבניית מודלים מתמטיים וסטטיסטיים) ותואר B.A. בכלכלה (עם התמחות בכלים ושיטות לאנליזה ותחקור מידע), שניהם בהצטיינות ומאוניברסיטת בן-גוריון בנגב.

■ מחזיק בהסמכות הבאות מטעם האיגוד הישראלי למדעני נתונים מקצועיים (Professional Data Scientists' Israel Association): "מפתח פייתון מוסמך" (Certified Python Developer), "דאטה אנליסט בפייתון" (Python Data Analyst), "מומחה ללמידת מכונה" (Machine Learning Specialist), "מוסמך בלמידה עמוקה" (Accredited in Deep Learning) ו- "דאטה סיינטיסט מקצועי" (Professional Data Scientist).



■ מחזיק בהסמכות מקצועיות שונות, כגון: הסמכה בינלאומית "מנהל סיכונים פיננסיים" (Financial Risk Manager), המעידה על כך שהמחזיק בה בקיא בפיתוח, יישום ותיקוף מודלים סטטיסטיים ואלגוריתמים מתמטיים כגון K-Means, SVM ו- KNN למדידה וניהול סיכונים (מטעם האיגוד העולמי למומחי סיכונים (Global Association of Risk Professionals), הסמכת "אקטואר מלא" (Fellow Actuary), המעידה על כך שהמחזיק בה בקיא בפיתוח, יישום ותיקוף מודלים סטטיסטיים ואלגוריתמים מתמטיים כגון RF, GLM ו- NN לקביעת פרמיות בביטוח כללי) מטעם לשכת מעריכי השווי והאקטוארים הפיננסיים בישראל (Israel Association of Valuators and Financial Actuaries) והסמכת "מנהל סיכונים מוסמך" (Certified Risk Manager), המעידה על כך שהמחזיק בה בקיא בפיתוח, יישום ותיקוף מודלים סטטיסטיים ואלגוריתמים מתמטיים כגון DT, NB ו- PCA לניהול סיכונים תפעוליים) מטעם האיגוד הישראלי למנהלי סיכונים (Israeli Association of Risk Managers).

■ חבר באיגוד הבינלאומי למנהלי סיכונים מקצועיים (Professional Risk Managers' International Association), מה שמעיד עליו כבקיא במחקר, פיתוח, כתיבה ומימוש אלגוריתמים של בינה מלאכותית על כמויות גדולות של מידע).

■ למד אקטואריה (יישום טכניקות סטטיסטיות וכריית נתונים לבעיות ניתוח סדרות עתיות, צמצום מימדים, אופטימיזציה וסימולציה) במסגרת תוכנית ההכשרה היוקרתית בת ה- 250 שעות של אוניברסיטת חיפה, למד ניהול סיכונים פיננסיים (בניית מודלים סטטיסטיים מנבאים והסתברותיים לבעיות חיזוי, סיווג, ניתוח אשכולות וזיהוי אנומליות) במסגרת תוכנית ההכשרה היוקרתית בת ה- 250 שעות של אוניברסיטת אריאל וכן למד למידת מכונה ולמידה עמוקה (בניית מערכות המלצה ואימון רשתות נוירונים לעיבוד תמונה ו- NLP) במסגרת תוכנית ההכשרה היוקרתית בת ה- 500 שעות של מכללת John Bryce (התוכנית נקראת Data Science, Machine Learning & Deep Learning with Python).

■ בוגר השתלמויות מקצועיות שונות במכללת John Bryce, כגון: השתלמות מקצועית בנושא Introduction to Machine Learning, AI & Data Visualization for Managers and Architects, השתלמות מקצועית בנושא Practical Machine Learning, AI & Deep Learning with Python for Algorithm Developers & Data Scientists, השתלמות מקצועית בנושא Azure Data Fundamentals: Relational Data, Non-Relational Data and Modern Data Warehouse, השתלמות מקצועית בנושא Analytics in Azure, השתלמות מקצועית בנושא Azure AI Fundamentals: Azure Tools for ML, Automated ML & Visual Tools for ML and Deep Learning.

■ בוגר השתלמויות מקצועיות שונות של PDSIA, כגון: השתלמות מקצועית בנושא Neural Networks and Deep Learning, השתלמות מקצועית בנושא Big Data and Cloud Services והשתלמות מקצועית בנושא Natural Language Processing and Text Mining.