

סיכום שיעור 72

למידה עמוקה והקדמה לפרספטרון

Deep Learning and Perceptron Introduction

מרצה: ד"ר יורם סגל

Dr. Yoram Segal

מילות מפתח

בינה מלאכותית (AI) – למידת מכונה (Machine Learning) – למידה עמוקה (Deep Learning) – פרספטרון (Perceptron) – רשתות נוירונים (Neural Networks) – פרופגציה לאחור (Back Propagation) – רשתות קונבולוציה (CNN) – רשתות נשנות (RNN) – טרנספורמרים (Transformers) – מנגנון תשומת לב (Attention) – GPU – Hyperparameters

1 מבוא: היררכיית הבינה המלאכותית

בינה מלאכותית (AI - Artificial Intelligence) היא התחום הרחב ביותר, הכולל את כל הטכנולוגיות המאפשרות למחשבים לבצע משימות הדורשות אינטליגנציה. בתוך תחום זה נמצאת למידת מכונה (Machine Learning), ובתוכה נמצאת למידה עמוקה (Deep Learning).

1.1 הקשר בין התחומים

- **בינה מלאכותית (AI):** התחום הכללי - כולל רובוטיקה, ראייה ממוחשבת, עיבוד שפה טבעית, חיפוש, תכנון ועוד
- **למידת מכונה (ML):** תת-תחום של AI - אלגוריתמים הלומדים מנתונים, כולל סטטיסטיקה, רגרסיה, סיווג ועוד
- **למידה עמוקה (DL):** תת-תחום של ML - מבוססת על רשתות נוירונים עם שכבות רבות

2 ארכיטקטורת הלמידה

2.1 משתנים בלמידה

בכל בעיית למידה יש שלושה סוגי משתנים:

1. **משתנה הכניסה (X):** הנתונים שאנו מזינים למודל
2. **משתנה היציאה (Y):** התוצאה הרצויה (ה-Ground Truth)
3. **משתנה הפרמטרים (W):** המשקולות שהמודל לומד

2.2 היפר-פרמטרים (Hyperparameters)

היפר-פרמטרים הם פרמטרים שאינם נלמדים אלא נקבעים מראש על ידי המתכנת. הם קובעים את המבנה הכללי של המודל, כגון:

- מספר השכבות ברשת
- מספר הנוירונים בכל שכבה
- קצב הלמידה (Learning Rate)
- גודל האצווה (Batch Size)

3 היסטוריה של הפרספטרון והלמידה העמוקה

3.1 שנת 1958 – לידת הפרספטרון

פרנק רוזנבלט (Frank Rosenblatt) המציא את הפרספטרון, שהיה הצעד הראשון לקראת הרשתות הנוירונליות. הפרספטרון היה מודל פשוט של נוירון מלאכותי שיכול לבצע סיווג בינארי.

3.2 שנת 1969 – המשבר הראשון

מינסקי ופאפרט (Minsky and Papert) הראו שהפרספטרון לא יכול לפתור בעיות שאינן ניתנות להפרדה לינארית (כמו XOR). זה גרם לירידה חדה במימון ובמחקר בתחום.

3.3 שנת 1989 – התחייה הראשונה

יאן לקון (Yann LeCun) הציג את רשתות הקונבולוציה (CNN) לזיהוי ספרות בכתב יד. זו הייתה פריצת דרך משמעותית.

3.4 שנת 2006 – הפריצה

ג'פרי הינטון (Geoffrey Hinton) הציג את רשת האמונה העמוקה (Deep Belief Network) והוכיח שניתן לאמן רשתות עמוקות בהצלחה באמצעות טכניקות חדשות.

3.5 שנת 2012 ואילך – עידן הלמידה העמוקה

מאז 2012, עם הזמינות של GPU חזקים וכמויות גדולות של נתונים, הלמידה העמוקה הפכה לכלי המרכזי בבינה מלאכותית.

4 אלגוריתם הפרופגציה לאחור (Back Propagation)

4.1 עקרון הפעולה

הפרופגציה לאחור היא האלגוריתם המאפשר לאמן רשתות נוירונים. העיקרון:

1. מעבר קדימה (Forward Pass): חישוב הפלט בהתבסס על הכניסות והמשקולות
2. חישוב השגיאה: ההפרש בין הפלט בפועל לפלט הרצוי
3. מעבר לאחור (Backward Pass): חישוב הגרדיאנטים לכל משקולת
4. עדכון המשקולות: שינוי המשקולות בכיוון שמקטין את השגיאה

4.2 גרדיאנט דיסנט (Gradient Descent)

שיטת ירידת הגרדיאנט משמשת לעדכון המשקולות:

$$(1) \quad W_{new} = W_{old} - \eta \cdot \frac{\partial E}{\partial W}$$

כאשר η הוא קצב הלמידה ו- E היא פונקציית השגיאה.

5 סוגי רשתות נוירונים

5.1 רשתות מחוברות לחלוטין (Fully Connected / MLP)

- כל נוירון בשכבה מחובר לכל הנוירונים בשכבה הבאה
- מתאימות לנתונים טבלאיים
- מספר רב של פרמטרים

5.2 רשתות קונבולוציה (CNN - Convolutional Neural Networks)

- מותאמות לעיבוד תמונות
- משתמשות בפילטרים (גרעינים) המזהים דפוסים מקומיים
- מספר פרמטרים קטן יחסית
- יכולת לזהות אובייקטים ללא תלות במיקומם בתמונה

5.3 רשתות נשנות (RNN - Recurrent Neural Networks)

- מותאמות לנתונים סדרתיים (טקסט, זמן)
- יש להן "זיכרון" המאפשר לזכור מידע מצעדים קודמים
- LSTM הוא וריאנט משופר עם יכולת לזכור מידע לטווח ארוך

5.4 טרנספורמרים (Transformers)

- ארכיטקטורה מודרנית המבוססת על מנגנון תשומת לב (Attention)
- מאפשרים עיבוד מקבילי יעיל
- הבסיס למודלים כמו GPT ו-BERT
- משמשים ב-NLP, ראייה ממוחשבת ועוד

6 מנגנון תשומת הלב (Attention)

מנגנון ה-Attention מאפשר למודל "לשים לב" לחלקים רלוונטיים בקלט. במקום לעבד את כל הקלט באופן שווה, המודל לומד לאילו חלקים לתת יותר משקל.

יתרונות:

- יכולת להתמקד במידע הרלוונטי
- עיבוד מקבילי (בניגוד ל-RNN)
- ביצועים מעולים במשימות NLP

7 חשיבות ה-GPU בלמידה עמוקה

GPU (יחידת עיבוד גרפי) חיוני ללמידה עמוקה מכיוון ש:

- מאפשר חישובים מקביליים מאסיביים
 - מאיץ את תהליך האימון פי עשרות עד מאות
 - ניתן להשתמש בשירותי ענן כמו Google Colab
- דוגמה:** אימון שלוקח שבועיים על CPU יכול לקחת יום אחד על GPU.

8 סיכום

הלמידה העמוקה היא כלי רב-עוצמה המאפשר פתרון בעיות מורכבות שלא ניתן היה לפתור בעבר. מהפרספטרון הפשוט של 1958, דרך רשתות הקונבולוציה, הרשתות הנשנות ועד הטרנספורמרים של היום – התחום התפתח באופן דרמטי. הבנת העקרונות הבסיסיים היא המפתח להבנת הטכנולוגיות המתקדמות ביותר כיום.