Discussion of Random Projection Ensemble Classification by Timothy I. Cannings and Richard J. Samworth

Wicher Bergsma and Haziq Jamil
London School of Economics and Political Science

March 30, 2017

The authors present impressive theoretical results. However, we are not yet convinced about their practical use: as summarized in Table 1, we ran our own favourite classifier — Gaussian process regression using fractional Brownian motion (GPR-FBM) — and for five of the eight data sets we obtained results unequivocally better than random projection (RP) ensembles, and for the remaining three we obtained mixed results. Furthermore, our preliminary analyses indicate RP ensembles worsen GPR-FBM classification, but this could be due to the small B_1 and B_2 we chose due to time constraints ($B_1 = 30$ and $B_2 = 5$). Thus, although RP ensemble methods can demonstrably improve frequently poor methods such LDA and $k$nn, we wonder if they can improve good methods. If not, what then is the advantage of using RP ensembles?

Our methodology was as follows. We fitted the model

$$y_i = f(x_i) + \varepsilon_i$$

where $y_i \in \{0, 1\}$ is the class of unit i, the ε_i are i.i.d. normal errors, the x_i are p-dimensional covariates, and f has a zero mean Gaussian process distribution. In Table 1 we show results for GPR-linear, with covariance kernel

$$K(x, x') = \lambda(x - \bar{x})^\top (x' - \bar{x})$$

and for GPR-FBM-γ with covariance kernel

$$K_\gamma(x, x') = \tilde{K}_\gamma(x, x') - \frac{1}{n}\sum_{j=1}^{n} \tilde{K}_\gamma(x, x_j) - \frac{1}{n}\sum_{i=1}^{n} \tilde{K}_\gamma(x_i, x') + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \tilde{K}_\gamma(x_i, x_j)$$

where

$$\tilde{K}_\gamma(x, x') = \frac{\lambda}{2}\left(\|x\|^{2\gamma} + \|x'\|^{2\gamma} - \|x - x'\|^{2\gamma}\right)$$

The parameter $\gamma \in (0, 1)$ is the Hurst coefficient, and we took $\gamma = 1/2$ and $\gamma = \hat{\gamma}$, where $\hat{\gamma}$ is the maximum likelihood estimator of γ. We also estimated scale parameter λ using maximum likelihood. Note that GPR-linear is essentially ridge regression. We omitted GPR-FBM-0.99999 from the table, which gives results competitive with the RP ensembles for the Hill-Valley data. Code for replication of most results is provided at https://haziqjamil.github.io/rec-jrss-reply/ (some results were obtained with Mathematica code).

In summary, it appears there is a mismatch between theory and practice. Theory tells us that the curse of dimensionality is a problem for high-dimensional classification; for example, according to Hastie and Tibshirani (1986, page 305) in their seminal paper, "the chief motivation for the additive model" is that "it is well known that smoothers break down in higher dimensions [because] the curse of dimensionality takes its toll". However, in view of the success of, e.g., SVMs and GPR, and the results in Table 1, it appears to us that in practice smoothers do *not* break down in high dimensions. So it might be wondered, is the curse of dimensionality a straw man undeserving of the broad attention it is receiving?

Finally, we remark that our own GPR methodology, related to the above, will appear soon on arxiv. In particular, we will propose a flexible empirical Bayes methodology based on the Fisher information for the regression function, which in a well-defined sense can improve on Tikhonov regularization, and can further improve some of the GPR results in Table 1.

References

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, *1*(3), 297-310.

Method	Eye state data			Ionosphere data		
	$n = 50$	$n = 200$	$n = 1000$	$n = 50$	$n = 100$	$n = 200$
RP5-LDA	$42.1_{0.38}$	$38.6_{0.29}$	$36.3_{0.21}$	$13.1_{0.38}$	$10.8_{0.25}$	$9.8_{0.26}$
RP5-QDA	$39.0_{0.39}$	$32.4_{0.42}$	$30.9_{0.87}$	$8.1_{0.37}$	$6.2_{0.37}$	$5.2_{0.20}$
RP5-$k$nn	$39.4_{0.39}$	$26.9_{0.27}$	$13.5_{0.19}$	$13.1_{0.46}$	$7.4_{0.25}$	$5.4_{0.19}$
RP5-GPR-linear				$36.0_{0.11}$	$35.9_{0.16}$	$35.5_{0.28}$
RP5-GPR-FBM-1/2				$36.2_{0.10}$	$35.7_{0.16}$	$35.8_{0.29}$
GPR-linear	$46.6_{0.92}$	$42.3_{0.95}$	$37.5_{0.48}$	$17.3_{0.30}$	$15.6_{0.22}$	$13.7_{0.24}$
GPR-FBM-1/2	$37.0_{0.27}$	$24.0_{0.13}$	$10.3_{0.08}$	$9.6_{0.04}$	$6.6_{0.07}$	$5.2_{0.08}$
GPR-FBM-$\hat{\gamma}$				$12.1_{0.32}$	$8.6_{0.26}$	$6.4_{0.20}$

Method	Mice data			Hill-valley data		
	$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
RP5-LDA		$25.2_{0.30}$	$23.6_{0.26}$	$36.8_{0.84}$	$36.5_{0.85}$	$32.6_{1.06}$
RP5-QDA		$18.2_{0.29}$	$16.1_{0.24}$	$44.4_{0.34}$	$43.6_{0.31}$	$41.1_{0.33}$
RP5-$k$nn		$11.2_{0.29}$	$2.2_{0.10}$	$49.1_{0.24}$	$47.3_{0.26}$	$36.4_{0.29}$
RP5-GPR-linear						
RP5-GPR-FBM-1/2						
GPR-linear*	$6.2_{0.23}$	$3.2_{0.12}$	$2.2_{0.19}$	$50.2_{0.14}$	$50.0_{0.20}$	$48.5_{0.59}$
GPR-FBM-1/2*	$8.5_{0.26}$	$3.0_{0.15}$	$0.3_{0.07}$	$45.3_{0.09}$	$49.8_{0.08}$	$50.7_{0.12}$
GPR-FBM-$\hat{\gamma}$*	$4.5_{0.21}$	$1.0_{0.11}$	$0.0_{0.00}$	$45.0_{0.09}$	$49.7_{0.10}$	$50.7_{0.13}$

Method	Musk data			Arrhythmia data		
	$n = 100$	$n = 200$	$n = 500$	$n = 50$	$n = 100$	$n = 200$
RP5-LDA	$14.6_{0.31}$	$12.2_{0.23}$	$10.2_{0.15}$	$33.2_{0.42}$	$30.2_{0.35}$	$27.5_{0.30}$
RP5-QDA	$12.1_{0.27}$	$9.9_{0.18}$	$8.6_{0.13}$	$30.5_{0.33}$	$28.3_{0.26}$	$26.3_{0.28}$
RP5-$k$nn	$11.8_{0.27}$	$9.7_{0.21}$	$8.0_{0.15}$	$33.5_{0.40}$	$30.2_{0.33}$	$27.1_{0.31}$
RP5-GPR-linear	$15.2_{0.11}$	$15.5_{0.09}$	$15.5_{0.09}$	$47.3_{0.32}$	$47.5_{0.40}$	$46.2_{0.33}$
RP5-GPR-FBM-1/2	$15.3_{0.10}$	$15.5_{0.10}$	$15.2_{0.11}$	$47.1_{0.33}$	$46.7_{0.27}$	$46.4_{0.25}$
GPR-linear	$14.7_{0.37}$	$12.3_{0.52}$	$9.2_{0.37}$	$42.2_{0.65}$	$33.2_{0.51}$	$27.7_{0.34}$
GPR-FBM-1/2	$9.8_{0.11}$	$7.0_{0.07}$	$5.0_{0.06}$	$32.8_{0.46}$	$27.8_{0.23}$	$25.1_{0.24}$
GPR-FBM-$\hat{\gamma}$				$28.4_{0.27}$	$25.2_{0.22}$	

Method	Activity recognition data			Gisette data		
	$n = 50$	$n = 200$	$n = 1000$	$n = 50$	$n = 200$	$n = 1000$
RP5-LDA	$0.18_{0.02}$	$0.10_{0.01}$	$0.01_{0.00}$	$15.8_{0.41}$	$10.6_{0.17}$	$9.4_{0.15}$
RP5-QDA	$0.15_{0.02}$	$0.09_{0.01}$	$0.00_{0.00}$	$15.5_{0.40}$	$10.5_{0.19}$	$9.4_{0.16}$
RP5-$k$nn	$0.21_{0.02}$	$0.11_{0.01}$	$0.01_{0.00}$	$16.0_{0.46}$	$11.1_{0.17}$	$9.6_{0.16}$
RP5-GPR-linear						
RP5-GPR-FBM-1/2						
GPR-linear	$0.06_{0.01}$	$0.00_{0.00}$	0_{0}	$12.4_{0.34}$	$6.9_{0.07}$	$4.5_{0.08}$
GPR-FBM-1/2	$48.44_{0.26}$	$7.39_{0.26}$	$0.04_{0.00}$	$13.2_{0.25}$	$7.4_{0.08}$	$4.5_{0.09}$
GPR-FBM-$\hat{\gamma}$						

Table 1: Misclassification rates for eight data sets: GPR versus RP ensembles. Blank spaces indicate results are currently unavailable.
*: missing values removed (GPR for mice data)