Intrinsically Interpretable Machine Learning In Computer Aided Diagnosis

S. Ghosh, Sreejita

DOI:
10.33612/diss.175627883

Publication date:
2021

Bibliography

Arlt, W., Walker, E. A., Draper, N., Ivison, H. E., Ride, J. P., Hammer, F., Chalder, S. M., Borucka-Mankiewicz, M., Hauffa, B. P., Malunowicz, E. M. et al.: 2004, Congenital adrenal hyperplasia caused by mutant p450 oxidoreductase and human androgen synthesis: analytical study, *The Lancet* **363**(9427), 2128–2135.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R. et al.: 2020, Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, *Information Fusion* **58**, 82–115.

Artelt, A. and Hammer, B.: 2019, On the computation of counterfactual explanations–a survey, *arXiv preprint arXiv:1911.07749* .

Azur, M. J., Stuart, E. A., Frangakis, C. and Leaf, P. J.: 2011, Multiple imputation by chained equations: what is it and how does it work?, *International journal of methods in psychiatric research* **20**(1), 40–49.

Backhaus, A. and Seiffert, U.: 2014, Classification in high-dimensional spectral data: Accuracy vs. interpretability vs. model size, *Neurocomputing* **131**, 15–22.

Baranowski, E., Bunte, K., Shackleton, C., Taylor, A., Hughes, B., Biehl, M., Tino, P., Guran, T. and Arlt, W.: 2016, Steroid metabolomics for diagnosis of inborn steroidogenic disorders-bridging the gap between clinician and scientist through computational approaches, *Society for Endocrinology BES 2016*, Vol. 44, BioScientifica.

Baranowski, E. S., Arlt, W. and Idkowiak, J.: 2018, Monogenic disorders of adrenal steroidogenesis, *Hormone research in paediatrics* **89**(5), 292–310.

Baranowski, E. S., Ghosh, S., Shackleton, C. H., Taylor, A. E., Hughes, B. A., Gilligan, L. C., Utari, A., Faradz, S. M., der Grinten Hedi, L., Biehl, M. et al.: 2019, Steroid metabolomics: a rapid computational approach for accurate differentiation of inborn disorders of steroidogenesis, *21st European Congress of Endocrinology*, Vol. 63, BioScientifica.

Bibal, A. and Frénay, B.: 2016, Interpretability of machine learning models and representations: an introduction., *ESANN*.

Biehl, M., Hammer, B. and Villmann, T.: 2013, Distance measures for prototype based classification, *International Workshop on Brain-Inspired Computing*, Springer, pp. 100–116.

Breiman, L.: 1996, Bagging predictors, *Machine learning* **24**(2), 123–140.

Breiman, L.: 2001, Random forests, *Machine learning* **45**(1), 5–32.

Bunte, K., Baranowski, E., Arlt, W. and Tino, P.: n.d., Relevance learning vector quantization in variable dimensional spaces, *in* B. Hammer, T. Martinetz and T. Villmann (eds), *NC^2*, Workshop of the GI-Fachgruppe Neuronale Netze and the German Neural Networks Society in connection to GCPR 2016, Hannover, Germany, pp. 20–23.

Bunte, K., Hammer, B., Wismüller, A. and Biehl, M.: 2010, Adaptive local dissimilarity measures for discriminative dimension reduction of labeled data, *Neurocomputing* **73**(7-9), 1074–1092.

Bunte, K., Schneider, P., Hammer, B., Schleif, F., Villmann, T. and Biehl, M.: 2012, Limited rank matrix learning -discriminative dimension reduction and visualization, *Neural Networks* **26**(4), 159–173.
URL: *http://dx.doi.org/10.1016/j.neunet.2011.10.001*

Bunte, K. and Tino, P.: 2020, Geodesic average over model parameters. In preparation for publishing.
URL: *https://github.com/kbunte/geodesicLVQ_toolbox*

Carvalho, D. V., Pereira, E. M. and Cardoso, J. S.: 2019, Machine learning interpretability: A survey on methods and metrics, *Electronics* **8**(8), 832.

Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P.: 2002, Smote: synthetic minority over-sampling technique, *Journal of artificial intelligence research* **16**, 321–357.

Chechik, G., Heitz, G., Elidan, G., Abbeel, P. and Koller, D.: 2008, Max-margin classification of data with absent features, *Journal of Machine Learning Research* **9**(Jan), 1–21.

Doshi-Velez, F. and Kim, B.: 2017, Towards a rigorous science of interpretable machine learning, *arXiv preprint arXiv:1702.08608* .

Driscoll, T. A., Hale, N. and Trefethen, L. N.: 2014, Chebfun guide.
URL: *https://nl.mathworks.com/matlabcentral/fileexchange/47023-chebfun-current-version*

Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B. and Herrera, F.: 2018, *Learning from imbalanced data sets*, Springer.

Fisher, A., Rudin, C. and Dominici, F.: 2019, All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously., *Journal of Machine Learning Research* **20**(177), 1–81.

Fletcher, P. T., Lu, C., Pizer, S. M. and Joshi, S.: 2004, Principal geodesic analysis for the study of nonlinear statistics of shape, *IEEE Trans. on Medical Imaging* **23**(8), 995–1005.

García-Laencina, P. J., Sancho-Gómez, J.-L. and Figueiras-Vidal, A. R.: 2010, Pattern classification with missing data: a review, *Neural Computing and Applications* **19**(2), 263–282.
URL: *https://doi.org/10.1007/s00521-009-0295-6*

Ghosh, S., Baranowski, E., van Veen, R., de Vries, G., Biehl, M., Arlt, W., Tino, P. and Bunte, K.: 2017, Comparison of strategies to learn from imbalanced classes for computer aided diagnosis of inborn steroidogenic disorders, *Proc. of the European Symposium on Artificial Neural Networks*.

Ghosh, S., Bunte, K., Tino, P., Biehl, M., deHaan, S. and Diaconou, C.: 2021, Probabilistic learning vector quatization with kl divergence.

Ghosh, S., Tino, P. and Bunte, K.: 2020, Visualisation and knowledge discovery from interpretable models, *International Joint Conference on Neural Networks, IJCNN 2020 Glasgow, UK, July 19-24, 2020*, IEEE.

Hammer, B., Strickert, M. and Villmann, T.: 2005, On the generalization ability of grlvq networks, *Neural Processing Letters* **21**(2), 109–120.

Hammer, B. and Villmann, T.: 2002, Generalized relevance learning vector quantization, *Neural Networks* **15**(8–9), 1059 – 1068.
URL: *http://www.sciencedirect.com/science/article/pii/S0893608002000795*

Hegde, H., Shimpi, N., Panny, A., Glurich, I., Christie, P. and Acharya, A.: 2019, Mice vs ppca: Missing data imputation in healthcare, *Informatics in Medicine Unlocked* **17**, 100275.
URL: *http://www.sciencedirect.com/science/article/pii/S2352914819302783*

Holzinger, A., Biemann, C., Pattichis, C. S. and Kell, D. B.: 2017, What do we need to build explainable ai systems for the medical domain?, *arXiv preprint arXiv:1712.09923* .

Hutter, F., Kotthoff, L. and Vanschoren, J.: 2019, *Automated machine learning: methods, systems, challenges*, Springer Nature.

Janosi, A., Steinbrunn, W., Pfisterer, M. and Detrano, R.: 1988, Heart disease data set, UCI machine learning repository.
URL: *https://archive.ics.uci.edu/ml/datasets/heart+Disease*

Kahramanli, H. and Allahverdi, N.: 2008, Design of a hybrid system for the diabetes and heart diseases, *Expert systems with applications* **35**(1-2), 82–89.

Kohonen, T.: 2012, *Self-organizing maps*, Vol. 30, Springer Science & Business Media.

Kubat, M.: 2017, *An introduction to machine learning*, Springer.

Lall, U. and Sharma, A.: 1996, A nearest neighbor bootstrap for resampling hydrologic time series, *Water Resources Research* **32**(3), 679–693.

Little, R. J.: 1988, A test of missing completely at random for multivariate data with missing values, *Journal of the American statistical Association* **83**(404), 1198–1202.

Little, R. J. and Rubin, D. B.: 2019, *Statistical analysis with missing data*, Vol. 793, John Wiley & Sons.

Lundberg, S. M. and Lee, S.-I.: 2017, A unified approach to interpreting model predictions, *Advances in neural information processing systems*, pp. 4765–4774.

Marlin, B. M.: 2008, *Missing Data Problems in Machine Learning*, PhD thesis, Toronto, Ont., Canada, Canada. AAINR57898.

Minka, T. et al.: 2005, Divergence measures and message passing, *Technical report*, Technical report, Microsoft Research.

Nahar, J., Imam, T., Tickle, K. S. and Chen, Y.-P. P.: 2013, Computational intelligence for heart disease diagnosis: A medical knowledge driven approach, *Expert Systems with Applications* **40**(1), 96–104.

Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S.: 2019, Dissecting racial bias in an algorithm used to manage the health of populations, *Science* **366**(6464), 447–453.

Palaniappan, S. and Awang, R.: 2008, Intelligent heart disease prediction system using data mining techniques, *2008 IEEE/ACS international conference on computer systems and applications*, IEEE, pp. 108–115.

Parsons, S.: 2005, Introduction to machine learning by ethem alpaydin, mit press, 0-262-01211-1, 400 pp, *The Knowledge Engineering Review* **20**(4), 432–433.

Pfannschmidt, L., Göpfert, C., Neumann, U., Heider, D. and Hammer, B.: 2019, Fri-feature relevance intervals for interpretable and interactive data exploration, *2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, IEEE, pp. 1–10.

Reis, I., Baron, D. and Shahaf, S.: 2018, Probabilistic random forest: A machine learning algorithm for noisy data sets, *The Astronomical Journal* **157**(1), 16.

Ribeiro, M. T., Singh, S. and Guestrin, C.: 2016, " why should i trust you?" explaining the predictions of any classifier, *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144.

Royston, P., White, I. R. et al.: 2011, Multiple imputation by chained equations (mice): implementation in stata, *J Stat Softw* **45**(4), 1–20.

Sato, A. S. and Yamada, K.: 1996, Generalized learning vector quantization, *Advances in Neural Information Processing Systems*, Vol. 8, pp. 423–429.

Schneider, P.: 2010, *Advanced methods for prototype-based classification*, Atto Producties Europe.

Schneider, P., Biehl, M. and Hammer, B.: 2007, Relevance matrices in learning vector quantization, *in* M. Verleysen (ed.), *Proc. of the 15th European Symposium on Artificial Neural Networks (ESANN)*, d-side publishing, Bruges, Belgium, pp. 37–43.

Schneider, P., Biehl, M. and Hammer, B.: 2009, Adaptive relevance matrices in learning vector quantization, *Neural computation* **21**(12), 3532–3561.

Schneider, P., Geweniger, T., Schleif, F.-M., Biehl, M. and Villmann, T.: 2011, Multivariate class labeling in robust soft lvq., *ESANN*.

Schulz, A., Gisbrecht, A. and Hammer, B.: 2015, Using discriminative dimensionality reduction to visualize classifiers, *Neural Processing Letters* **42**(1), 27–54.

Severson, K. A., Molaro, M. C. and Braatz, R. D.: 2017, Principal component analysis of process datasets with missing values, *Processes* **5**(3), 38.

Shouman, M., Turner, T. and Stocker, R.: 2012, Applying k-nearest neighbour in diagnosing heart disease patients, *International Journal of Information and Education Technology* **2**(3), 220–223.

Storbeck, K.-H., Schiffer, L., Baranowski, E. S., Chortis, V., Prete, A., Barnard, L., Gilligan, L. C., Taylor, A. E., Idkowiak, J., Arlt, W. et al.: 2019, Steroid metabolome analysis in disorders of adrenal steroid biosynthesis and metabolism, *Endocrine Reviews* **40**(6), 1605–1625.

Tipping, M. E. and Bishop, C. M.: 1999, Mixtures of probabilistic principal component analyzers, *Neural computation* **11**(2), 443–482.

van Veen, R.: 2016, *Analysis of missing data imputation applied to heart failure data*, Masters thesis, University of Groningen.

Verma, S., Dickerson, J. and Hines, K.: 2020, Counterfactual explanations for machine learning: A review, *arXiv preprint arXiv:2010.10596* .

Villmann, A., Kaden, M., Saralajew, S. and Villmann, T.: 2018, Probabilistic learning vector quantization with cross-entropy for probabilistic class assignments in classification learning, *International Conference on Artificial Intelligence and Soft Computing*, Springer, pp. 724–735.

Wang, F., Kaushal, R. and Khullar, D.: 2020, Should health care demand interpretable artificial intelligence or accept black box medicine?

Wilson, R. C., Hancock, E. R., Pekalska, E. and Duin, R. P. W.: 2014, Spherical and hyperbolic embeddings of data, *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(11), 2255–2269.
URL: *http://dx.doi.org/10.1109/TPAMI.2014.2316836*