

High-Dimensional Explainable ANOVA Approximation

Michael Schmischke¹

¹) TU Chemnitz
michael.schmischke@math.tu-chemnitz.de

Building models for creating predictions based on empirical data is a current and active research topic with numerous applications. The amount of data we collect today increases rapidly resulting in high-dimensional datasets and corresponding regression or classification problems. There is a number of classical machine learning methods like support vector machines, neural networks, and decision trees, see e.g. [15, 1, 5], to tackle these problems. However, the question of how the predictions come to pass, i.e., the interpretability of these models, is ever more important. With this information one may omit certain (possibly expensive) measurements if they have a small influence or use it to influence predictions to a desired outcome. While there is new research in the area of interpretability of the classical methods, see e.g. [32, 23], those models do not intrinsically allow for it.

In this talk we present the approximation method introduced and applied in [26, 27, 28]. It is based on the analysis of variance (ANOVA) decomposition, cf. [6, 29, 21, 19, 16, 13], for functions in \mathcal{L}_2 spaces and corresponding complete orthonormal systems. The ANOVA decomposition in this setting uniquely decomposes a function into terms that correspond to variable couplings or variable interactions. The assumption we use for tackling the curse of dimensionality is that the function has a low superposition dimension, i.e., it can be explained well by low-dimensional interactions, cf. [6, 19, 10, 24, 14]. In [26] we showed that functions of certain smoothness types have this property. Moreover, it has been theorized that most real world applications consist only of low-order interactions relating to sparsity-of-effects, cf. [35], or the Pareto principle. The method pairs this idea with the concept of grouped index sets together with grouped transformations we presented in [2]. We are able to use different orthonormal systems that allow for fast transformations. In particular, the grouped transformations are based on the non-equispaced fast Fourier transform and the non-equispaced fast cosine transform, see [18, 25, 17].

From the model we immediately obtain importance information on the variable couplings by using global sensitivity indices or Sobol indices, cf. [33, 34, 21], and sensitivity analysis, see [31]. Therefore, we are able to interpret the model and gain information about the importance of attributes and attribute interactions. Moreover, we improve it by a number of techniques using this information. By the utilization of attribute rankings we can remove an unimportant variable entirely and reduce the dimensionality of the problem. It is also possible to find that the representation of the data in the groups, i.e., the ANOVA decomposition, is sparse. In this case we have that some variable interactions do not influence the approximation (significantly) and can therefore be discarded. These techniques allow us to build an *active set* of couplings that will in the end represent the model we use for predictions. This simultaneously gives us a control mechanism for the complexity of the model and combat overfitting.

The technique is also related to low-dimensional structures and active subspace methods [12, 8, 9] as well as random features [30, 7, 36, 20, 14]. The main difference to random features is that random features draws weights or in our language indices/frequencies at random and uses a different optimization problem.

We present numerical results, cf. [28], for synthetic test functions, the Friedman functions, see [22, 3, 4], as an example of how it performs on synthetic data with Gaussian noise and compare our findings to previously obtained benchmark results in the same setting. Moreover, we show performance on application datasets from the UCI database [11] and other sources. Each datasets provides a different challenge and we compare our results to different machine learning methods. We observe very promising results and in many cases outperform previous benchmark experiments.

References

- [1] C. C. Aggarwal. *Data Classification: Algorithms and Applications*. Chapman & Hall/CRC, 1st edition, 2014.
- [2] F. Bartel, D. Potts, and M. Schmiscke. Grouped transformations in high-dimensional explainable ANOVA approximation. *ArXiv e-prints 2010.10199*, 2020.
- [3] G. Beylkin, J. Garcke, and M. Mohlenkamp. Multivariate regression and machine learning with sums of separable functions. *SIAM J. Scientific Computing*, 31:1840–1857, 2009.
- [4] P. Binev, W. Dahmen, and P. Lamby. Fast high-dimensional approximation with sparse occupancy trees. *J. Comput. Appl. Math.*, 235(8):2063 – 2076, 2011.
- [5] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer New York, Berlin-Heidelberg, 2016.
- [6] R. Caffisch, W. Morokoff, and A. Owen. Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension. *J. Comput. Finance*, 1(1):27–46, 1997.
- [7] R. Chitta, R. Jin, and A. K. Jain. Efficient kernel clustering using random Fourier features. In *2012 IEEE 12th International Conference on Data Mining*. IEEE, 2012.
- [8] P. G. Constantine, E. Dow, and Q. Wang. Active subspace methods in theory and practice: Applications to kriging surfaces. *SIAM J. Sci. Comput.*, 36(4):A1500–A1524, 2014.
- [9] P. G. Constantine, A. Eftekhari, J. Hokanson, and R. A. Ward. A near-stationary subspace for ridge approximation. *Comput. Methods Appl. Mech. Engrg.*, 326:402–421, 2017.
- [10] R. DeVore, G. Petrova, and P. Wojtaszczyk. Approximation of functions of few variables in high dimensions. *Constr. Approx.*, 33(1):125–143, 2010.

- [11] D. Dua and C. Graff. UCI machine learning repository, 2017.
- [12] M. Fornasier, K. Schnass, and J. Vybiral. Learning functions of few arbitrary linear parameters in high dimensions. *Found. Comput. Math.*, 12(2):229–262, 2012.
- [13] C. Gu. *Smoothing Spline ANOVA Models*. Springer New York, 2013.
- [14] A. Hashemi, H. Schaeffer, R. Shi, U. Topcu, G. Tran, and R. Ward. Function approximation via sparse random features. *ArXiv e-prints 2103.03191*, 2021.
- [15] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning - Data Mining, Inference, and Prediction*. Springer Science & Business Media, Berlin Heidelberg, 2013.
- [16] M. Holtz. *Sparse grid quadrature in high dimensions with applications in finance and insurance*, volume 77 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2011.
- [17] J. Keiner, S. Kunis, and D. Potts. NFFT 3.5, C subroutine library. <http://www.tu-chemnitz.de/~potts/nfft>. Contributors: F. Bartel, M. Fenn, T. Görner, M. Kircheis, T. Knopp, M. Quellmalz, M. Schmischke, T. Volkmer, A. Vollrath.
- [18] J. Keiner, S. Kunis, and D. Potts. Using NFFT3 - a software library for various nonequispaced fast Fourier transforms. *ACM Trans. Math. Software*, 36:Article 19, 1–30, 2009.
- [19] F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, and H. Woźniakowski. On decompositions of multivariate functions. *Math. Comp.*, 79(270):953–966, 2009.
- [20] Z. Li, J.-F. Ton, D. Oglic, and D. Sejdinovic. Towards a unified analysis of random Fourier features. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 3905–3914. PMLR, 09–15 Jun 2019.
- [21] R. Liu and A. B. Owen. Estimating mean dimensionality of analysis of variance decompositions. *J. Amer. Statist. Assoc.*, 101(474):712–721, 2006.
- [22] D. Meyer, F. Leisch, and K. Hornik. The support vector machine under test. *Neurocomputing*, 55(1):169 – 186, 2003. Support Vector Machines.
- [23] G. Montavon, W. Samek, and K.-R. Müller. Methods for interpreting and understanding deep neural networks. *Digit. Signal Process.*, 73:1–15, 2018.
- [24] A. Owen. Effective dimension of some weighted pre-Sobolev spaces with dominating mixed partial derivatives. *SIAM J. Numer. Anal.*, 57(2):547–562, 2019.
- [25] G. Plonka, D. Potts, G. Steidl, and M. Tasche. *Numerical Fourier Analysis. Applied and Numerical Harmonic Analysis*. Birkhäuser, 2018.

- [26] D. Potts and M. Schmischke. Approximation of high-dimensional periodic functions with Fourier-based methods. *SIAM J. Numer. Anal. (accepted)*, 2019. , arXiv: 1907.11412 [math.NA].
- [27] D. Potts and M. Schmischke. Learning multivariate functions with low-dimensional structures using polynomial bases. *ArXiv e-prints 1912.03195*, 2019.
- [28] D. Potts and M. Schmischke. Interpretable approximation of high-dimensional data. *ArXiv e-prints 2103.13787*, 2021.
- [29] H. Rabitz and O. F. Alis. General foundations of high dimensional model representations. *J. Math. Chem.*, 25:197–233, 1999.
- [30] A. Rahimi and B. Recht. Random features for large-scale kernel machines. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2008.
- [31] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, Ltd., 2008.
- [32] W. Samek, T. Wiegand, and K.-R. Müller. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *ArXiv e-prints 1708.08296*, 2017.
- [33] I. M. Sobol. On sensitivity estimation for nonlinear mathematical models. *Keldysh Applied Mathematics Institute*, 1:112–118, 1990.
- [34] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simulation*, 55(1-3):271–280, 2001.
- [35] C. F. J. Wu and M. S. Hamada. *Experiments - Planning, Analysis, and Optimization*. John Wiley & Sons, New York, 2011.
- [36] T. Yang, Y.-F. Li, M. Mahdavi, R. Jin, and Z.-H. Zhou. Nyström method vs random Fourier features: A theoretical and empirical comparison. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS’12, page 476–484, Red Hook, NY, USA, 2012. Curran Associates Inc.