# Review 10: Meta-Learning with Implicit Gradients

**Paper: https://arxiv.org/abs/1909.04630v1**

תחום: meta deep learning

תקציר מאמר: המאמר המכליל את MAML - model-agnostic meta learning. השיטה המוצעת במאמר הנסקר נקראת באופן לא מפתיע iMAML, כאשר האות הראשונה היא ממילה implicit. בגדול מאוד הרעיון ב- MAML המקורי זה לאמן רשת (החיצונית) לחישוב משקלים לרשתות הפנימיות למגוון רחב של טאסקים, דומיינים ארכיטקטורות הרשת. הרשת מאומנת בשביל למזער את הלוס הממוצע של הרשתות הפנימיות. הבעיה העיקרית בגישה הזו היא לבצע אופטימיזציה (גרדיאנט דסצנט) על הרשת החיצונית כי בעצם בשביל לחשב את הלוס לכל טאסק בכל רשת חיצונית צריך לבצע כמה איטרציות של באקפרופ. כמובן קשה מאוד לגלגל את האיטרציות האלו לבאקפרופ של הרשת החיצונית. אפשר לעשות קירוב מסדר ראשון אבל זה לא עובד כל כך טוב.

רעיון של מאמר: מה ש- imaml מציעים זה בעצם שני דברים:

1. לאמן את הרשת החיצונית כך שהמשקלים שהיא מוציאה יהיו כמה שיותר קרובים (מרחק ריבוע) ל"משקלים אופטימליים" של כל הרשתות הפנימיות. פשוט מוסיפים איבר רגולריזציה שמודד את המרחק הזה.

2. המשקלים האופטימליים של כל רשת פנימית הן אלו המביאים למינימום את הלוס (ראו תמונה המצורפת) עכשיו זה נראה מהמבט הראשון בלתי אפשרי לאמן דבר כזה כי לך תדע כמה איטרציות צריך לבצע בכל רשת פנימית וחישוב הגרדיינט מסתבך עוד יותר. וכאן בא הטריק המתמטי האלגנטי שמוכיח שאם מוסיפים את איבר רגולריזציה ריבועי שמבטל את הצורך לחשב אותו מפורשות (מכאן ה implicit). יש ביטוי סגור לגרדיאנט הזה.

יש עוד לא מעט טריקים מתמטיים במאמר הזה שבעזרתן הם מציעים אלגוריתם בר מימוש (אחד מהם זה חישוב היפוך של מטריצות מאוד גדולות בעזרת (conjugate gradient)

הם מראים תוצאות לא רעות דרך אגב.

בלוג המסביר את המאמר:blog

הרצאה מצוינת של יענק:youtube explanation