Indefinite Kernel Fisher Discriminant

Bernard Haasdonk
Institute of Numerical and Applied Mathematics
University of Münster, Germany
haasdonk@math.uni-muenster.de

Elżbieta Pękalska
School of Computer Science
University of Manchester, United Kingdom
pekalska@cs.man.ac.uk

Abstract

Indefinite kernels arise in practice, e.g. from problem-specific kernel construction. Therefore, it is necessary to understand the behavior and suitability of classifiers in the corresponding indefinite inner product spaces. In this paper we address the Indefinite Kernel Fisher Discriminant (IKFD). First, we give the geometric interpretation of the Fisher Discriminant in indefinite inner product spaces. We show that IKFD is closely related to the well-known formulation of the traditional Kernel Fisher Discriminant derived for positive definite kernels. Practical implications are that IKFD can be directly applied to indefinite kernels without manipulation of the kernel matrix. Experiments demonstrate the geometrically intuitive classification and enable comparisons to other indefinite kernel classifiers.

1. Introduction

Indefinite kernels arise in practice when derived from non-Euclidean distances, by incorporation of invariance into similarity measures, via robust template matching procedures, or by combination of multiple kernel matrices. However, kernel methods are mostly (and frequently wrongly) considered to be restricted to positive definite kernels. The mathematical elegance of kernel-based approaches can be extended to indefinite kernels, leading to powerful learning methods. It is, therefore, crucial to understand the geometry, behavior and suitability of kernel algorithms in the corresponding indefinite inner product spaces. This has successfully been demonstrated for, e.g. Generalized Nearest Mean Classifiers [9], Support Vector Machines [4] or regression methods [8, 1].

Kernel Fisher Discriminant (KFD) has been proposed in [6, 7] as a kernel-based formulation of the traditional Fisher Linear Discriminant [2]. In this paper, we address the Indefinite Kernel Fisher Discriminant (IKFD) as a natural extension of KFD to indefinite kernels. We will provide a geometrical interpretation, a kernelized algorithm and explain the relation to other relevant techniques in the indefinite case.

2. Data Representation in Indefinite Spaces

The so-called *pseudo-Euclidean (pE) spaces* [3] with their non-degenerate indefinite inner product are suitable for representing indefinite kernels for finite data. We briefly recall some basics of these spaces required for the further derivations; see [3] for details. A pE space is a direct sum of two Euclidean spaces $\mathbb{R}^{(p,q)} := \mathbb{R}^p \oplus \mathbb{R}^q$ for $p, q \in \mathbb{N}_0$ with the inner product $\langle \cdot,\cdot \rangle_{\mathrm{pE}} := \langle \cdot,\cdot \rangle_{\mathbb{R}^p} - \langle \cdot,\cdot \rangle_{\mathbb{R}^q}$ based on the traditional inner products of $\mathbb{R}^p$ and $\mathbb{R}^q$. The pair (p, q) is the *signature* of the pE space. For $p, q > 0$, the inner product is indefinite and can be written as a matrix-vector multiplication by $\langle \mathbf{x}, \mathbf{x}' \rangle_{\mathrm{pE}} = \mathbf{x}^T \mathbf{J} \mathbf{x}'$ with the matrix $\mathbf{J} = \mathrm{diag}(\mathbf{1}_p, -\mathbf{1}_q)$, where $\mathbf{1}_p \in \mathbb{R}^p$ denotes the p-dimensional vector of all ones. Hence, the inner product is positive definite on $\mathbb{R}^p$ and negative definite on $\mathbb{R}^q$. The indefinite inner product defines a squared norm, as usual, $\|\mathbf{x}\|^2_{\mathrm{pE}} := \langle \mathbf{x}, \mathbf{x} \rangle_{\mathrm{pE}}$, which consequently defines the squared distances $\|\mathbf{x} - \mathbf{x}'\|^2_{\mathrm{pE}}$. The space $\mathbb{R}^{p+q}$ with the usual positive definite inner product $\langle \mathbf{x}, \mathbf{x}' \rangle := \mathbf{x}^T \mathbf{x}'$ is called the *associated Euclidean space*. Linear decision functions in pE spaces are based on a normal vector $\mathbf{w} \in \mathbb{R}^{(p,q)}$ and bias $b \in \mathbb{R}$ as $f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle_{\mathrm{pE}} + b = \mathbf{w}^T \mathbf{J} \mathbf{x} + b$. These spaces face some geometrical particularities, such as vectors with negative squared norm or negative squared distances, and notions of orthogonality which differ from the usual Euclidean notion. In particular, the pE orthogonality corresponds to Euclidean orthogonality after reflection with the elementary symmetry $\mathbf{J}$, which is important for understanding linear classifiers, cf. Fig. 1. These spaces are relevant for pattern analysis as they allow embeddings of arbitrary symmetric dissimilarity data [3, 9] or data with indefinite kernels. Hence, they provide the geometry for data analysis algorithms. We

978-1-4244-2175-6/08/$25.00

assume to have the training data $X = \{x_i\}_{i=1}^n \subset \mathcal{X}$ from a general object space $\mathcal{X}$ with binary labels $\mathbf{y} = (y_i)_{i=1}^n, y_i \in \{\pm 1\}$. The number of positive and negative examples is n_+ and n_-, respectively. The set of indices of each class are $I_+ := \{i : y_i = +1\}$ and $I_- := \{i : y_i = -1\}$. A symmetric but possibly indefinite kernel function $k(x, x')$ induces the kernel matrix $\mathbf{K} = (k(x_i, x_j))_{i,j=1}^n$. By singular value decomposition of $\mathbf{K}$ one can construct an embedding $\Phi : X \to \mathbb{R}^{(p,q)}$ into a pE-space $\mathbb{R}^{(p,q)}$ such that k is the indefinite inner product $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle_{\mathrm{pE}}$. We abbreviate the matrix of mapped training data as $\mathbf{\Phi} = [\Phi(x_1), \ldots, \Phi(x_n)]$, which implies that $\mathbf{K} = \mathbf{\Phi}^T \mathbf{J} \mathbf{\Phi}$. The class means are $\boldsymbol{\mu}_+ := \frac{1}{n_+} \sum_{i \in I_+} \Phi(x_i)$ and $\boldsymbol{\mu}_- := \frac{1}{n_-} \sum_{i \in I_-} \Phi(x_i)$.

3. Indefinite KFD

Given the indefinite kernel matrix and the embedded data in a pE space, the linear Fisher Discriminant $f(x) = \langle \mathbf{w}, \Phi(x) \rangle_{\mathrm{pE}} + b$ is based on a weight vector $\mathbf{w}$ such that the between-class scatter is maximized while the within-class scatter is minimized along $\mathbf{w}$. This direction is obtained by maximizing the Fisher criterion

$$J(\mathbf{w}) := \frac{\left\langle \mathbf{w}, \Sigma_{\mathrm{pE}}^B \mathbf{w} \right\rangle_{\mathrm{pE}}}{\left\langle \mathbf{w}, \Sigma_{\mathrm{pE}}^W \mathbf{w} \right\rangle_{\mathrm{pE}}}. \tag{1}$$

Here, the operation of the pseudo-Euclidean between-class scatter matrix is given by $\Sigma_{\mathrm{pE}}^B \mathbf{w} = (\boldsymbol{\mu}_+ - \boldsymbol{\mu}_-) \left\langle \boldsymbol{\mu}_+ - \boldsymbol{\mu}_-, \mathbf{w} \right\rangle_{\mathrm{pE}} = (\boldsymbol{\mu}_+ - \boldsymbol{\mu}_-)(\boldsymbol{\mu}_+ - \boldsymbol{\mu}_-)^T \mathbf{J} \mathbf{w}$. Hence, we can write $\Sigma_{\mathrm{pE}}^B = \Sigma^B \mathbf{J}$, where $\Sigma^B = (\boldsymbol{\mu}_+ - \boldsymbol{\mu}_-)(\boldsymbol{\mu}_+ - \boldsymbol{\mu}_-)^T$ is the Euclidean between-class scatter matrix in the associated Euclidean space. Similarly, the within-class scatter matrix can be expressed as $\Sigma_{\mathrm{pE}}^W := \Sigma^W \mathbf{J}$ with the Euclidean within-class scatter matrix $\Sigma^W := \sum_{i \in I_+} (\Phi(x_i) - \boldsymbol{\mu}_+)(\Phi(x_i) - \boldsymbol{\mu}_+)^T + \sum_{i \in I_-} (\Phi(x_i) - \boldsymbol{\mu}_-)(\Phi(x_i) - \boldsymbol{\mu}_-)^T$. The bias in the classifier can be chosen as $b = -\frac{1}{2} \left\langle \mathbf{w}, \boldsymbol{\mu}_+ + \boldsymbol{\mu}_- \right\rangle_{\mathrm{pE}}$.

The first important result is a geometric interpretation of the indefinite Fisher Discriminant: inserting these representations and substituting $\mathbf{v} = \mathbf{J}\mathbf{w}$ into the Fisher criterion (1) and the discriminant function yields $J(\mathbf{w}) = \mathbf{v}^T \Sigma^B \mathbf{v} / (\mathbf{v}^T \Sigma^W \mathbf{v})$ while $f(x) = \mathbf{v}^T \Phi(x) + b$ with $b = -\frac{1}{2} \mathbf{v}^T (\boldsymbol{\mu}_+ + \boldsymbol{\mu}_-)$. This means that the Fisher Discriminant in the pE space $\mathbb{R}^{(p,q)}$ is identical to the Fisher Discriminant in the associated Euclidean space $\mathbb{R}^{p+q}$. This is by far not clear a-priori and not valid for other indefinite kernel classifiers, e.g. indefinite SVM.

A kernel method should avoid such explicit embeddings into a pE space and constructions of new inner products based on eigendecompositions. The kernel function should only be used, instead. And indeed, IKFD can exactly be obtained in such a kernelized form by using the original indefinite kernel. The normal $\mathbf{w} \in \mathbb{R}^{(p,q)}$ can be expressed as a linear combination of the training data $\Phi(x_i)$. So, we have $\mathbf{w} = \sum_{i=1}^n \alpha_i \Phi(x_i) = \mathbf{\Phi} \boldsymbol{\alpha}$ with the coefficient vector $\boldsymbol{\alpha} = (\alpha_i)_{i=1}^n \in \mathbb{R}^n$. In addition, we also introduce the coefficient vector $\mathbf{c}_+$ with the entries $(\mathbf{c}_+)_i = 1/n_+$ for $i \in I_+$ and $(\mathbf{c}_+)_i = 0$, otherwise and, similarly, $\mathbf{c}_-$ with the entries $(\mathbf{c}_-)_i = 1/n_-$ for $i \in I_-$ and $(\mathbf{c}_-)_i = 0$, otherwise. This allows us to compactly write $\boldsymbol{\mu}_+ - \boldsymbol{\mu}_- = \mathbf{\Phi}(\mathbf{c}_+ - \mathbf{c}_-)$. By using $\mathbf{C} := (\mathbf{c}_+ - \mathbf{c}_-)(\mathbf{c}_+ - \mathbf{c}_-)^T$, as a result, the numerator of (1) can be rewritten as

$$\begin{aligned}\left\langle \mathbf{w}, \Sigma_{\mathrm{pE}}^B \mathbf{w} \right\rangle_{\mathrm{pE}} &= (\mathbf{\Phi}\boldsymbol{\alpha})^T \mathbf{J} (\boldsymbol{\mu}_+ - \boldsymbol{\mu}_-)(\boldsymbol{\mu}_+ - \boldsymbol{\mu}_-)^T \mathbf{J} \mathbf{\Phi} \boldsymbol{\alpha} \\ &= \boldsymbol{\alpha}^T \mathbf{\Phi}^T \mathbf{J} \mathbf{\Phi} (\mathbf{c}_+ - \mathbf{c}_-)(\mathbf{c}_+ - \mathbf{c}_-)^T \mathbf{\Phi}^T \mathbf{J} \mathbf{\Phi} \boldsymbol{\alpha} \\ &= \boldsymbol{\alpha}^T \mathbf{\Phi}^T \mathbf{J} \mathbf{\Phi} \mathbf{C} \mathbf{\Phi}^T \mathbf{J} \mathbf{\Phi} \boldsymbol{\alpha} = \boldsymbol{\alpha}^T \mathbf{K} \mathbf{C} \mathbf{K} \boldsymbol{\alpha}\end{aligned}$$

thanks to $\mathbf{K} = \mathbf{\Phi}^T \mathbf{J} \mathbf{\Phi}$ being the given indefinite kernel matrix. Concerning the denominator, we decompose $\mathbf{\Phi}$ into images of the positive class $\mathbf{\Phi}_+ \in \mathbb{R}^{n \times n_+}$ and images of the negative class $\mathbf{\Phi}_- \in \mathbb{R}^{n \times n_-}$. Similarly, the kernel matrix is decomposed into $\mathbf{K}_+ = \mathbf{\Phi}^T \mathbf{J} \mathbf{\Phi}_+ \in \mathbb{R}^{n \times n_+}$ and $\mathbf{K}_- = \mathbf{\Phi}^T \mathbf{J} \mathbf{\Phi}_- \in \mathbb{R}^{n \times n_-}$. Using these definitions and some elementary reformulations, the within-class scatter matrix for the positive class is

$$\sum_{i \in I_+} (\Phi(x_i) - \boldsymbol{\mu}_+)(\Phi(x_i) - \boldsymbol{\mu}_+)^T = \mathbf{\Phi}_+ (\mathbf{I}_{n_+} - \frac{1}{n_+} \mathbf{1}_{n_+} \mathbf{1}_{n_+}^T) \mathbf{\Phi}_+^T = \mathbf{\Phi}_+ \mathbf{H}_+ \mathbf{\Phi}_+^T$$

with the centering matrix $\mathbf{H}_+ = \mathbf{I}_{n_+} - \frac{1}{n_+} \mathbf{1}_{n_+} \mathbf{1}_{n_+}^T$, and identity matrix $\mathbf{I}_{n_+} \in \mathbb{R}^{n_+ \times n_+}$. A similar derivation with an analogous definition of $\mathbf{H}_-$ can be done for the within-class scatter matrix of the other class. Consequently, the denominator of (1) becomes

$$\begin{aligned}&\left\langle \mathbf{w}, \Sigma_{\mathrm{pE}}^W \mathbf{w} \right\rangle_{\mathrm{pE}} \\ &= \boldsymbol{\alpha}^T \mathbf{\Phi}^T \mathbf{J} \left(\mathbf{\Phi}_+ \mathbf{H}_+ \mathbf{\Phi}_+^T + \mathbf{\Phi}_- \mathbf{H}_- \mathbf{\Phi}_-^T \right) \mathbf{J} \mathbf{\Phi} \boldsymbol{\alpha} \\ &= \boldsymbol{\alpha}^T \left(\mathbf{K}_+ \mathbf{H}_+ \mathbf{K}_+^T + \mathbf{K}_- \mathbf{H}_- \mathbf{K}_-^T \right) \boldsymbol{\alpha}.\end{aligned}$$

The objective function to be maximized for IKFD is now identical to the original KFD criterion (valid for the positive definite case), but involves an indefinite kernel matrix: $J(\boldsymbol{\alpha}) = \frac{\boldsymbol{\alpha}^T \mathbf{M} \boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \mathbf{N} \boldsymbol{\alpha}}$ with the matrices $\mathbf{M} := \mathbf{K}\mathbf{C}\mathbf{K}$ and $\mathbf{N} := \mathbf{K}_+ \mathbf{H}_+ \mathbf{K}_+^T + \mathbf{K}_- \mathbf{H}_- \mathbf{K}_-^T$. Note that $\mathbf{M}$ and $\mathbf{N}$ are positive semidefinite by construction. Since $\mathbf{N}$ is in general singular, a slight regularization is used, e.g. $\mathbf{N}_\beta := \mathbf{N} + \beta \mathbf{I}_n$ with a parameter $\beta > 0$. We know that the solution $\boldsymbol{\alpha}$ can be obtained as the leading eigenvector of $\mathbf{N}_\beta^{-1} \mathbf{M}$. A simplified solution is based on the insight that $\mathrm{rank}(\mathbf{M}) = 1$, mapping vectors to the span of $\mathbf{K}(\mathbf{c}_+ - \mathbf{c}_-)$. Hence, one

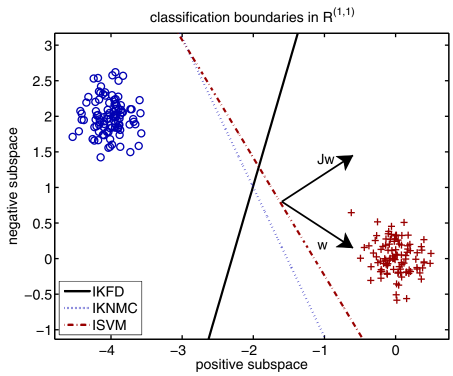

Figure 1. Linear separation based on an indefinite linear kernel.

can directly obtain the solution $\boldsymbol{\alpha} = \mathbf{N}_{\beta}^{-1}\mathbf{K}(\mathbf{c}_{+} - \mathbf{c}_{-})$. After training of the IKFD in the kernel form, classification also naturally relies on taking the sign of the kernel expansion $f(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x) + b$, where $b = -\frac{1}{2}\boldsymbol{\alpha}^T(\frac{1}{n_+}\mathbf{K}_+\mathbf{1}_{n_+} + \frac{1}{n_-}\mathbf{K}_-\mathbf{1}_{n_-})$. Starting with the original KFD formulation for positive definite kernels one could presume that the KFD might be heuristically applicable for indefinite kernels. Our derivation shows that this is based on a theoretic basis and an intuitive geometrical interpretation. Hence, IKFD is a sound method for indefinite kernels.

4. Experiments

All experiments are performed by using the MATLAB-toolbox PRtools (`http://prtools.org`). We use two reference kernel classifiers: indefinite SVM (ISVM) and indefinite kernel nearest mean classifier (IKNMC), based on the decision function $f(x) := \left\|\boldsymbol{\mu}_- - \Phi(x)\right\|^2_{\mathrm{pE}} - \left\|\boldsymbol{\mu}_+ - \Phi(x)\right\|^2_{\mathrm{pE}}$ and becomes a linear discriminant with $\mathbf{w} := \boldsymbol{\mu}_+ - \boldsymbol{\mu}_-$ and b as for IKFD.

The first example is an artificial indefinite data set consisting of two normally distributed point clouds in $\mathbb{R}^{(1,1)}$ plotted in Fig. 1. The horizontal axis is the subspace with positive signature, the vertical axis the space with negative signature. The pE inner product is taken as a kernel, which results in linear classification boundaries for all considered classifiers.

The intuitively good, and "Euclidean-like" classification of the indefinite data set by IKFD is obvious. This is a confirmation of the geometrical interpretation from Sec. 3. Other classifiers such as IKNMC and ISVM suffer from the implicit reflection of the decision normal $\mathbf{w}$ by the multiplication with $\mathbf{J}$ in the indefinite inner product. Note, that in the given example, both ISVM and IKFD will converge towards the IKNMC boundary if the regularization parameters $C \to 0$ and $\beta \to \infty$, respectively.

As a further example we consider an artificial 4×4 checkerboard data set based on a uniform distribution on $[-2, 2]^2 \subset \mathbb{R}^2$, cf. Fig. 2. A practical source of indefiniteness is incorporation of invariance into kernels. Here, it is done by combining different kernels into a new one. Let us denote $d(x, x') := \sum_{i=1,2} |(x)_i - (x')_i|^2$ and the kernel $k(x, x') := \exp(-d(x, x')^2/\sigma^2)$. As prior knowledge we observe that the problem is invariant wrt. the point reflection $\tau(x) := -x$ through the origin. We incorporate this prior knowledge by combining two kernels into a new one: $\bar{k}(x, x') := \max(k(x, x'), k(x, \tau(x')))$, which can alternatively be motivated by invariant distances [5]. Application on a random training data set of $50 + 50$ samples yields an indefinite kernel matrix and corresponding data representations in pE spaces $\mathbb{R}^{(p,q)}$ for each σ. In addition to the signature (p, q), Tab. 1 gives further quantifications of the indefiniteness by $r := \left(\sum_{\lambda_i < 0} |\lambda_i|\right) / \left(\sum_i |\lambda_i|\right)$, the ratio of negative variance to overall variance measured by the sums of eigenvalues λ_i of $\mathbf{K}$. A 10-fold cross-validation was performed on the training set for each of the listed σ to determine the regularization parameter for IKFD among $\beta = 10^{-4}, 10^{-3}, \ldots, 10^2$ and for ISVM among $C = 10^{-2}, 10^{-1}, \ldots, 10^4$. The chosen regularization parameters and the corresponding test errors ($500 + 500$ samples) are reported in Tab. 1. The resulting classifiers (with σ also being selected in cross-validation) are illustrated in Fig. 2. The parameters are: $\beta = 1, \sigma = 0.5$ for IKFD, $C = 1, \sigma = 0.1$ for ISVM, and

Table 1. Checkerboard example. Measures of indefiniteness and test errors.

σ	$r\ (p,q)$	IKFD (β)	ISVM (C)	IKNMC
0.010	0.000 (98,2)	0.336 (10)	0.323 (10)	0.340
0.050	0.022 (82,18)	0.145 (10)	0.134 (10)	0.173
0.100	0.055 (66,34)	0.121 (10^{-1})	0.121 (1)	0.201
0.500	0.125 (51,49)	0.083 (1)	0.168 (1)	0.384
1.000	0.132 (52,48)	0.091 (10^{-3})	0.418 (1)	0.486
5.000	0.107 (50,50)	0.132 (10^{-2})	0.480 (1)	0.497
10.00	0.062 (51,49)	0.159 (10^{-3})	0.373 (10^2)	0.494

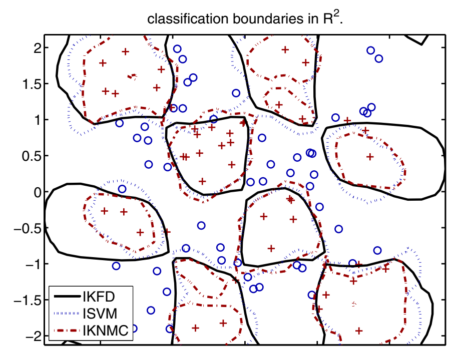

Figure 2. Indefinite invariant-kernel classifiers for the checkerboard data.

Table 2. Polygon example. Average signatures and test-errors.

γ	mean (p, q)	IKFD	ISVM	IKNMC
0.2	(99.0,1.0)	0.021±0.006	0.021±0.006	0.089±0.027
0.5	(99.0,1.0)	0.019±0.006	0.018±0.004	0.110±0.034
0.7	(98.9,1.1)	0.020±0.004	0.018±0.004	0.118±0.037
1.0	(85.9,14.1)	0.019±0.009	0.029±0.007	0.129±0.041
2.0	(48.9,51.1)	0.017±0.008	0.094±0.057	0.152±0.051
5.0	(44.8,55.2)	0.102±0.021	0.131±0.030	0.218±0.081
7.0	(47.4,52.6)	0.111±0.027	0.237±0.058	0.253±0.093

$\sigma = 0.05$ for IKNMC. The test-errors equal $0.083, 0.121$ and 0.173, respectively. The perfect point symmetry of all classifiers in Fig. 2 occurs thanks to the invariant kernel. The table shows that IKNMC is here consistently worse than IKFD. ISVM performs as well or better than IKFD for marginal indefiniteness (here $\sigma \leq 0.1$). For predominantly indefinite data (here $\sigma > 0.1$), IKFD outperforms ISVM.

The last two examples are based on non-Euclidean dissimilarities, a common cause of indefiniteness. We consider a data set of $2000+2000$ polygons corresponding to two classes of polygons with five and seven vertices, respectively. The modified Hausdorff-distance is applied for computing the pairwise distances, see [10]. We convert this dissimilarity d into similarity by considering the kernel $k(x, x') := -d(x, x')^\gamma$ for $\gamma > 0$. The experiment setting is as before, taking $50 + 50$ samples for a 10-fold cross-validated parameter selection and training, and then testing on the remaining 3900 examples. In order to address the statistical significance, we repeat this 10 times. The resulting mean and standard deviations of the test-errors, as well as of the signatures are reported in Tab. 2 We see that IKFD and ISVM clearly outperform IKNMC. In the dominant positive definite case, $\gamma \leq 0.7$, there is no significant difference between the performance of IKFD and ISVM, while in the remaining indefinite cases IKFD is obviously beneficial, with the overall best result for $\gamma = 2$. The final 5-class data set, called *chicken* data [11], is given as a dissimilarity matrix between 446 objects based on an asymmetric measure d. This give the symmetric but indefinite kernel $k(x, x') = -((d(x, x') + d(x', x))/2)^2$. The cross-validation, training and test-error determination is based on a 75%/25% train/test partition. As average test errors over 20 repetitions we obtain 0.0785 ± 0.034 (IKFD), 0.1029 ± 0.0273 (ISVM) and 0.2966 ± 0.0496 (IKNMC). This again confirms the superiority of IKFD over ISVM in the indefinite cases.

5. Conclusions

We have derived the Fisher Discriminant in a pE space and shown its equivalence to the Fisher Discriminant in the associated Euclidean space. This is an appealing nontrivial property which is invalid for several other indefinite classifiers, such as ISVM or IKNMC. Consequently, we obtain more intuitive classification boundaries in pE spaces than in case of other indefinite classifiers. Further on, Indefinite Kernel Fisher Discriminant (IKFD) is proposed as an extension of the traditional KFD. In agreement with the derivation in a pE space, this leads to the usual KFD algorithm, but now based on an *indefinite* kernel matrix. Consequently, IKFD has a sound theoretical basis and can be applied to indefinite kernels without manipulation of the kernel matrix. We have demonstrated the classification behavior of IKFD in indefinite spaces and concluded superiority over ISVM and IKNMC in case of essential indefiniteness. This is of practical importance for learning from non-Euclidean dissimilarities, kernel combinations or incorporation of prior-knowledge, e.g. invariances, into kernels.

References

[1] S. Canu, C. Ong, and X. Mary. Splines with non positive kernels. In *International ISAAC Congress*. World Scientific, 2005.

[2] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. John Wiley & Sons, Inc., 2nd edition, 2001.

[3] L. Goldfarb. A new approach to pattern recognition. In L. Kanal and A. Rosenfeld, editors, *Progress in Pattern Recognition*, volume 2, pages 241–402. Elsevier Science Publishers BV, 1985.

[4] B. Haasdonk. Feature space interpretation of SVMs with indefinite kernels. *IEEE TPAMI*, 27(4):482–492, 2005.

[5] B. Haasdonk and H. Burkhardt. Invariant kernel functions for pattern analysis and machine learning. *Machine Learning*, 68(1):35–61, 2007.

[6] S. Mika, G. Rätsch, B. Schölkopf, A. Smola, J. Weston, and K.-R. Müller. Invariant feature extraction and classification in kernel spaces. In *NIPS 12*. MIT Press, 2000.

[7] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R. Müller. Fisher discriminant analysis with kernels. In *NNSP*, pages 41–48, 1999.

[8] C. S. Ong, X. Mary, S. Canu, and A. J. Smola. Learning with non-positive kernels. In *ICML*, pages 639–646. ACM Press, 2004.

[9] E. Pękalska and R. Duin. *The Dissimilarity Representation for Pattern Recognition. Foundations and Applications*. World Scientific, Singapore, 2005.

[10] E. Pękalska, R. Duin, and P. Paclík. Prototype selection for dissimilarity-based classifiers. *Pattern Recognition*, 39(2):189–208, 2006.

[11] E. Pękalska, A. Harol, R. Duin, B. Spillmann, and H. Bunke. Non-euclidean or non-metric measures can be informative. In *S+SSPR 2006*, 2006.