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

Rico Blaser & Piotr Fryzlewicz
Department of Statistics
London School of Economics
Houghton Street
London WC2A 2AE
UK

February 26, 2017

We would like to congratulate the authors on their thought-provoking and well-written paper in this promising area of research.

In our work on random rotation ensembles (Blaser and Fryzlewicz, 2016) we randomly rotated the feature space prior to applying a high-dimensional classifier. In the present paper, the high-dimensional feature space is projected into a lower-dimensional space prior to applying a low-dimensional classifier. The two strategies are closely related.

In the current paper, projections are performed randomly under the Haar measure. Interestingly, a random rotation followed by a random axis-aligned projection in which d-of-p features are retained is identical to the random projection described in the paper. Our tree-based ensemble classifiers perform axis-aligned projections after rotation and thus effectively describe a random-projection ensemble, whereby the final classification is restricted to a tree-based model.

More generally, we believe decoupling rotation from dimension reduction and dimension reduction from classification is desirable. In particular, such a decomposition addresses the question, if the benefit of a particular random projection arises from an advantageous viewpoint at the problem due to the rotation or from an effective dimension reduction due to the feature selection, as the two operations can be analysed and optimised separately.

The authors also provide interesting insights on the selection of retained projections and the determination of the voting threshold. They note that most random projections are unhelpful in classification, a pattern that we have also observed for random rotations.

Hence, a natural question to ask is how we can identify (or explicitly generate) only the most helpful projections.

One way to address this issue is by performing a large number of candidate projections and retaining only the most successful candidates. The authors of the present paper recommend retaining 2% of the generated projections by default, substantially fewer than the 90% of rotations we examined. More accuracy is achieved at the expense of a higher overhead. Alternatively, analytical methods such as PCA, can be used to determine successful rotations. In Rodriguez et al. (2006), this approach is used for random subsets of the features.

Different subsets of the data frequently benefit from different rotations; for non-linear decision boundaries this is quite evident. Hence, it might also be useful to construct a classifier that rotates different sections of the data independently.

The data-driven selection of the voting threshold suggested by the authors is insightful but is not straight-forward to generalise to multi-class problems.

References

Blaser, R. and Fryzlewicz, P. (2016) Random rotation ensembles. *Journal of Machine Learning Research* 17, 1-26.

Rodriguez, JJ. et al. (2006) Rotation forest: a new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 1619-1630.