Discussion of "Random-projection ensemble classification" by Timothy I. Cannings and Richard J. Samworth

J. R. Statist. Soc. B, Vol. **79**, Part 4, pp. 1–38 (2017)

Miguel de Carvalho (University of Edinburgh), **Garritt L. Page**, and **Bradley Barney** (Brigham Young University).

We congratulate the authors for proposing a sturdy method based on randomly compressing feature vectors prior to classification. Below, we focus on connecting the random-projection ensemble classifier with ideas and concepts from *compressed classification* and *compressed regression* methods. Let $\mathcal{A} = \{A \in \mathbb{R}^{d\times p} : AA^{\mathrm{T}} = I_{d\times d}\}$ be the so-called Steifel manifold. Similarly to Page *et al.* (2013), in the paper under discussion the author's first compress the covariates by using projection matrices, but a key difference is that here the authors consider a set of independent projections, $A_1, \dots, A_{B_1} \in \mathcal{A}$, whereas in the latter paper a single projection matrix $A \in \mathcal{A}$ is considered—and treated as a Bayesian parameter. In particular, Page *et al.* (2013) considers a nonparametric Bayesian approach which leads to a principal subspace classifier for a setting similar to the one in the current manuscript, and assigns to A a (conjugate) von Mises–Fisher prior distribution on the Steifel manifold. In an analogy to the author's claim that "in a similar spirit to subsampling and bootstrap sampling, we can can think of each random projection as a perturbation of the original data," the compressing paradigms described above—based on a single but random A—keep the data as fixed, and posterior sampling about good directions along which to project the data is itself target. Both compressing principles (single $A \in \mathcal{A}$ as Bayesian parameter, *vs* ensemble of random $A_1, \dots, A_{B_1} \in \mathcal{A}$) seem to have their own merits, and we wonder if the authors could comment on this remark. On another note, the recently proposed compressed regression approach by Guhaniyogi and Dunson (2015) is even closer to the authors proposal, in the sense that it projects data into an ensemble of directions and uses model averaging to arrive at a final regression model. The focus of the latter paper is on regression itself though, but we also wonder about the author's view on this. Finally, the practitioner could be left with the question: "How likely is it for the ensemble classifier to improve over the base classifier on the original feature vectors?"

References

Page, G. L., Bhattacharya, A., and Dunson, D. B. (2013). Classification via Bayesian nonparametric learning of affine subspaces. *Journal of the American Statistical Association*, **108**, 187–201.

Guhaniyogi, R. and Dunson, D. B. (2015). Bayesian compressed regression. *Journal of the American Statistical Association*, **110**, 1500–1514.