JMLR: Workshop and Conference Proceedings 11 (2010) 4–11 Workshop on Applications of Pattern Analysis

O-IPCAC and its Application to EEG Classification

Alessandro Rozza ROZZA@DSI.UNIMI.IT
Gabriele Lombardi LOMBARDI@DSI.UNIMI.IT
Marco Rosa ROSA@DSI.UNIMI.IT
Elena Casiraghi CASIRAGH@DSI.UNIMI.IT
Università degli Studi di Milano, Dipartimento di Scienze dell'Informazione

Editors: Tom Diethe, Nello Cristianini, John Shawe-Taylor

Abstract

In this paper we describe an online/incremental linear binary classifier based on an interesting approach to estimate the Fisher subspace. The proposed method allows to deal with datasets having high cardinality, being dynamically supplied, and it efficiently copes with high dimensional data without employing any dimensionality reduction technique. Moreover, this approach obtains promising classification performance even when the cardinality of the training set is comparable to the data dimensionality.

We demonstrate the efficacy of our algorithm by testing it on EEG data. This classification problem is particularly hard since the data are high dimensional, the cardinality of the data is lower than the space dimensionality, and the classes are strongly unbalanced. The promising results obtained in the MLSP competition, without employing any feature extraction/selection step, have demonstrated that our method is effective; this is further proved both by our tests and by the comparison with other well-known classifiers.

Keywords: Fisher subspace, Online learning, EEG classification.

1. Introduction

Given a set of training vectors $\mathcal{P} = \bigcup_{c=1}^{C} \mathcal{P}_c = \{\boldsymbol{p}_i\}_{i=1}^{N}$ sampled from a set of linearly separable classes, where c is the label of the class and $\boldsymbol{p}_i \in \Re^D$, Fisher Linear Discriminant Analysis (`FLDA`) is often applied, due to its simplicity and efficacy, to project the data on the $c-1$ dimensional linear subspace where the data separability is maximal.

In (Rozza et al., 2009) the authors exploit the appealing properties of `FLDA` by proposing a binary classifier, called Isotropic `PCA` Classifier (`IPCAC`), that exploits theoretical results presented in (Brubaker and Vempala, 2008) to efficiently estimate the Fisher Subspace (`FS`). More precisely, Brubaker and Vempala demonstrate that, given a set of clustered points sampled from an isotropic Mixture of Gaussians (`MoG`), `FS` corresponds to the span of the class means; as a consequence, when a binary classification problem is considered, `FS` is spanned by $\boldsymbol{f} = \frac{\boldsymbol{\mu}_A - \boldsymbol{\mu}_B}{||\boldsymbol{\mu}_A - \boldsymbol{\mu}_B||}$, being A/B the two classes, and $\boldsymbol{\mu}_{A/B}$ the class means.

`IPCAC` exploits this result by firstly whitening the training set $\mathcal{P}$, computing the unit vector $\boldsymbol{f}$, and then classifying a new point $\boldsymbol{p}$ by thresholding its projection on `FS` as follows:

$$(\boldsymbol{W}_N^T \boldsymbol{f}) \cdot \boldsymbol{p} - \gamma = \boldsymbol{w} \cdot \boldsymbol{p} - \gamma < 0\,; \qquad \gamma = \left\langle \arg \max_{\{\bar{\gamma}\} \subseteq \{\boldsymbol{w} \cdot \boldsymbol{p}_i\}} Score(\bar{\gamma}) \right\rangle \tag{1}$$

where the matrix $\boldsymbol{W}_N$ represents the whitening transformation estimated on the N training points, $Score(\bar{\gamma})$ computes the number of correctly classified training points when $\bar{\gamma}$ is used as threshold, and $\langle\cdot\rangle$ represents the average operator.

Despite the effectiveness of IPCAC and its improvements (Rozza et al., 2010), when the dataset is high dimensional the computation of $\boldsymbol{W}_N$ is impractical. Moreover, as other techniques based on the estimation of FS, IPCAC fails when the training-set cardinality is equal or lower than the space dimensionality.

To address these problems, in this work we propose an improvement of IPCAC, that will be referred as O-IPCAC (Online IPCAC). In O-IPCAC two main improvements have been defined: firstly, the data whitening has been replaced by a process that whitens the data in a linear subspace $\pi_d = \textbf{Span}\langle \boldsymbol{v}_1, \ldots, \boldsymbol{v}_d \rangle, d \ll D$, while maintaining unaltered the information related to the orthogonal subspace $(\pi_d)^{\perp} = \textbf{Span}\langle \boldsymbol{v}_{d+1}, \ldots, \boldsymbol{v}_D \rangle$; secondly, the classification algorithm has been designed to perform online/incremental training.

We demonstrate the efficacy of our algorithm by employing it on EEG classification. Recently this problem is raising a wide interest since it is the fundamental step of Brain to Computer Interface (BCI) systems, which are based on the translation of brain activity into commands for the computer. The task of EEG classification is a hard problem, since the two classes are often highly unbalanced, the selection of discriminative information is difficult, the data are high dimensional, and the cardinality of the training sets are often lower than their dimensionality (for a survey see Lotte et al. (2007)). To deal with these problems, feature extraction/selection techniques are generally used to compute a small number of features representing the data; unfortunately, this approach causes loss of discriminative information, and might affect the classification accuracy. Note that, while dimensionality reduction is exploited by several EEG classification systems (Wei et al., 2007), so that their performance mainly depends on the quality of the used features, O-IPCAC can be applied on the raw data since it has been developed to deal with high dimensional datasets whose cardinality is lower than the space dimensionality.

The EEG dataset employed in this work has been distributed for the MLSP 2010 competition (Hild et al., Sept. 2010); it consists in a multi-channel time-series containing measures of brain electrical activity recorded while a subject viewed satellite images. The classifier must analyze the brain activity to recognize those images containing a predefined target.

The promising results achieved in the MLSP competition, and the comparison with state-of-the-art methods, prove the efficacy of our approach.

2. Online IPCAC

Given the matrix $\boldsymbol{P} \in \Re^{D \times N}$, representing a training dataset $\mathcal{P} = \mathcal{P}_A \cup \mathcal{P}_B, |\mathcal{P}| = N = N_A + N_B$, let α be the ratio D/N; when $\alpha \approx 1$ the performance of IPCAC deteriorates dramatically since the sample covariance matrix $S_N = \frac{1}{N-1}\boldsymbol{P}\boldsymbol{P}^T$ is not a consistent estimator of the population covariance matrix Σ (see Johnstone and Lu (2004)). More precisely, assuming that $\Sigma = \Sigma^* + \sigma^2 I$, where Σ^* has rank $k < D$ and $\sigma^2 I$ represents the contribution of a zero mean Gaussian noise affecting the data, calling $\sigma^2 = \lambda_1 = \ldots = \lambda_{D-k-1} < \ldots < \lambda_D$ the ordered eigenvalues of Σ, and denoting with $l_1 < \ldots < l_D$ the ordered eigenvalues of S_N, in (Paul, 2007) it is proved that only the portion of the spectrum of Σ above $\sigma^2 + \sqrt{\alpha}$ can be correctly estimated from the sample. Furthermore, when $\alpha \approx 1$ the estimates of the smallest eigenvalues l_i can be much larger than the real ones, and the corresponding estimated eigenvectors are uncorrelated with the real ones. These results motivate our choice of improving IPCAC by considering only the largest eigenvalues; this is obtained by substituting the whitening step by a "partial" whitening with respect to the first $d \leq D$ principal components of $\boldsymbol{P}$, where d is a parameter to be set[1].

To estimate the linear transformation $\boldsymbol{W}$, which represents the partial whitening operator, we apply the Truncated Singular Value Decomposition(TSVD, Hansen (1986)), obtaining the low-rank factorization $\boldsymbol{P} \simeq \boldsymbol{U}_d \boldsymbol{Q}_d \boldsymbol{V}_d^T$. The d largest singular values on the diagonal of $\boldsymbol{Q}_d$, and the associated left singular vectors, are employed to project the points in $\boldsymbol{P}$ on the subspace $\mathcal{SP}_d$ spanned by the columns of $\boldsymbol{U}_d$, and to perform the whitening, as follows:

$$\bar{\boldsymbol{P}}_{\boldsymbol{W}_d} = q_d \boldsymbol{Q}_d^{-1} \boldsymbol{P}_{\perp \mathcal{SP}_d} = q_d \boldsymbol{Q}_d^{-1} \boldsymbol{U}_d^T \boldsymbol{P} = \boldsymbol{W}_d \boldsymbol{P} \tag{2}$$

where q_d is the smallest singular value of the points projected in $\mathcal{SP}_d$. Note that, to obtain points whose covariance matrix best resembles a multiple of the identity, we have chosen to set the value of the d largest singular values to q_d instead of 1, thus avoiding the gap between the d-th and the $(d+1)$-th singular value.

The obtained matrix $\boldsymbol{W}_d$ projects and whitens the points in the linear subspace $\mathcal{SP}_d$; however, dimensionality reduction might delete discriminative information, decreasing the classification performance. As an example, consider two classes with the shape of two parallel pancakes in $\Re^D$: if the direction defined by the two class means in the original space $(\boldsymbol{\mu}_A - \boldsymbol{\mu}_B)$ is orthogonal to the subspace π_d defined by the first $d \leq D$ principal components, the dimensionality reduction process projects the data on π_d, obtaining an isotropic mixture of two completely overlapped Gaussian distributions.

To avoid this information loss, we add to the partially whitened data the residuals ($\boldsymbol{R}$) of the points in $\boldsymbol{P}$ with respect to their projections on $\mathcal{SP}_d$:

$$\boldsymbol{R} = \boldsymbol{P} - \boldsymbol{U}_d \boldsymbol{P}_{\perp \mathcal{SP}_d} = \boldsymbol{P} - \boldsymbol{U}_d \boldsymbol{U}_d^T \boldsymbol{P}$$
$$\bar{\boldsymbol{P}}_{\boldsymbol{W}_D} = \boldsymbol{U}_d \bar{\boldsymbol{P}}_{\boldsymbol{W}_d} + \boldsymbol{R} = \boldsymbol{U}_d \boldsymbol{W}_d \boldsymbol{P} + \boldsymbol{P} - \boldsymbol{U}_d \boldsymbol{U}_d^T \boldsymbol{P} = \left(q_d \boldsymbol{U}_d \boldsymbol{Q}_d^{-1} \boldsymbol{U}_d^T + \boldsymbol{I} - \boldsymbol{U}_d \boldsymbol{U}_d^T \right) \boldsymbol{P} = \boldsymbol{W} \boldsymbol{P}$$

where $\boldsymbol{W} \in \Re^{D \times D}$ represents the linear transformation that whitens the data along the first d principal components, while keeping unaltered the information along the remaining components. FS is estimated by exploiting the whitened class means, $\boldsymbol{\mu}_A$ and $\boldsymbol{\mu}_B$, obtained by the class means in the original space $\hat{\boldsymbol{\mu}}_A$ and $\hat{\boldsymbol{\mu}}_B$ as follows:

$$\boldsymbol{\mu}_A = \boldsymbol{W} \hat{\boldsymbol{\mu}}_A = \left(q_d \boldsymbol{U}_d \boldsymbol{Q}_d^{-1} \boldsymbol{U}_d^T + \boldsymbol{I} - \boldsymbol{U}_d \boldsymbol{U}_d^T \right) \hat{\boldsymbol{\mu}}_A = q_d \boldsymbol{U}_d \boldsymbol{Q}_d^{-1} \boldsymbol{U}_d^T \hat{\boldsymbol{\mu}}_A + \hat{\boldsymbol{\mu}}_A - \boldsymbol{U}_d \boldsymbol{U}_d^T \hat{\boldsymbol{\mu}}_A \tag{3}$$

The same calculation is done for $\boldsymbol{\mu}_B$. Using these quantities we estimate $\boldsymbol{f} = \frac{\boldsymbol{\mu}_A - \boldsymbol{\mu}_B}{\|\boldsymbol{\mu}_A - \boldsymbol{\mu}_B\|}$. Then, we process an unknown point $\boldsymbol{p}$ by transforming it with $\boldsymbol{W}$, and projecting it on $\boldsymbol{f}$; both these steps are performed by the inner product $\boldsymbol{w} \cdot \boldsymbol{p}$, where:

$$\boldsymbol{w} = \boldsymbol{W}^T \boldsymbol{f} = q_d \boldsymbol{U}_d^T \boldsymbol{Q}_d^{-1} \boldsymbol{U}_d \boldsymbol{f} + \boldsymbol{f} - \boldsymbol{U}_d^T \boldsymbol{U}_d \boldsymbol{f} \tag{4}$$

Finally, given γ as in Equation (1), $\boldsymbol{p}$ is assigned to class A if $\boldsymbol{w} \cdot \boldsymbol{p} < \gamma$, to class B otherwise.

1. We empirically chose $d = \min(\log_2^2 N, D)$ since we noticed that in our tests, on both synthetic and real data, the generalization capability of O-IPCAC remains approximately maximal by employing this value.

Notice that we never explicitly compute the matrix $\boldsymbol{W}$, but we perform the matrix times vector operations reported in Equations (3) and (4), thus preventing a quadratic time/space complexity. After the training phase, the classification model is represented by $\boldsymbol{w}$ and γ.

With training sets of high cardinality, or when mini-batches of training data $\mathcal{B}_k = \mathcal{B}_{A,k} \cup \mathcal{B}_{B,k}$ are dynamically supplied, subsequent training phases must be applied to update the classification model. To this aim, the algorithm has been extended to perform **online/incremental** training by updating the following parameters:

$N_k, N_{A,k}, N_{B,k}$: number of training points seen until the k-th training phase;

$\boldsymbol{\mu}_k, \hat{\boldsymbol{\mu}}_{A,k}, \hat{\boldsymbol{\mu}}_{B,k}$: the means employed to obtain the centered sets $\mathcal{P}_k$, $\mathcal{P}_{A,k}$, and $\mathcal{P}_{B,k}$ respectively;

$\boldsymbol{U}_{d_k}, \boldsymbol{Q}_{d_k}, \boldsymbol{V}_{d_k}$: the `SVD` matrices related to $\mathcal{P}_k$, truncated to d_k principal components;

σ_A, σ_B : the standard deviations of the projections $\boldsymbol{w}_k^T \boldsymbol{P}_{A,k}$ and $\boldsymbol{w}_k^T \boldsymbol{P}_{B,k}$.

Consider the case when k-1 training phases have been performed and a new mini-batch $\mathcal{B}_k$ is provided, where $|\mathcal{B}_k| = n_k = n_{A,k} + n_{B,k}$. Firstly, we update $N_k, N_{A,k}, N_{B,k}$, $\boldsymbol{\mu}_k, \hat{\boldsymbol{\mu}}_{A,k}, \hat{\boldsymbol{\mu}}_{B,k}$, and we center the points in $\mathcal{B}_k$ around the old mean: $\bar{\mathcal{B}}_k = \{\boldsymbol{p} - \boldsymbol{\mu}_{k-1} | \boldsymbol{p} \in \mathcal{B}_k\}$. Secondly, we update the `TSVD` matrices by means of the algorithm described in (Brand, 2006) and by exploiting the information carried by $\bar{\mathcal{B}}_k$, so that: $\bar{\boldsymbol{P}}_k \simeq \boldsymbol{U'}_{d_k} \boldsymbol{Q'}_{d_k} \boldsymbol{V'}^T_{d_k}$, where $\bar{\boldsymbol{P}}_k$ represents the set $\bar{\mathcal{P}}_k = \mathcal{P}_{k-1} \cup \bar{\mathcal{B}}_k$. Finally, considering that the updated `TSVD` matrices are related to the points $\bar{\mathcal{P}}_k$ that are centered on $\boldsymbol{\mu}_{k-1}$, a re-centering operation is required to obtain `TSVD` matrices related to points centered on $\boldsymbol{\mu}_k$; this is done by applying the re-centering rank-one modification described in (Brand, 2006), and choosing as translation the quantity $\frac{n}{N} \langle \bar{\mathcal{B}}_k \rangle$. Given the updated means and `TSVD` matrices, we can: estimate the whitened means $\boldsymbol{\mu}_{A,k}$ and $\boldsymbol{\mu}_{B,k}$ by employing Equation (3); obtain the updated vector $\boldsymbol{f}_k$; compute the the new vector $\boldsymbol{w}_k$ through Equation (4). Notice that these computations require to store only $O(Dn_k + Dd_k)$ real values per mini-batch.

Regarding the update of the thresholding value γ_k we have not employed Equation (1), since it requires to store the whole training set, and it is not able to handle unbalanced classes; therefore, we have chosen to set the value of γ_k so that it corresponds to the point having the same Mahalanobis distance $|\xi|$ from the projections of the mean vectors on the `FS`: $\bar{\mu}_{A,k} = \boldsymbol{f}_k \cdot \boldsymbol{\mu}_{A,k}$ and $\bar{\mu}_{B,k} = \boldsymbol{f}_k \cdot \boldsymbol{\mu}_{B,k}$. More precisely, we impose that $\bar{\mu}_{A,k} + \xi\ \sigma_{A,k} = \bar{\mu}_{B,k} - \xi\ \sigma_{B,k}$, where $\sigma_{A,k}$ and $\sigma_{B,k}$ are the standard deviations of the projections of the whitened points on $\boldsymbol{f}_k$. Defining $\gamma_k = \bar{\mu}_{A,k} + \xi\ \sigma_{A,k}$ we obtain $\gamma_k = \bar{\mu}_{A,k} + \sigma_{A,k} \left(\bar{\mu}_{B,k} - \bar{\mu}_{A,k}\right) / \left(\sigma_{A,k} + \sigma_{B,k}\right)$. The employed quantities are updated by using the relation $\sigma^2 = \mathbb{E}\left[x^2\right] - \mathbb{E}\left[x\right]^2$:

$$
\begin{aligned}
\bar{\mu}_{A,k} &= \frac{N_{A,k-1}\ \bar{\mu}_{A,k-1} + n_{A,k} \left\langle \boldsymbol{w}_k^T \mathcal{B}_{A,k} \right\rangle}{N_{A,k}} \\
\bar{\mu}_{2A,k-1} &= \mathbb{E}\left[\left\{ \left(\boldsymbol{w}_k^T \boldsymbol{p}\right)^2 \mid \boldsymbol{p} \in \mathcal{P}_{A,k-1} \right\}\right] = \bar{\mu}_{A,k-1}^2 + \sigma_{A,k-1}^2 \\
\bar{\mu}_{2A,k} &= \mathbb{E}\left[\left\{ \left(\boldsymbol{w}_k^T \boldsymbol{p}\right)^2 \mid \boldsymbol{p} \in \mathcal{P}_{A,k} \right\}\right] = \frac{N_{A,k-1}\ \bar{\mu}_{2A,k-1} + n_{A,k} \left\langle \left(\boldsymbol{w}_k^T \mathcal{B}_{A,k}\right)^2 \right\rangle}{N_{A,k}} \\
\sigma_{A,k} &= \sqrt{\bar{\mu}_{2A,k} - \bar{\mu}_{A,k}^2}
\end{aligned}
$$

where we have denoted with $\left(\boldsymbol{w}_k^T \mathcal{B}_{A,k}\right)^2$ the set of real values $\{(\boldsymbol{w}_k \cdot \boldsymbol{p})^2 | \boldsymbol{p} \in \mathcal{B}_{A,k}\}$. The updated quantities $\bar{\mu}_{B,k}$, $\bar{\mu}_{2B,k}$, and $\sigma_{B,k}$ are similarly evaluated.

The computational cost of the training algorithm is dominated by the incremental rank-d SVD that requires $O(DNd)$ operations for $d \leq \sqrt{\min(N, D)}$ (see Brand (2006)). Regarding the memory requirements, O-IPCAC stores $O(Dn_k + Dd_k)$ real values during the training tasks, and only $O(D)$ values during classification.

3. Data Description and Pre-processing

The data used in our tests have been distributed by the organizers of the MLSP 2010 competition for research purposes[2], and consist of EEG brain signals collected while the subject viewed satellite images and tried to detect those containing a predefined target. There are 64 channels of EEG data, the total number of samples is 176378, and the sampling rate is 256Hz. During the EEG recording 2775 satellite images were shown, partitioned in 75 activation blocks with 37 images per block. All images within a block were consecutively displayed for 100 ms (an "image trigger" is provided to indicate the time samples corresponding to the on-set of each image). Each block was initiated by the subject after a rest period, the length of which was not specified in advance. The classifier must analyze the brain activity to recognize those images containing the target.

We pre-processed each channel with a Gaussian filter with cut-frequency of 2.2Hz, and we subtracted the filtered data from the original one to obtain high-pass filtered signals. These signals were then used to extract 64×97 image blocks, where each image block starts exactly 65 time samples ($\approx$ 250ms) after the corresponding image trigger. We underline that each image block covers a time window approximately located between 250ms and 550ms after the image trigger, since this range contains the P300 waves (Picton, 1992) and other possible brain activations (Chiappa, 1997). The extracted blocks are serialized in 2775 vectors in $\Re^{6208}$, of which only 58 points represent images with target.

4. Results

In this section we consider both the results achieved in the MLSP 2010 competition, and the tests we performed to compare O-IPCAC with state-of-the-art online/incremental classifiers. Notice that the considered high dimensional EEG data cannot be processed by most batch algorithms such as LDA or SVM, due to either memory requirements, or training time. On the other side, online algorithms such as O-IPCAC can handle this problem since they perform subsequent training phases on mini-batches of training data. Therefore we compare our method with: Perceptron, Second Order Perceptron (SOP, Cesa-Bianchi et al. (2005)), Online Independent SVM (OISVM, Orabona et al. (2007)), Passive Aggressive (PA, Crammer et al. (2006)), Alma (Gentile, 2002), and Incremental LDA (ILDA, Kim et al. (2007))[3].

To evaluate the performance of our classifier comparing it with the other methods, we employed the dataset described in Section 3, we computed the Receiver Operating Characteristic (ROC) curve, and we estimated the Area Under the Curve (AUC). To obtain an unbiased evaluation, we performed ten-fold cross validation, and we averaged the computed sensitivity and specificity values.

2. http://www.bme.ogi.edu/∼hildk/mlsp2010Competition.html.
3. We appropriately tuned all the parameters of the employed algorithms. Notice that for the kernel methods the best results are achieved by choosing the linear kernel.

Figure 1: ROC curves

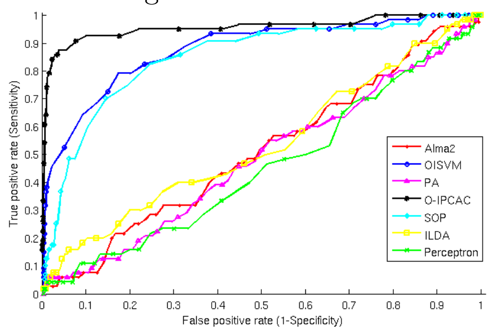

Figure 2: AUC per classifier

Classifier	AUC
O-IPCAC	**0.9541**
OISVM	0.8766
SOP	0.8479
ILDA	0.5315
Alma	0.5110
PA	0.4835
Perceptron	0.4507

Figures 1 and 2 show the obtained ROC curves and the AUC of the tested classifiers; it can be noted that O-IPCAC achieves the best results. Furthermore, these results demonstrate that the first order techniques (Perceptron, Alma, and PA) cannot discriminate the two classes, while O-IPCAC, SOP, and OISVM, which are second order techniques, achieve good results. Note that ILDA, despite being a second order method, obtains bad results since it is not able to manage datasets whose cardinality is lower than their dimensionality. We underline that our method is able to handle strongly unbalanced classes thanks to the thresholding method based on the Mahalanobis distance that avoids any experimental setup.

Regarding the MLSP competition, to evaluate the various approaches the organizers have acquired the training set (described in Section 3) and the test set by performing two different experiments. The test data differ from the training one since different image durations (50ms, 100ms, 150ms, and 200ms) are used in different activation blocks. While the training set was distributed by the MLSP organizers, the test set was not.

In the MLSP competition the efficacy of the proposed approaches were evaluated by estimating the ROC curves, and comparing the AUCs. Our algorithm covered the 80% of the area, ranking seventh among the 35 participants (Hild et al., Sept. 2010). It is worth noting that the first 10 algorithms have very close classification performance (between 82% and 79%). Considering that we performed just a high-pass filtering as pre-processing step, avoiding any kind of bootstrap aggregating technique, we believe that the achieved results are very promising and they confirm the quality of the proposed classifier.

5. Conclusions and Future Works

This work proposes an online/incremental linear binary classifier that has been developed to deal with: high dimensional data, classification problems where the cardinality of the point set is high or the data are dynamically supplied, and highly unbalanced training sets whose cardinality is lower than the dimensionality.

We evaluated the performance of our algorithm by executing experiments on EEG data distributed by the MLSP competition. It is important to underline that, instead of focusing on complex features extraction/selection techniques, we propose a classifier that is able to deal with the MLSP raw data achieving good results, as demonstrated in Section 4.

In future works we want to apply our method to datasets characterized by a very large ratio between dimension and training points, such as Microarray data. Furthermore, to cope with classification problems where the probability distribution underlying the data changes with time, we want to develop an adaptive version of `O-IPCAC`.

Acknowledgments

The authors would like to thank Professor Paola Campadelli for her invaluable support.

References

M. Brand. Fast low-rank modifications of the thin singular value decomposition. *Linear Algebra and Its Applications*, 2006.

S. C. Brubaker and S. Vempala. Isotropic pca and affine-invariant clustering. *Foundations of Computer Science, Annual IEEE Symposium on*, 2008.

N. Cesa-Bianchi, A. Conconi, and C. Gentile. A second-order perceptron algorithm. *SIAM J. Comput.*, 2005.

K.H. Chiappa. Evoked potentials in clinical medicine. In *Lippincott-Raven*, 1997.

K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer. Online passive-aggressive algorithms. *JMLR*, 7, 2006.

C. Gentile. A new approximate maximal margin classification algorithm. *JMLR*, 2002.

P. C. Hansen. The truncated svd as a method for regularization. Technical report, Stanford University, Stanford, CA, USA, 1986.

K. Hild, M. Kurimo, and V. Calhoun. The sixth annual mlsp competition. In *MLSP '10*, Sept. 2010. URL `http://www.bme.ogi.edu/~hildk/Hild_competition_mlsp2010.pdf`.

I. M Johnstone and A. Y. Lu. Sparse principal components analysis. *Journal of the American Statistical Association*, 2004.

T. Kim, S. Wong, B. Stenger, J. Kittler, and R. Cipolla. Incremental linear discriminant analysis using sufficient spanning set approximations. In *CVPR*. IEEE Computer Society, 2007.

F Lotte, M Congedo, A Lécuyer, F Lamarche, and B Arnaldi. A review of classification algorithms for eeg-based brainâĂŞcomputer interfaces. *Journal of Neural Engineering*, 2007.

F. Orabona, C. Castellini, B. Caputo, J. Luo, and G. Sandini. Indoor place recognition using online independent support vector machines. In *BMVC '07*, pages 1090–1099, 2007.

D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 2007.

T.W. Picton. The p300 wave of the human event related potential. In *Journal of Clinical Neurophysiology*, 1992.

A. Rozza, G. Lombardi, and E. Casiraghi. Novel ipca-based classifiers and their application to spam filtering. In *ISDA '09*. IEEE Computer Society, 2009.

A. Rozza, G. Lombardi, and E. Casiraghi. Pipcac: A novel binary classifier assuming mixtures of gaussian functions. In *AIA '10*. ACTA press, 2010.

Q. Wei, Y. Wang, X. Gao, and S. Gao. Amplitude and phase coupling measures for feature extraction in an eeg-based brain computer interface. *Journal of Neural Engineering*, 2007.