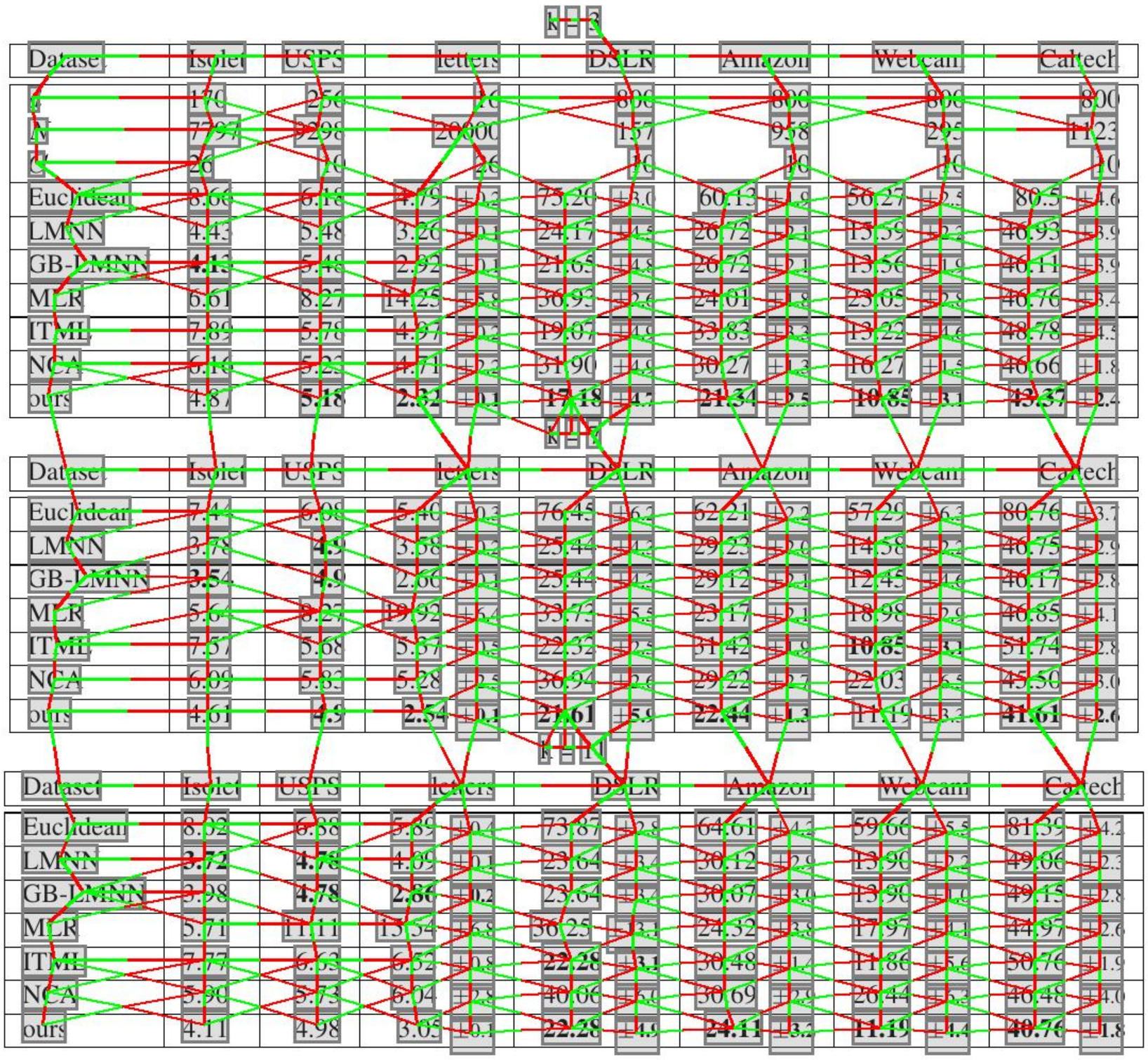

Table 1: kNN error,for k=3, 7 and 11. Features were scaled by z-scoring. Mean and standard deviation are shown for data sets on which 5-fold partition was used. Best performing methods are shown in bold. Note that the only non-linear metric learning method in the above is GB-LMNN.

than previously proposed similarity learning methods. Our learning algorithm is simple yet efficient, converging on all the data sets we have experimented upon in reasonable time as compared to the competing methods.

Our choice of Frobenius regularizer is motivated by desire to control model complexity without biasing towards a particular form of the matrix. We have experimented with alternative regularizers, both the trace norm of $\mathbf{W}$ and the shrinkage towards Euclidean distance, $\|\mathbf{W}-\mathbf{I}\|_F^2$, but found both to be inferior to $\|\mathbf{W}\|_F^2$. We suspect that often the optimal $\mathbf{W}$ corresponds to a highly anisotropic scaling of data dimensions, and thus bias towards $\mathbf{I}$ may be unhealthy.

The results in this paper are restricted to Mahalanobis metric, which is an appealing choice for a number of reasons. In particular, learning such metrics is equivalent to learning linear embedding of the data, allowing very efficient methods for metric search. Still, one can consider non-linear embeddings $\mathbf{x} \to \boldsymbol{\phi}(\mathbf{x};\mathbf{w})$ and define the distance D in terms of the embeddings, for example, as $D(\mathbf{x},\mathbf{x}_i) = \|\boldsymbol{\phi}(\mathbf{x}) - \boldsymbol{\phi}(\mathbf{x}_i)\|$ or as $-\boldsymbol{\phi}(\mathbf{x})^T\boldsymbol{\phi}(\mathbf{x}_i)$. Learning S in the latter form can be seen as learning a kernel with discriminative objective of improving kNN performance. Such a model would be more expressive, but also more challenging to optimize. We are investigating this direction.

Acknowledgments

This work was partly supported by NSF award IIS-1409837.