

# LEARNING FROM A LOT: EMPIRICAL BAYES IN HIGH-DIMENSIONAL PREDICTION SETTINGS

Wiel, Mark van de

Vrije Universiteit, Amsterdam

Empirical Bayes is a versatile approach to ‘learn from a lot’ in two ways: first, from a large number of variables and second, from a potentially large amount of prior information, e.g. stored in public repositories. We review applications of a variety of empirical Bayes methods to a broad spectrum of prediction methods including penalized regression, random forest, linear discriminant analysis, and Bayesian models with sparse or dense priors. We discuss ‘formal’ empirical Bayes methods which maximize the marginal likelihood, but also more informal approaches based on other data summaries. Empirical Bayes is contrasted to cross-validation and full Bayes. Hybrid approaches will be discussed.

Empirical Bayes is particularly useful when the prior or penalty contains multiple parameters that model a priori information on the variables, termed ‘co-data’. In practice, one often seeks for predictors that require measuring a few variables only. We will show that in such a context the combination of empirical Bayes and co-data can render a large improvement in predictive performance and typically stabilizes variable selection. Methods will be illustrated on several cancer genomics applications. Finally, we shortly discuss extensions to other problems such as network reconstruction and drug target models.