Table 1. Classification results on the UCI datasets (# points correctly classified / # points for testing).

data set	input space		feature space	
	NM	NN	NM	NN
pima	463/638	432/638	477/638	428/638
soybean	36/37	35/37	37/37	37/37
wine	86/118	77/118	113/118	115/118
breast	430/469	420/469	448/469	451/469
ionosphere	159/251	212/251	201/251	224/251

ing methods by means of the kernel methods and MDS techniques. In the classification scenario, we defined discriminant kernels on the joint space of input and output spaces, and presented a specific family of the discriminant kernels. This family of the discriminant kernels is attractive because the induced metrics are Euclidean and Fisher separable.

Acknowledgments

The author would like to thank Dit-Yan Yeung and James T. Kwok for fruitful discussions. Many thanks to the anonymous reviewers for their useful comments.

References

Bach, F. R., & Jordan, M. I. (2003). Learning graphical models with Mercer kernels. *Advances in Neural Information Processing Systems 15*. Cambridge, MA: MIT Press.

Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling*. New York: Springer-Verlag.

Bousquet, O., & Herrmann, D. J. L. (2003). On the complexity of learning the kernel matrix. *Advances in Neural Information Processing Systems 15*. Cambridge, MA: MIT Press.

Cox, T. F., & Ferry, G. (1993). Discriminant analysis using non-metric multidimensional scaling. *Pattern Recognition*, *26*, 145–153.

Cristianini, N., Kandola, J., Elisseeff, A., & Shawe-Taylor, J. (2002). On kernel target alignment. *Advances in Neural Information Processing Systems 14*. Cambridge, MA: MIT Press.

Gower, J. C., & Legendre, P. (1986). Metric and Euclidean properties of dissimilarities coefficients. *Journal of Classification*, *3*, 5–48.

Haussler, D. (1999). *Convolution kernels on discrete structures* (Technical Report UCSC-CRL-99-10). Department of Computer Science, University of California at Santa Cruz.

Kandola, J., Shawe-Taylor, J., & Cristianini, N. (2002a). *On the extensions of kernel alignment* (Technical Report 2002-120). NeuroCOLT.

Kandola, J., Shawe-Taylor, J., & Cristianini, N. (2002b). *Optimizing kernel alignment over combinations of kernels* (Technical Report 2002-121). NeuroCOLT.

Koontz, W. L. G., & Fukunaga, K. (1972). A nonlinear feature extraction algorithm using distance information. *IEEE Transactions on Computers*, *21*, 56–63.

Lanckriet, G. R. G., Cristianini, N., Ghaoui, L. E., Bartlett, P., & Jordan, M. I. (2002). Learning the kernel matrix with semi-definite programming. *The 19th International Conference on Machine Learning*.

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, *290*, 2323–2326.

Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels*. The MIT Press.

Tenenbaum, J. B., de Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, *290*, 2319–2323.

Vapnik, V. (1995). *The nature of statistical learning theory*. New York: Springer-Verlag.

Wahba, G. (1990). *Spline models for observational data*. Philadelphia: SIAM.

Webb, A. R. (1995). Multidimensional scaling by iterative majorization using radial basis functions. *Pattern Recognition*, *28*, 753–759.

Xing, E. P., Ng, A. Y., Jordan, M. I., & Russell, S. (2003). Distance metric learning, with application to clustering with side-information. *Advances in Neural Information Processing Systems 15*. Cambridge, MA: MIT Press.

Zhang, Z., Kwok, J. T., & Yeung, D. Y. (2003). *Parametric distance metric with label information* (Technical Report HKUST-CS03-02). Department of Computer Science, Hong Kong University of Science and Technology.