Discussion on "Random projection ensemble classification" by Drs. Cannings and Samworth

Dr. Xin Tong

Assistant Professor
Mailing Address: BRI 308
Department of Data Sciences and Operations
Marshall School of Business
University of Southern California
Los Angeles, CA 90089, USA
Email: xint@marshall.usc.edu

Dr. Jingyi Jessica Li

Assistant Professor
Mailing Address: 8125 Math Sciences Bldg.
Department of Statistics
University of California, Los Angeles
Los Angeles, CA 90095-1554, USA
Email: jli@stat.ucla.edu

We congratulate Dr. Timmothy I. Cannings and Dr. Richard J. Samworth for their innovative and thought-provoking read paper titled "Random projection ensemble classification." In recent years, many classification methods have been developed for high-dimensional settings, where the feature dimension p is comparable to or larger than the sample size n. In the literature, most of the existing work aimed to build a specific procedure (e.g., screening and penalization approaches) to effectively reduce model complexity. From a different and novel perspective, Cannings and Samworth's work developed a theory-backed ensemble classification procedure, which first projects features into many lower dimensional spaces so that "base" classifiers (e.g., LDA and QDA) can be applied to the projected data without any modification, and second aggregates these low-dimensional classifiers via a proper voting scheme. Their paper lays a good foundation that motivates us to think about many questions including the following three.

(I) What is the consequence of relaxing the assumption (A.3)? This assumption states: "There exists a projection $A^* \in \mathcal{A}$ such that

$$P_X(\{x \in \mathbb{R}^p : \eta(x) \geq 1/2\}\Delta\{x \in \mathbb{R}^p : \eta^{A^*}(A^*x) \geq 1/2\}) = 0\,,$$

where Δ denotes the symmetric difference between sets." Essentially, this assumption means there exists one projection that leads to an oracle decision boundary essentially the same as the oracle decision boundary in the original feature space. While this is a reasonable and convenient assumption, it can probably be relaxed, and the discrepancy between the two oracle decision boundaries, that is the discrepancy between the original Bayes classifier (in $\mathbb{R}^p$) and the best projected Bayes classifier (in $\mathbb{R}^d$), can perhaps show up in the upper bound of the excess error.

(II) In the paper, the voting threshold α is to mimic α^* in equation (12):

$$\alpha^* = \operatorname*{arg\,min}_{\alpha' \in [0,1]} \left[\pi_1 G_{n,1}(\alpha') + \pi_2 \left\{ 1 - G_{n,2}(\alpha') \right\} \right] .$$

This is a very natural choice when the classification target is to minimize the classification error (i.e., risk) and when the empirical proportions $\hat{\pi}_1$ and $\hat{\pi}_2$ are good estimates of of π_1 and π_2 respectively. How would the authors suggest the choice for α when we are interested in a type I/II error weighting different from that implied by the class priors, or when there lack good estimates of π_1 and π_2?

(III) This comment is related to the first one. If we relax (A.3) to achieve the best performance bounds in Section 4, we no longer prefer d as small as possible (while validating (A.3)). We expect that a best choice of d will depend on the discrepancy between the two Bayes classifiers in R^p and R^d.