

REFERENCES

1. Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov, *Enriching word vectors with subword information*, 2016.
2. Michael A. A. Cox and Trevor F. Cox, *Multidimensional scaling*, pp. 315–347, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
3. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, *BERT: pre-training of deep bidirectional transformers for language understanding*, CoRR **abs/1810.04805** (2018).
4. Wlodzislaw Duch, *Filter methods*, Feature Extraction: Foundations and Applications (Studies in Fuzziness and Soft Computing) (I. Guyon, S. Gunn, M. Nikravesh, and L.A. Zadeh, eds.), Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2006, pp. 89–117.
5. Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep learning*, MIT Press, Cambridge, MA, USA, 2016.
6. I. Guyon, S. Gunn, M. Nikravesh, and L.A. Zadeh, *Feature extraction: Foundations and applications (studies in fuzziness and soft computing)*, Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2006.
7. Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik, *Gene selection for cancer classification using support vector machines*, Mach. Learn. **46** (2002), no. 1–3, 389–422.
8. Yaakov HaCohen-Kerner, Daniel Miller, and Yair Yigal, *The influence of preprocessing on text classification using a bag-of-words representation*, PLOS ONE **15** (2020), no. 5, 1–22.
9. Mark Hall and L.A. Smith, *Feature subset selection: A correlation based filter approach*, Proc. Intl. Conf. Neural Inform. Processing Intell. Inform. Syst., 1997, pp. 855–858.
10. Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The elements of statistical learning, 2nd edition*, Springer Series in Statistics, Springer New York Inc., New York, NY, USA, 2017.
11. Ting-Yao Hu, Mohammadreza Armandpour, Ashish Shrivastava, Jen-Hao Rick Chang, Hema Koppula, and Oncel Tuzel, *Synt++: Utilizing imperfect synthetic data to improve speech recognition*, ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 7682–7686.
12. George H. John, Ron Kohavi, and Karl Pfleger, *Irrelevant features and the subset selection problem*, MACHINE LEARNING: PROCEEDINGS OF THE ELEVENTH INTERNATIONAL, Morgan Kaufmann, 1994, pp. 121–129.
13. Philipp Koehn, *Europarl: A parallel corpus for statistical machine translation*, <https://www.statmt.org/europarl/>, 2012.
14. Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Mallocci, Alexander Kolesnikov, and et al., *The open images dataset v4*, International Journal of Computer Vision **128** (2020), no. 7, 1956–1981.
15. Yann LeCun, John S. Denker, and Sara A. Solla, *Optimal brain damage*, Advances in Neural Information Processing Systems 2 (D. S. Touretzky, ed.), Morgan-Kaufmann, 1990, pp. 598–605.
16. John A. Lee and Michel Verleysen, *Nonlinear dimensionality reduction*, 1st ed., Springer Publishing Company, Incorporated, 2010.
17. Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts, *Learning word vectors for sentiment analysis*, 2011.
18. Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, *Efficient estimation of word representations in vector space*, 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings (Yoshua Bengio and Yann LeCun, eds.), 2013.
19. K. Pearson, *On lines and planes of closest fit to systems of points in space*, Philosophical Magazine **2** (1901), 559–572.
20. Jeffrey Pennington, Richard Socher, and Christopher D. Manning, *Glove: Global vectors for word representation*, Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1532–1543.
21. Juha Reunanen, *Overfitting in making comparisons between variable selection methods*, J. Mach. Learn. Res. **3** (2003), 1371–1382.
22. John W. Sammon, *A nonlinear mapping for data structure analysis*, IEEE Transactions on computers **18** (1969), no. 5, 401–409.

23. Bernhard Scholkopf, Alexander Smola, and Klaus-Robert Müller, *Kernel principal component analysis*, ADVANCES IN KERNEL METHODS - SUPPORT VECTOR LEARNING, MIT Press, 1999, pp. 327–352.
24. Luciana O. Silva and Luis E. Zárate, *A brief review of the main approaches for treatment of missing data*, Intell. Data Anal. **18** (2014), no. 6, 1177–1198.
25. Petr Somol, Jana Novovicova, and Pavel Pudil, *Efficient feature subset selection and subset size optimization*, Pattern Recognition (Adam Herout, ed.), IntechOpen, Rijeka, 2010.
26. Joshua B. Tenenbaum, Vin de Silva, and John C. Langford, *A global geometric framework for nonlinear dimensionality reduction*, Science **290** (2000), no. 5500, 2319.
27. W. S. Torgerson, *Multidimensional scaling: I. theory and method*, Psychometrika **17** (1952), 401–419.
28. L.J.P. van der Maaten and G.E. Hinton, *Visualizing high-dimensional data using t-sne*, (2008).
29. Christopher K. I. Williams, *On a connection between kernel pca and metric multidimensional scaling*, Advances in Neural Information Processing Systems 13 (T. K. Leen, T. G. Dietterich, and V. Tresp, eds.), MIT Press, 2001, pp. 675–681.
30. G. Young and A. Householder, *Discussion of a set of points in terms of their mutual distances*, Psychometrika **3** (1938), 19–22.