Matching Pursuit Kernel Fisher Discriminant Analysis

Tom Diethe, Zakria Hussain, David R. Hardoon, John Shawe-Taylor
Centre for Computational Statistics and Machine Learning
University College London
London WC1E 6BT, UK

Abstract

Kernel Fisher Discriminant Analysis (KFDA) allows us to carry out Fisher linear discriminant analysis in a high dimensional feature space defined implicitly by a kernel [Mika et al., 1999]. One drawback, as with most kernel methods, is that storing large kernel matrices is computationally prohibitive. Several authors have made attempts at addressing this issue by creating low-rank kernel matrices behaving similarly to the full-ranked ones, such as Smola and Schölkopf [2000] who devise a method motivated by a greedy approach called Matching Pursuit (MP) Mallat and Zhang [1993].

We derive a novel sparse version of KDFA using an approach based on MP. We call this algorithm Matching Pursuit Kernel Fisher Discriminant Analysis (MPKFDA). We provide generalisation error bounds based on the analysis of the Robust Minimax algorithm by Shawe-Taylor and Cristianini [2003] together with a sample compression bounding technique as in Littlestone and Warmuth [1986]. However, the algorithm does not form a traditional compression scheme so we use a similar idea to that of Hussain and Shawe-Taylor [2008] and bound the generalisation error in the sparsely defined subspace by amalgamating both theories mentioned above.

We present experimental results on real world datasets, which show that MPKFDA is competitive with the KFDA and the SVM on UCI datasets, and additional experiments that show that the MPKFDA on average outperforms KFDA and SVM in extremely high dimensional settings.

We believe this general approach of using Matching Pursuit can be applied to other learning algorithms, resulting in sparse greedy forms of these algorithms. It would be conceivable to apply the method to Logistic Regression.

This work is based on a paper of the same title to appear at AISTATS 2009.

References

Zakria Hussain and John Shawe-Taylor. Theory of matching pursuit. *Neural Information Processing Systems*, 2008.

Nick Littlestone and Manfred Warmuth. Relating data compression and learnability. Technical report, University of California Santa Cruz, 1986.

Stéphane Mallat and Zhifeng Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

Sebastian Mika, Gunnar Rätsch, Jason Weston, Bernhard Schölkopf, and Klaus-Robert Müller. Fisher discriminant analysis with kernels. In E. Wilson Y. H. Hu, J. Larsen and S. Douglas, editors, *Proc. NNSP'99*, pages 41–48. IEEE, 1999.

John Shawe-Taylor and Nello Cristianini. Estimating the moments of a random vector. In *Proceedings of GRETSI 2003 Conference*, volume 1, page 4752, 2003.

Alex J. Smola and Bernhard Schölkopf. Sparse greedy matrix approximation for machine learning. In *Proceedings of 17th International Conference on Machine Learning*, pages 911–918. Morgan Kaufmann, San Francisco, CA, 2000.