המאמר היומי של מייק: 26.07.25

Building Bridges between Regression, Clustering, and Classification

מזמן לא סקרתי מאמר שלא מופיעה בו גם מילה LLM וגם diffusion models - תתפלאו אבל יש עדיין כאלו ואני חייב להודות שזה היווה אחת הסיבות לבחירתו. המאמר דן בבעיה די מעניינת היא המרה של בעיות רגרסיה לבעיות סיווג (בתחום למידה עמוקה).

מרבית המודלים העמוקים שלנו היום, כמו llms, מודלים ויזואליים ומולטימודליים הם מודלי סיווג במהותם כלומר הפלט שלהם חי במרחב דיסקרטי כלשהו למשל טוענים טקסטואליים או פיקסלים. אז זה נשמע די טבעי לקחת בעיה שהפלט שלה רציף (חד או רב מימדי), להמיר אותה לבעיית סיווג ולבנות (לאמן) מודל סיווג במקום מודל רגרסיה. זה נעשה בד״כ על ידי חלוקה(binning) של מרחב הפלט לכמה תת-מרחבים זרים ואז כל פלט ממופה למספר תת-המרחב שהוא שייך אליו. ככה בעיית רגרסיה הופכת להיות בעיית סיווג. לאחר אימון המודל ניתן להמיר את הערך הדיסקרטי בחזרה למרחב הרציף תוך שימוש חיזוי המודל (לרוב סופטמקס).

המאמר שנסקור היום מציע גישה כללית לפיתוח מודלי סיווג לבעיות רציפות. המחברים מציעים כמה מודלים שמאומנים בצוותא לפתרון בעיה זו. המודל הראשון, האנקודר, לוקח את הקלט מעביר אותה למרחב הלטנטי ובנוסף מאמנים שכבה שחוזה את התפלגות הקטגוריות עבור הקלט (אחרי ההמרה).

המודל השני לוקח את הפלט ומעביר אותו למרחב החדש של הקטגוריות. הקטגוריה של הפלט יכולה להיות רכה או soft - כלומר להוות התפלגות לא מנוונת(לא וקטור one-hot) מעל כל הקטגוריות. משמעות הדבר שהתפלגות יעד של הקטגוריה עבור פלטים מסוימים, הקרובים לכמה מרכזי קלסטרים, תשקף את זה בצורה הסתברותית. מה שמאומן במודל הזה זה מרכזי הקלסטרים. התפלגות קטגוריות עבור הפלט מחושבת למשל עם פונקציית סופטמקס המשקללת את הסיכוי של הפלט שייך לקלסטר המחושב באמצעות התפלגות גאוסית (למשל). שני המודלים אלו מאומנים יחד כאשר פונקציית לוס הוא מרחק KL בין התפלגויות הקטגוריות שהן מוציאות.

שני מודלים נוספים הם הדקודרים עם משקלים משותפים(בעלי שכבה אחת בלבד כל אחד). הראשון לוקח את הפלט של אנקודר הפלט ומעביר אותו בחזרה למרחב המקורי (עם לוס ריבועי למשל). הדקודר השני לוקח את חיזוי עבור הפלט ומעביר אותם לרחב המקורי של הפלט.

וזה וזה - מאמר נחמד ולא רגיל, מומלץ בחום

https://arxiv.org/pdf/2502.02996