ABSTRACT

In high-dimensional classification settings, we wish to seek a balance between attaining high power while ensuring control over a target loss function. In many problems, the points most likely to be misclassified are those who lie near the decision boundary of the given classification method. Often, these uninformative points should not be classified as they are noisy and do not exhibit strong signals. In this paper, we introduce the Thresholding Method to parameterize the problem of determining which points exhibit strong signals and should be classified. We demonstrate the empirical performance of the method in providing loss function control at a desired level, as well as explore how the method assuages the effect of overfitting. Finally, we show the flexibility of the Thresholding Method through the application of the method to a variety of real data and simulated data to control various loss functions and provide control to multiple classification models.