Learning Under Weak Supervision:
Some Insights From Decision Theory

Julian Rodemann

Department of Statistics, LMU Munich
department@stat.uni-muenchen.de

Labeled data are scarce in a good deal of applied classification settings. This has given rise to the paradigm of semi-supervised learning (SSL), where information from unlabeled data is (partly) taken into account to improve inference drawn from labeled data in a supervised learning framework. Within SSL, an intuitive and widely used approach is referred to as self-training or pseudo-labeling [16, 5, 8, 9]. The idea is to fit an initial model to labeled data and iteratively assign pseudo-labels to some of the unlabeled data according to the model’s predictions. This process requires a criterion for pseudo-label selection (PLS), that is, the selection of unlabeled instances to be pseudo-labeled and added to the training data.

In my talk as part of the Young Statisticians Lecture Series, I will argue that PLS is nothing but a decision problem. This perspective clears the way for deploying several decision-theoretic approaches — first and foremost, finding Bayes optimal actions (selections of pseudo-labels) under common loss/utility functions. With the joint likelihood as utility, the Bayes optimal criterion turns out to be the posterior predictive of pseudo-samples [13]. Since the latter requires computing a possibly intractable integral, I will spotlight some analytical approximations based on Laplace’s method that circumvent expensive sampling-based evaluations of the posterior predictive [13]. Empirical evidence suggests that such a Bayesian approach to PLS can mitigate the confirmation bias in self-training that results from overfitting initial models [1]. Notably, the decision-theoretic embedding of PLS also yields the framework of optimistic/pessimistic superset learning [3, 4, 15] as max-max-/min-max-actions.

In the second part of my talk, I will discuss some extensions of Bayesian PLS based on generalized Bayesian decision theory [14]. They aim at robustifying PLS w.r.t. model selection, accumulation of errors and covariate shift [12]. What is more, I present Bayesian PLS with convex sets of prior and a regret-based updating rule. The set of priors can reflect uncertainty regarding prior information, but might as well represent priors near ignorance, see e.g. [2, 7, 6, 10, 11]. The talk will conclude with several applications of (generalized) Bayesian PLS on real-world classification problems under weak supervision.

Keywords: Semi-Supervised Learning · Self-Training · Approximate Inference · Bayesian Decision Theory · Generalized Bayes · Robustness
References

12. Rodemann, J., Fischer, S., Schneider, L., Nalenz, M., Augustin, T.: Not all data are created equal: Lessons from sampling theory for adaptive machine learning (2022), Poster presented at International Conference on Statistics and Data Science (ICSDS), Institute of Mathematical Statistics (IMS)