

Table A.1-1: P-values between the performance of DRVML vs. other methods on each dataset as calculated via Welch's two-sided t-test. Bold numbers denote higher performance by the proposed method. Values calculated on results shown in Table 3.

	KNN-EUC	LMNN	RVML-classes	RVML-transport
Balance	.0147	.2117	.0259	.0424
Credit	.1340	.4947	.0341	.2474
Digits	.7983	.7751	.1789	.0477
German	.3481	.2671	.9573	.7982
Mice	.0074	.6304	.0282	.0023
Segmentation	.0007	.4423	.1460	.0012
Wisconsin	.0866	.0070	.4467	.4752

A.2 Visualization

Results of the visualization experiment are shown as Figure A.2-1.

REFERENCES

- [1] M. Aharon, M. Elad, and A. Bruckstein. 2006. *K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation*. *IEEE Transactions on Signal Processing* 54, 11 (Nov 2006), 4311–4322. <https://doi.org/10.1109/TSP.2006.881199>
- [2] C. Bao, H. Ji, Y. Quan, and Z. Shen. 2016. Dictionary Learning for Sparse Coding: Algorithms and Convergence Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 7 (July 2016), 1356–1369. <https://doi.org/10.1109/TPAMI.2015.2487966>
- [3] Amir Beck and Marc Teboulle. 2009. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM J. Img. Sci.* 2, 1 (March 2009), 183–202. <https://doi.org/10.1137/080716542>
- [4] Aurélien Bellet, Amaury Habrard, and Marc Sebban. 2013. A Survey on Metric Learning for Feature Vectors and Structured Data. *CoRR* abs/1306.6709 (2013). <http://arxiv.org/abs/1306.6709>
- [5] Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. Representation Learning: A Review and New Perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 8 (aug 2013), 1798–1828. <https://doi.org/10.1109/TPAMI.2013.50>
- [6] Kevin S. Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft. 1999. When Is “Nearest Neighbor” Meaningful?. In *Proceedings of the 7th International Conference on Database Theory (ICDT '99)*. Springer-Verlag, London, UK, UK, 217–235. <http://dl.acm.org/citation.cfm?id=645503.656271>
- [7] Jock A. Blackard and Denis J. Dean. 1999. Comparative accuracies of artificial neural networks and discriminant analysis in predicting forest cover types from cartographic variables. *Computers and Electronics in Agriculture* 24, 3 (1999), 131–151. [https://doi.org/10.1016/S0168-1699\(99\)00046-0](https://doi.org/10.1016/S0168-1699(99)00046-0)
- [8] Olivier Bousquet and André Elisseeff. 2002. Stability and Generalization. *J. Mach. Learn. Res.* 2 (Mar 2002), 499–526. <https://doi.org/10.1162/153244302760200704>
- [9] Kenneth P. Burnham and David R. Anderson. 2004. Multimodal Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research* 33, 2 (2004), 261–304. <http://dx.doi.org/10.1177/0049124104268644>
- [10] Gavin C. Cawley and Nicola L.C. Talbot. 2010. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. *J. Mach. Learn. Res.* 11 (Aug. 2010), 2079–2107. <http://dl.acm.org/citation.cfm?id=1756006.1859921>
- [11] T. Cover and P. Hart. 2006. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theor.* 13, 1 (Sept. 2006), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>
- [12] Marco Cuturi. 2013. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'13)*. Curran Associates Inc., USA, 2292–2300. <http://dl.acm.org/citation.cfm?id=2999792.2999868>
- [13] Amir Globerson and Sam T. Roweis. 2005. Metric Learning by Collapsing Classes.. In *NIPS*. 451–458. <http://dblp.uni-trier.de/db/conf/nips/nips2005.html#GlobersonR05>
- [14] Jacob Goldberger, Geoffrey E Hinton, Sam T. Roweis, and Ruslan R Salakhutdinov. 2005. Neighbourhood Components Analysis. In *Advances in Neural Information Processing Systems 17*, L. K. Saul, Y. Weiss, and L. Bottou (Eds.). MIT Press, 513–520. <http://papers.nips.cc/paper/2566-neighbourhood-components-analysis.pdf>
- [15] Clara Higuera, Kathleen J. Gardiner, and Krzysztof J. Cios. 2015. Self-Organizing Feature Maps Identify Proteins Critical to Learning in a Mouse Model of Down Syndrome. *PLOS ONE* 10, 6 (06 2015), 1–28. <https://doi.org/10.1371/journal.pone.0129126>
- [16] Elad Hoffer and Nir Ailon. 2015. Deep Metric Learning Using Triplet Network.. In *SIMBAD (Lecture Notes in Computer Science)*, Aasa Feragen, Marcello Pelillo, and Marco Loog (Eds.), Vol. 9370. Springer, 84–92. <http://dblp.uni-trier.de/db/conf/simbad/simbad2015.html#HofferA15>
- [17] Z. Jiang, Z. Lin, and L. S. Davis. 2013. Label Consistent K-SVD: Learning a Discriminative Dictionary for Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 11 (Nov 2013), 2651–2664. <https://doi.org/10.1109/TPAMI.2013.88>
- [18] Purushottam Kar and Prateek Jain. 2011. Similarity-based Learning via Data Driven Embeddings. In *Proceedings of the 24th International Conference on Neural Information Processing Systems (NIPS'11)*. Curran Associates Inc., USA, 1998–2006. <http://dl.acm.org/citation.cfm?id=2986459.2986682>
- [19] C Kaynak. 1995. *Methods of combining multiple classifiers and their applications to handwritten digit recognition*. Master's thesis. Institute of Graduate Studies in Science and Engineering, Bogazici University.
- [20] Brian Kulis. 2013. Metric Learning: A Survey. *Foundations and Trends in Machine Learning* 5, 4 (2013), 287–364. <https://doi.org/10.1561/22000000019>
- [21] M. Lichman. 2013. UCI Machine Learning Repository. (2013). <http://archive.ics.uci.edu/ml>
- [22] P. C. Mahalanobis. 1936. On the generalised distance in statistics. In *Proceedings National Institute of Science, India*, Vol. 2. 49–55. <http://ir.isical.ac.in/dspace/handle/1/1268>
- [23] Olvi L. Mangasarian, W. Nick Street, and William H. Wolberg. 1995. Breast Cancer Diagnosis and Prognosis Via Linear Programming. *Operations Research* 43, 4 (1995), 570–577. <https://doi.org/10.1287/opre.43.4.570>
- [24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [25] Michaël Perrot and Amaury Habrard. 2015. Regressive Virtual Metric Learning. In *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (Eds.). Curran Associates, Inc., 1801–1809. <http://papers.nips.cc/paper/5687-regressive-virtual-metric-learning.pdf>
- [26] J. R. Quinlan. 1987. Simplifying Decision Trees. *Int. J. Man-Mach. Stud.* 27, 3 (Sept. 1987), 221–234. [https://doi.org/10.1016/S0020-7373\(87\)80053-6](https://doi.org/10.1016/S0020-7373(87)80053-6)
- [27] Thomas R. Shultz, Denis Mareschal, and William C. Schmidt. 1994. Modeling Cognitive Development on Balance Scale Phenomena. *Machine Learning* 16, 1 (01 Jul 1994), 57–86. <https://doi.org/10.1023/A:1022630902151>
- [28] Joseph St.Amand and Jun Huan. 2017. Sparse Compositional Local Metric Learning. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17)*. ACM, New York, NY, USA, 1097–1104. <https://doi.org/10.1145/3097983.3098153>
- [29] Robert Tibshirani. 1994. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society, Series B* 58 (1994), 267–288.
- [30] Andrey N. Tikhonov and Vasily Y. Arsenin. 1977. *Solutions of ill-posed problems*. V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York. xiii+258 pages. Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics.
- [31] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9 (2008), 2579–2605. <http://www.jmlr.org/papers/v9/vandermaaten08a.html>
- [32] Cédric Villani. 2008. *Optimal Transport: Old and New* (2009 ed.). Springer. <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/3540710493>
- [33] K.Q. Weinberger and L.K. Saul. 2009. Distance metric learning for large margin nearest neighbor classification. *The Journal of Machine Learning Research* 10 (2009), 207–244.
- [34] Kilian Q Weinberger, John Blitzer, and Lawrence K. Saul. 2006. Distance Metric Learning for Large Margin Nearest Neighbor Classification. In *Advances in Neural Information Processing Systems 18*, Y. Weiss, B. Schölkopf, and J. C. Platt (Eds.). MIT Press, 1473–1480. <http://papers.nips.cc/paper/2795-distance-metric-learning-for-large-margin-nearest-neighbor-classification.pdf>
- [35] Kilian Q. Weinberger and Lawrence K. Saul. 2008. Fast Solvers and Efficient Implementations for Distance Metric Learning. In *Proceedings of the 25th International Conference on Machine Learning (ICML '08)*. ACM, New York, NY, USA, 1160–1167. <https://doi.org/10.1145/1390156.1390302>
- [36] B L Welch. 1947. The Generalization of ‘Student’s’ Problem when Several Different Population Variances are Involved. *Biometrika* 34, 1/2 (1947), 28–35. <https://doi.org/10.2307/2332510>
- [37] Huan Xu, Constantine Caramanis, and Shie Mannor. 2009. Robust Regression and Lasso. In *Advances in Neural Information Processing Systems 21*, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou (Eds.). Curran Associates, Inc., 1801–1808. <http://papers.nips.cc/paper/3596-robust-regression-and-lasso.pdf>