⚡️🚀המאמר היומי של מייק 17.07.24: ⚡️🚀

Learning Rate Curriculum

רוב המאמרים שסקרתי לאחרונה היו בנושא מודלי שפה והחלטתי לגוון טיפה ולסקור מאמרים בנושאים אחרים. מאמר שנסקור היום מדבר על שיטת אימון הנקראת למידת curriculum שבא אנו מאמנים את המודל כמו שאנו מלמדים חומר לתלמידים - מהקל לקשה. יש כמה וריאציות של למידת curriculum: באחת מהם אנו מתחילים לאמן מודל עם דוגמאות קלות ובהדרגה מעלים את קושי הדוגמאות. הוריאציה השניה אנו מתחילים ממשימה קלה יותר ומעלים את מורכבותה בהדרגה. בשלישית מאמנים מודל יחסית פשוט ומעלים את ״מורכבות״ של המודל.

המאמר מציע גישת curriculum אבל לקצב למידה. המחברים מציינים שלמשל ברשתות קונבולוציה עדיף בהתחלה להתמקד יותר בלמידה של השכבות הראשונות כי למעשה אם אלו לא נלמדו טוב ועדיין קרובים למצב האיתחול שלהם אז הם יוצרים דאטה ״רועש, מדי שזורם גם לשכבות הבאות שמתקשות להתמודד איתו (המאמר מציין כמה עבודות שחקרו את הנושא והגיעו למסקנות האלו). תופעה דומה מתרחשת גם כאשר אנו עושים פיין טיון למודל למשימה מסוימת כאשר המודל לפני זה אומן למשימה אחרת.

כדי להתמודד עם סוגיה זו המחברים מציעים להתחיל מקצב למידה גבוה עבור השכבות הראשונות (שיורד ככל שמתקדמים לשכבות עמוקות יותר). במהלך האיטרציות לעלות את קצב למידה בכל השכבות כך (קצב עלייה לא שווה בין השכבות) כך שעם הזמן (=איטרציות) קצבי הלמידה של כל השכבות משתוות. נציין שהמחברים מציעים שמספר האיטרציות הנדרש להשוואת קצב הלמידה עבור כל השכבות צריך להיות משמעותית קטן יותר מכמות האיטרציות הכולל הנדרש לאימון המודל. כלומר כל השיטה הזו מופעלת בשלב ה״חימום״ של הרשת.

https://arxiv.org/abs/2205.09180