kernDeepStackNet: An R package for tuning kernel deep stacking networks

Thomas Welchowski [1] *and Matthias Schmid* [1]

Kernel deep stacking networks [1] (KDSNs) are a novel method for supervised learning in biomedical research and belong to the class of deep learning methods. Deep learning uses multiple layers of non-linear transformations to derive higher abstractions of the input features [2]. These can more efficiently represent complex dependencies of joint distributions [3]. Training of deep artificial neural networks is a non-convex optimization problem, which may result in local optima and slow convergence. Kernel deep stacking networks are a computationally faster alternative, which is based on solving multiple convex optimization problems by combined kernel ridge regressions and random Fourier transformations.

Tuning of KDSNs is a challenging task, as there are multiple hyper parameters to tune. We propose a new data-driven tuning strategy for KDSNs using model based optimization (MBO) [4]. The performance criterion is RMSE on cross validation samples, and noisy Kriging is used as surrogate model. New design points are chosen by maximisation of the expected improvement criterion.

Numerical studies show, that the MBO approach is substantially faster than traditional grid search strategies. Further analysis of real data sets demonstrates, that tuned KDSNs are competetive to other state-of-art machine learning techniques in terms of prediction accuracy. The fitting and tuning procedures are implemented in the R package *kernDeepStackNet*.

New developments extend the KDSN tuning framework to include variable selection, dropout [5] and L1 penalization. Variable selection is based on the non-linear randomized dependence coefficient [6] in combination with sure independence screening [7]. Numerical simulations show that these sparse kernel deep stacking networks (SKDSN) improve generalization error compared to KDSNs. SKDSNs are also competitive to other state-of-the-art supervised learning methods in high dimensional environments, like random forests and gradient boosting.

[1] Department of Medical Biometry, Informatics and Epidemiology, Rheinische Friedrich-Wilhelms-Universität Bonn, Sigmund-Freud-Str. 25, 53127 Bonn,Germany

welchow@imbie.meb.uni-bonn.de, matthias.schmid@imbie.uni-bonn.de

References

1 Huang P-S, Deng L, Hasegawa-Johnson M, He X. , 2013, Random features for kernel deep convex network. In: Ward R, Deng L, editors. Proceedings of the IEEE international conference on acoustics, speech, and signal processing, NewYork, USA; pages 3143–7.

2 Li Deng, Dong Yu, 2014, Deep learning: Methods and Applications, Foundations and Trends® in Signal Processing, Volume 7, Issues 3-4

3 Bengio Y, Delalleau O. On the expressive power of deep architectures. In: Kivinen J, Szepesvari C, Ukkonen E, Zeugmann T, editors. Algorithmic learning theory. Berlin: Springer; 2011. p. 18–36.

4 Welchowski, Schmid, 2016, A framework for parameter estimation and model selection in kernel deep stacking networks, Artificial Intelligence in Medicine, Volume 70, Pages 31–40

5 Nitish Srivastava and Geoffrey Hinton and Alex Krizhevsky et al., 2014, Dropout: A simple way to prevent neural networks from overfitting, Journal of Machine Learning Research, Volume 15, Pages 1929-1958

6 David Lopez-Paz and Philipp Hennig and Bernhard Schölkopf, 2013, The randomized dependence coefficient, arXiv:1304.7717

7 Jianqing Fan, 2008, Sure independence screening for ultra high dimensional feature space, Journal of Royal Statistical Society Series B, Volume 70, Part 5, Pages 849–911