Simultaneous Model Selection and Learning through Parameter-free Stochastic Gradient Descent

Francesco Orabona
Yahoo! Labs NY, USA
francesco@orabona.com

Stochastic gradient descent algorithms for training linear and kernel predictors are gaining more and more importance, thanks to their scalability. While various methods have been proposed to speed up their convergence, the issue of the model selection phase has often been ignored in the literature. In fact, in theoretical works most of the time unrealistic assumptions are made, for example, on the prior knowledge of the norm of the optimal solution. Hence, costly validation methods remain the only viable approach in practical applications. In this talk, we show how a family of kernel-based stochastic gradient descent algorithms can perform model selection while training, with no parameters to tune, nor any form of cross-validation, and only one pass over the data. These algorithms are based on recent advancements in online learning theory in unconstrained settings. Optimal rates of convergence will be shown under standard smoothness assumptions on the target function, as well as empirical results.