

Meta-Learning for Few-shot Domain Adaptation

Aman Hussain, Ivan Bardarov, Omar Elbaghdadi, Balint Hompot
University of Amsterdam

Introduction

- Domain adaptation: test samples from different distribution.
- Example: train on book reviews, test on movie reviews.
- Meta-learning: adapt to new task with few samples.
- **Is meta-learning well-suited for domain adaptation?**

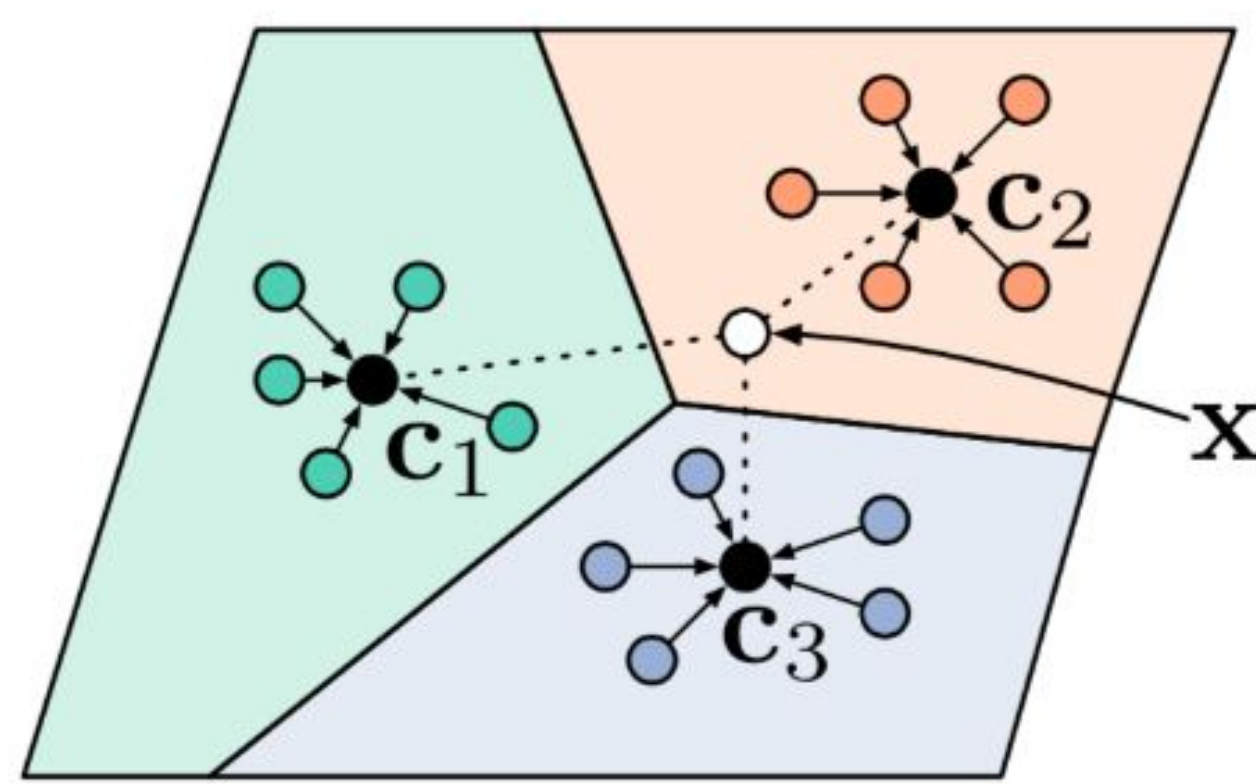


Figure 1. Prototypical networks

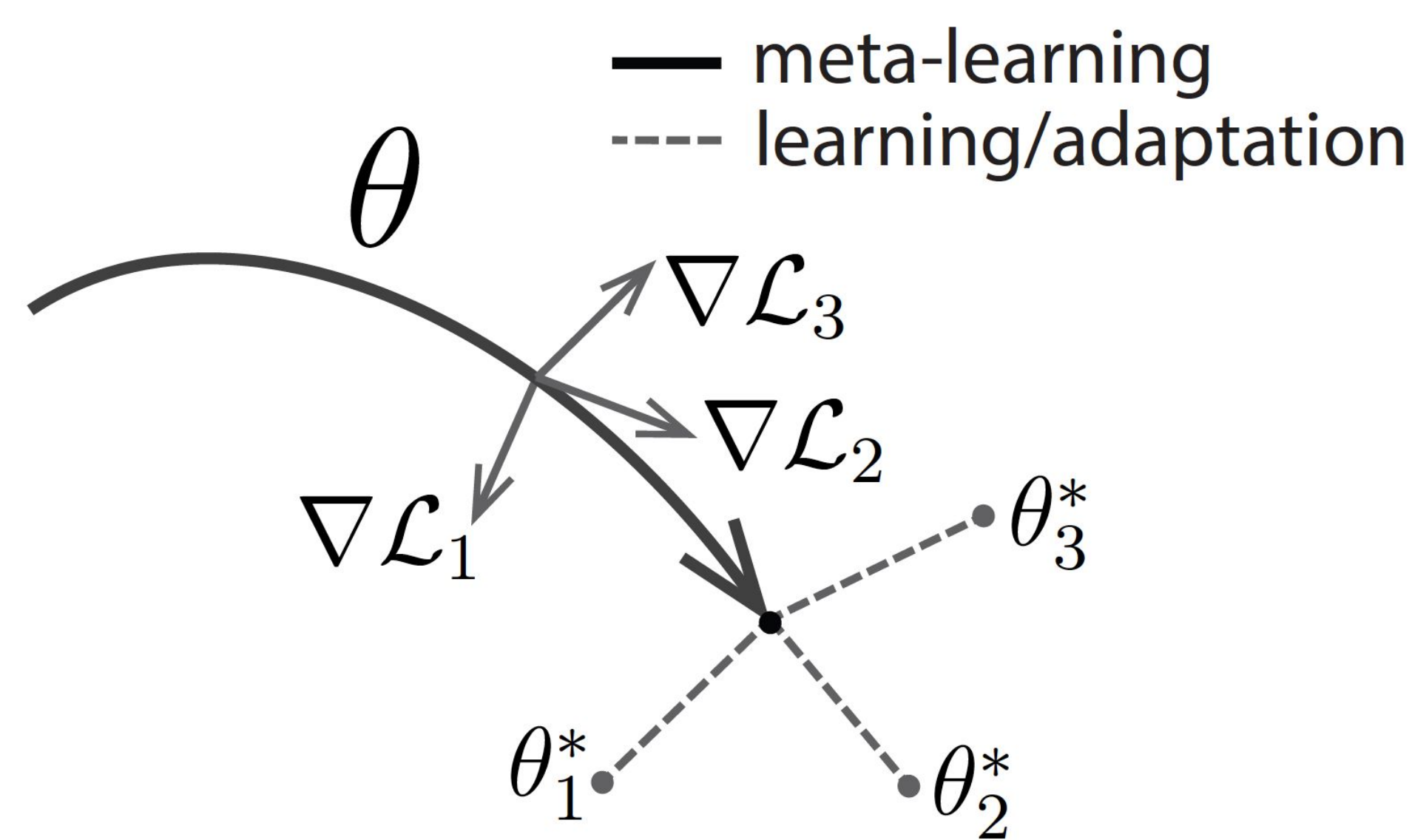
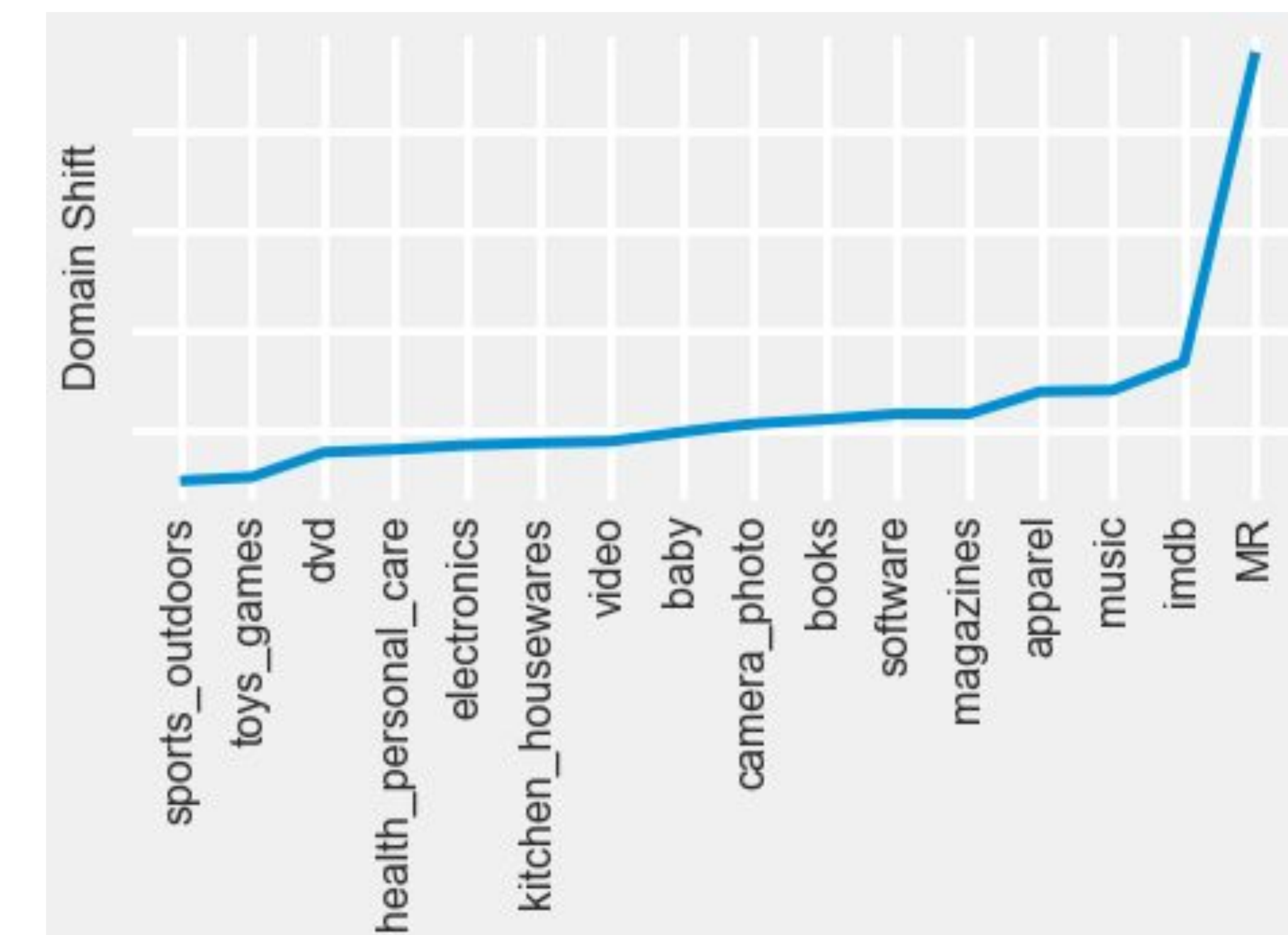


Figure 2. MAML

Methodology

- **Baseline: Multi-task Learning**
Classify sentiment regardless of domain.
- **Metric-based: Prototypical Networks**
Prototypes from k-shot examples of each sentiment from all domains.
- **Optimization-based: (FO)MAML**
Find a good starting point for k-shot fine-tuning to a new domain.



	DVD		MR	
	Limited data	Full data	Limited data	Full data
Baseline	83.7	92.3	75.6	82.6
Protonet	82.7	88.2	78.7	76.4
MAML	84.0	88.3	77.3	76.8

Table 1. Results

Experimental Setup

- **Data:** Sentiment of reviews from different domains.
- **16 Domains:** Train (14), Validation (1), Test (1).
- **Domain shift:** Low (dvd), High (MR).
- **5-shot 2-way classification:** Use 5 samples from each sentiment class.
- **Episodes:** domain specific *support* & *query* set.
- **Model:** BERT (layers 11-12 trained) + Feedforward.

Discussion & Conclusion

- MAML and Protonet similar in performance.
- **Multi-task better** than meta-learning on **full data**.
- Possible Reasons:
General domain-agnostic features already learnt by pre-training BERT.
- **Meta-learning better with limited data.**
- Possible Reasons:
Overfitting on train domains.

References

- [1] Snell, Jake, Kevin Swersky, and Richard Zemel. "Prototypical Networks for Few-Shot Learning,"
- [2] Finn, Chelsea, Pieter Abbeel, and Sergey Levine. "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks."
- [3] Van Asch, Vincent, and Walter Daelemans. "Using Domain Similarity for Performance Estimation."