

Application of regularized logistic regression for movement-related potentials-based EEG classification

Hu Chenchen

Wang Haixian

(Research Center for Learning Science, Southeast University, Nanjing 210096, China)

Abstract: In order to improve classification accuracy, the regularized logistic regression is used to classify single-trial electroencephalogram (EEG). A novel approach, named local sparse logistic regression (LSLR), is proposed. The LSLR integrates the locality preserving projection regularization term into the framework of sparse logistic regression. It tries to maintain the neighborhood information of original feature space, and, meanwhile, keeps sparsity. The bound optimization algorithm and component-wise update are used to compute the weight vector in the training data, thus overcoming the disadvantage of the Newton-Raphson method and iterative re-weighted least squares (IRLS). The classification accuracy of 80% is achieved using ten-fold cross-validation in the self-paced finger tapping data set. The results of LSLR are compared with SLR, showing the effectiveness of the proposed method.

Key words: logistic regression; locality preserving projection regularization; electroencephalogram

doi: 10.3969/j.issn.1003-7985.2013.01.008

Brain-computer interface (BCI) is a system to help patients with severe motor disabilities (e.g., amyotrophic lateral sclerosis, cerebral vascular accident and cerebral palsy) to express their intentions and manipulate the external environment. Moreover, it provides a new way of entertainment through which people can play electronic games without finger movement^[1-5]. A variety of noninvasive technologies of electroencephalography (EEG), magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI) can be applied to BCI. Nevertheless, the EEG is commercially affordable and friendly operable. It also has excellent temporal resolution^[6]. Our BCI research focuses on EEG signals.

The BCI system contains signal acquisition, preprocessing, feature extraction, feature selection and pattern

matching that converts input signals into commands. The key point of BCI is signal processing, which has two purposes: feature extraction and classification. Current research typically uses features extracted by principal component analysis (PCA)^[7], independent component analysis (ICA)^[8], and common spatial pattern (CSP)^[9]. The favorite classifications include linear discriminate analysis (LDA)^[10], neural network (NN), and support vector machine (SVM)^[11]. For the classification problem, the estimation of posterior probability plays an important role and the posterior probability can be directly used to predict labels of testing trials^[12]. Logistic regression (LR) is a popular binary classification method based on the posterior probability and is often applied in EEG signal processing. The main advantage of LR is that it computes a probability value rather than a score and no prior feature extraction is required^[13-14]. Meanwhile, the regression coefficients resemble CSP filters that transform original high-dimensional EEG data into a low-dimensional feature space which contains beneficial information for classification. However, LR is a global linear approach that ignores manifold information between different EEG trials. Besides, when few EEG trials contain outliers, the performance of LR may be deteriorated. Therefore, it is useful to incorporate the manifold regularization into LR^[15].

The family of algorithms based on manifold learning^[16], such as isometric map (ISOMAP), locally linear embedding (LLE), Laplace eigenmaps, and locality preserving projection (LPP)^[17], is developed to preserve the topological structure of data points. In this paper, we combine the LPP regularizer term with sparse logistic regression (SLR)^[18-19], called local sparse logistic regression (LSLR). The goal of the LSLR approach is to preserve the neighborhood information of original feature space in the framework of SLR. Computationally, to obtain the weighing vector, we use a bound optimization method with a component-wise update procedure. We demonstrate the effectiveness of our proposed method on an EEG-based BCI competition data set.

1 Methods

1.1 Sparse logistic regression

Let $D = \{(X_1, y_1), \dots, (X_n, y_n)\}$ be a given data set. X_i is a raw EEG signal denoted by a $T \times C$ matrix, where T

Received 2012-09-13.

Biographies: Hu Chenchen (1987—), female, graduate; Wang Haixian (corresponding author), male, doctor, associate professor, hxwang@seu.edu.cn.

Foundation items: The National Natural Science Foundation of China (No. 61075009), the Natural Science Foundation of Jiangsu Province (No. BK2011595), the Program for New Century Excellent Talents in University of China, the Qing Lan Project of Jiangsu Province.

Citation: Hu Chenchen, Wang Haixian. Application of regularized logistic regression for movement-related potentials-based EEG classification[J]. Journal of Southeast University (English Edition), 2013, 29(1): 38–42. [doi: 10.3969/j.issn.1003-7985.2013.01.008]

is the number of recording time samples, C is the number of EEG channels, and $y_i \in \{-1, 1\}$ is the class label for the i -th input trial. In this paper, $y_i = 1$ and $y_i = -1$ denote finger movement conditions of right hand and left hand, respectively. LR, considered in this paper, is a popular binary classification method^[20]. The goal of LR is to determine the posterior probability of the occurrence of an event. The probability that the i -th trial belongs to class y_i is

$$P(y_i | \mathbf{x}_i, \boldsymbol{\omega}) = \frac{1}{\exp(-y_i \boldsymbol{\omega}^T \mathbf{x}_i)} \quad (1)$$

where \mathbf{x}_i denotes a $T \times C$ vector which is formed by straightening the elements of \mathbf{X}_i ; $\boldsymbol{\omega}$ is a vector representing regression coefficients, each element of which corresponds to a weight of a feature. We estimate the weight matrix by maximizing the log-likelihood function:

$$l(\boldsymbol{\omega}) = \sum_{i=1}^n \log P(y_i | \mathbf{x}_i, \boldsymbol{\omega}) \quad (2)$$

While maximizing $l(\boldsymbol{\omega})$, regularization is constantly introduced to enable the weight vector to be sparse^[21]. The popular regularizers are the LASSO term and the ridge term. The former is a sparse regularizer, which makes the weights of irrelevant features be zero; the latter enables the weights of relevant features to be larger and irrelevant features to be smaller without an exact tendency to be zero^[19, 22–25]. However, the two regularizers might ignore the local structure of the EEG time course which contains discriminative information among patterns. To address this problem, we introduce a new regularizer and combine it with the SLR, which will be described in the next section.

1.2 Locality preserving logistic regression

1.2.1 Locality preserving projection

Locality preserving projection (LPP), a linear dimensionality reduction algorithm, is a useful and effective method for pattern recognition. It builds a projection to preserve the neighborhood structure of the original data based on Euclidean distances^[17]. He et al.^[26] showed that the LPP algorithm gave a better representation of the data structure and achieved higher accuracy in face recognition. Watanabe and Kurita^[13] introduced the locality into the regularization term of LR and applied it to standard benchmark datasets. Some encouraging results were obtained^[13, 27].

When defining the weight matrix, one usually introduces the following forms^[28]:

1) 0-1 weighting: $Q_{ij} = 1$ if and only if nodes i and j are connected by an edge.

2) Heat kernel weighing: If nodes i and j are connected,

$$Q_{ij} = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\tau}\right) \quad (3)$$

where τ is a hyper parameter and it is usually determined by a cross validation strategy.

3) Dot-product weighing: If nodes i and j are connected,

$$Q_{ij} = \mathbf{x}_i^T \mathbf{x}_j \quad (4)$$

Clearly, if $\|\mathbf{x}_i\|_2$ and $\|\mathbf{x}_j\|_2$ are both equal to 1, Q_{ij} is the cosine of the two vectors.

Owing to the tunable parameter τ of the heat kernel weighing, we adopt the second form above to define the weight matrix.

1.2.2 LPP regularizer

In this paper, we propose the LPP regularizer as follows:

$$E_{\text{LPP}} = \sum_i^n \sum_j^n (\boldsymbol{\omega}^T \mathbf{x}_i - \boldsymbol{\omega}^T \mathbf{x}_j)^2 Q_{ij} \quad (5)$$

Minimizing E_{LPP} is an attempt to ensure that if \mathbf{x}_i and \mathbf{x}_j are close in the original space, then they are close in the generation space. It gives a large value of Q_{ij} if the distance of two samples is close while it gives small weight when meeting distant samples. It shows the advantage of preserving the helpful neighborhood information of the original feature space.

By combining the E_{LPP} term and the LASSO term, we obtain the log-likelihood function:

$$L(\boldsymbol{\omega}) = l(\boldsymbol{\omega}) - \lambda_1 \|\boldsymbol{\omega}\|_1 - \frac{\lambda_2}{2} E_{\text{LPP}} \quad (6)$$

where λ_1 and $\frac{\lambda_2}{2}$ denote the degrees of L1 norm and LPP regularizer, respectively. When solving Eq. (6), the Newton-Raphson method or iterative re-weighted least squares (IRLS) is typically used. The crucial disadvantage of the Newton-Raphson method is that different matrices of the log-likelihood function must be inverted at each iteration. Although the IRLS algorithm can easily solve a Gaussian prior on $\boldsymbol{\omega}$, it cannot handle the other priors such as the Laplacian prior. Thus in the next section, we use a bound optimization algorithm to compute the weight matrix $\boldsymbol{\omega}$.

1.2.3 Bound optimization

With the bound optimization approach, $l(\boldsymbol{\omega})$ is optimized by iteratively maximizing a surrogate function S ,

$$\hat{\boldsymbol{\omega}}^{(t+1)} = \arg \max S(\boldsymbol{\omega} | \hat{\boldsymbol{\omega}}^{(t)}) \quad (7)$$

where $\hat{\boldsymbol{\omega}}^{(t)}$ is the value of the t -th iteration. If the surrogate function satisfies that $l(\boldsymbol{\omega}) - S(\boldsymbol{\omega} | \hat{\boldsymbol{\omega}}^{(t)})$ attains its minimum when $\boldsymbol{\omega} = \hat{\boldsymbol{\omega}}^{(t)}$, $l(\boldsymbol{\omega})$ will increase monotonically at each iteration^[18].

When $l(\boldsymbol{\omega})$ is concave, a bound is obtained based on its Hessian matrix $\mathbf{H}(\boldsymbol{\omega})$ so as to construct the surrogate function $S(\boldsymbol{\omega} | \hat{\boldsymbol{\omega}}^{(t)})$. If there exists a negative definite matrix \mathbf{B} which satisfies that $\mathbf{H}(\boldsymbol{\omega}) - \mathbf{B}$ is positive semi-definite, we can obtain a valid surrogate function as fol-

lows^[18]:

$$S(\boldsymbol{\omega} \mid \hat{\boldsymbol{\omega}}^{(t)}) = \boldsymbol{\omega}^T (\mathbf{g}(\hat{\boldsymbol{\omega}}^{(t)}) - \mathbf{B}\hat{\boldsymbol{\omega}}^{(t)}) + \frac{1}{2} \boldsymbol{\omega}^T \mathbf{B} \boldsymbol{\omega} \quad (8)$$

where $\mathbf{g}(\boldsymbol{\omega})$ denotes the gradient of $l(\boldsymbol{\omega})$. The matrix \mathbf{B} of two classes is given by^[19]

$$\mathbf{B} = -\frac{1}{4} \sum_{i=1}^n (\mathbf{x}_i \mathbf{x}_i^T) \quad (9)$$

Adding the regularizers of LASSO and LPP, the surrogate function of $L(\boldsymbol{\omega})$ is given by

$$S(\boldsymbol{\omega} \mid \hat{\boldsymbol{\omega}}^{(t)}) - \lambda_1 \|\boldsymbol{\omega}\|_1 - \frac{\lambda_2}{2} \boldsymbol{\omega}^T \mathbf{Q} \boldsymbol{\omega} \quad (10)$$

By maximizing the surrogate function in the framework

$$\hat{\boldsymbol{\omega}}_k^{(t+1)} = \text{soft} \left(\frac{\hat{\boldsymbol{\omega}}_k^{(t)} - \mathbf{g}_k(\hat{\boldsymbol{\omega}}^{(t)})/B_{kk} + \lambda_2 \mathbf{Q}_k^T \hat{\boldsymbol{\omega}}^{(t)}/B_{kk} - \lambda_2 \mathbf{Q}_{kk} \hat{\boldsymbol{\omega}}_k^{(t)}/B_{kk}}{1 - \lambda_2 \mathbf{Q}_{kk}/B_{kk}}, -\frac{\lambda_1/B_{kk}}{1 - \lambda_2 \mathbf{Q}_{kk}/B_{kk}} \right) \quad (13)$$

where B_{ij} denotes the (i, j) element of matrix \mathbf{B} ; \mathbf{Q}_{ij} denotes the (i, j) element of matrix \mathbf{Q} ; and $\mathbf{g}_k(\hat{\boldsymbol{\omega}}^{(t)})$ denotes the k -th element of the gradient vector $\mathbf{g}(\hat{\boldsymbol{\omega}}^{(t)})$. The soft function is defined as

$$\text{soft}(a; \delta) = \text{sign}(a) \max\{0, |a| - \delta\} \quad (14)$$

From the theory of the bound optimization, we know that, with the above procedures, the value of the surrogate function monotonically increases at each iteration. The LSLR algorithm can be obtained in the following steps. First, we record the EEG data and do data preprocessing. Next, we initialize the parameters of $\lambda_1, \lambda_2, \tau, \boldsymbol{\omega}_{\text{old}}$ and choose the best performance of the four parameters, respectively, by using cross-validation. Then we obtain $\boldsymbol{\omega}_{\text{new}}$ from $\boldsymbol{\omega}_{\text{old}}$ by using Eq. (13) until $\|\boldsymbol{\omega}_{\text{new}}\|_2 - \|\boldsymbol{\omega}_{\text{old}}\|_2 < 10^{-6}$.

2 Experiment

2.1 Self-paced finger tapping dataset

The self-paced finger tapping data set is an EEG data set from BCI competition 2003-dataset IV^[29-31]. Our goal is to predict the type of upcoming finger movement before it really happens. The subject presses the corresponding keys at an average speed of once per second, using either the index finger or the little finger of the right or the left hand. The data set consists of 416 trials, in which 316 trials are used for training and the remaining trials are used for testing. The duration length of each trial is 630 ms, in which just the beginning 500 ms is adopted because the ending 130 ms mainly contains artifacts from the electro-myogram (EMG). The data contains 28 channels and is down sampled from 1 000 to 100 Hz.

2.2 Preprocessing

Before applying our algorithm, we use a band-pass Butterworth filter with cut-off frequencies from 7 to 35 Hz

of bound optimization, we obtain the update equation:

$$\hat{\boldsymbol{\omega}}^{(t+1)} = (\mathbf{B} - \lambda_1 \mathbf{A}^{(t)} - \lambda_2 \mathbf{Q})^{-1} (\mathbf{B}\hat{\boldsymbol{\omega}}^{(t)} - \mathbf{g}(\hat{\boldsymbol{\omega}}^{(t)})) \quad (11)$$

$$\mathbf{A}^{(t)} = \text{diag}\{|\hat{\boldsymbol{\omega}}_1^{(t)}|, \dots, |\hat{\boldsymbol{\omega}}_{T \times C}^{(t)}|\} \quad (12)$$

Note that the dimension of $\boldsymbol{\omega}$, $T \times C$, is very large, which implies that the iteration (11) is intensively computational. To alleviate the computational burden, we adopt a component-wise update procedure which focuses on the update of one component of $\boldsymbol{\omega}$ while maximizing the surrogate function. Specifically, for $j \neq k$, $\hat{\boldsymbol{\omega}}_k^{(t+1)}$ is remained as its previous item $\hat{\boldsymbol{\omega}}_k^{(t)}$; and for $j = k$, $\hat{\boldsymbol{\omega}}_k^{(t+1)}$ is updated from $\hat{\boldsymbol{\omega}}_k^{(t)}$. The component-wise update equation of $\boldsymbol{\omega}$ is given by

on the raw EEG data. Meanwhile, the temporal interval is selected as follows: We use the latter 200 ms time samples in the whole 500 ms because the data from this time window are more close to the real action, and have achieved higher classification accuracy than that of the whole 500 ms.

2.3 Experimental setting and results

In this experiment, we use classification accuracy to compare the results of SLR with LSLR. First, we select each element as λ_1 from the set $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3, 10^4\}$ to test the accuracy of SLR because we found that when we fixed the value of λ_2 , different values of λ_1 made little difference in the accuracy of LSLR. Therefore, we test the accuracy of SLR with different values of λ_1 first in order to find the best performance of λ_1 . From Fig. 1, we find that, when the value of λ_1 is less than 10, the test accuracy changes slowly. Otherwise, it dramatically decreases. Then, on the basis of the optimal λ_1 , the same test is conducted on parameter λ_2 to do the LSLR experiment. The ten-fold cross-validation is used to evaluate the classification accuracy as follows. All the trials are divided into ten divisions, nine of which are used as the training set while the remaining one of which is used as the testing set. The average classifica-

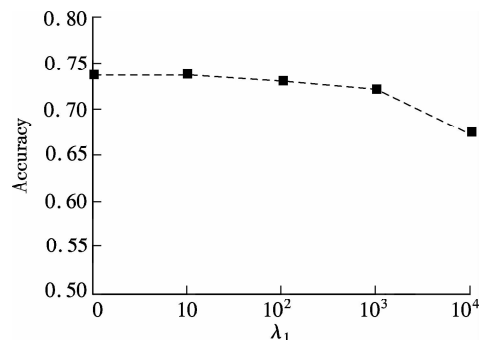


Fig. 1 Test accuracy of SLR vs. variation of parameter λ_1

tion accuracy of the ten folds is considered as the final recognition rate. Tab. 1 shows the performance of the LSLR method according to different values of the λ_2 penalty term.

Tab. 1 Average recognition values of SLR and LSLR using the ten-fold cross-validation

Methods	Recognition	λ_1	λ_2	Sparsity
SLR	0.78	10	0	0.460 7
	0.64		1	0.021 4
LSLR	0.66	10	0.1	0.071 4
	0.79		0.01	0.085 7
	0.80		0.001	0.248 2

We find that LSLR outperforms SLR when we keep the value of λ_1 invariant and select different values of λ_2 . The optimal LSLR performance is significantly above the average performance using SLR according to a t-test paired by cross-validation runs ($p < 0.05$). The improvement of classification performance comes from the local structure modeling of LSLR. However, the LPP regularizer yields a dense model whereas the LASSO term only selects 53.93% of the features. Fig. 2 contains box plots that compare the difference between SLR and LSLR in terms of test accuracy. The result strongly shows that our method outperforms SLR.

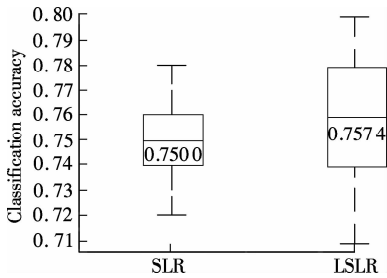


Fig. 2 Box plot of the accuracy using SLR and LSLR

Moreover, two other algorithms of LDA and LR are evaluated over the same dataset and the classification rates are demonstrated in Tab. 2. CSP is used to extract features before applying LDA as a classifier.

Tab. 2 Average recognition values of LDA, LR, SLR and LSLR

Algorithm	Average recognition value
LDA	0.57
LR	0.59
SLR	0.78
LSLR	0.80

From Tab. 2, it is noticed that the recognition values obtained by LDA and LR without regularization are low. The regularized logistic regression SLR and LSLR give better recognition rates than LR. Especially LSLR gives the best recognition rate for BCI competition 2003-dataset IV. These results prove that the LPP regularizer is effective for classification of neuroimaging data and the LPP regularizer can be a helpful regularizer.

Note that the recognition value of Wang et al.^[31], the winner in this dataset, achieved 84%. The algorithm combines common spatial subspace decomposition with LDA to extract features, and then uses a perceptron neural network as the classifier. In our algorithm, no prior feature extraction is required and we directly use LSLR to classify a single-trial EEG during the preparation of self-paced tapping.

2.4 Discussion

The LASSO term produces a sparse solution while the LPP regularizer has desirable properties. From Tab. 1 and Fig. 2, the improvement of classification performance comes from the LPP term, which provides the advantage of preserving the helpful neighborhood information of original feature space with less outlier distortion. It defines a small weight if the distance of two samples is large while it puts a large weight when meeting near neighbor samples. In our experiment, our classifier of the regularized logistic regression contains the step of feature extraction.

3 Conclusion

The goal of this paper is to demonstrate that the LPP regularizer, which is beneficial to the classification, can preserve the neighborhood information of the original feature space. Compared with SLR, LSLR has better discriminate performance. Actually, LSLR is the extension to logistic regression. Experimental results indicate that LPP regularization is a valuable regularization in single trial EEG data. Our future direction is to explore more useful regularizers and test their performance in EEG-based BCI application.

References

[1] Bashashati A, Fatourehchi M, Ward R K, et al. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals [J]. *J Neural Eng*, 2007, **4**(2): R32 – R57.

[2] Wolpaw J R, Birbaumer N, Heetderks W J, et al. Brain-computer interface technology: a review of the first international meeting [J]. *IEEE Trans Neural Syst Rehabil Eng*, 2000, **8**(2): 164 – 173.

[3] Birbaumer N, Ghanayim N, Hinterberger T, et al. A spelling device for the paralysed [J]. *Nature*, 1999, **398** (25): 297 – 298.

[4] Kübler A, Kotchoubey B, Kaiser J, et al. Brain-computer communication: unlocking the locked in [J]. *Psychol Bull*, 2001, **127**(3): 358 – 375.

[5] Wolpaw J R, Birbaumer N, McFarland D J, et al. Brain-computer interfaces for communication and control [J]. *Clin Neurophysiol*, 2002, **113**: 767 – 791.

[6] Haynes J D, Rees G. Decoding mental states from brain activity in humans [J]. *Nat Rev Neurosci*, 2006, **7**: 523 – 534.

[7] Wold S, Esbensen K, Geladi P. Principal component analysis [J]. *Chemometrics and Intelligent Laboratory*

- Systems, 1987, **2**(1/2/3): 37–52.
- [8] Hyvarinen A, Karhunen J, Oja E. Independent component analysis: algorithms and applications [J]. *Neural Computation*, 2001, **13**(4/5): 411–430.
- [9] Blankertz B, Tomioka R, Lemm S, et al. Optimizing spatial filters for robust EEG single-trial analysis [J]. *IEEE Signal Processing Magazine*, 2008, **25**(1): 41–56.
- [10] Mika S, Ratsch G, Weston J, et al. Fisher discriminant analysis with kernels [C]//*IEEE Signal Processing Society Workshop*. Madison, WI, USA, 1999: 41–48.
- [11] Cortes C, Vapnik V. Support-vector networks [J]. *Machine Learning*, 1995, **20**(3): 273–297.
- [12] Li T, Wang J, Wu X, et al. The estimate and application of posterior probability: based on kernel logistic regression [J]. *Pattern Recognition and Artificial Intelligence*, 2007, **19**(6): 689–695.
- [13] Watanabe K, Kurita T. Locality preserving multi-nominal logistic regression [C]//*19th International Conference on Pattern Recognition (ICPR 2008)*. Tampa, FL, USA, 2008: 1–4.
- [14] Aseervatham S, Antoniadis A, Gaussier E, et al. A sparse version of the ridge logistic regression for large-scale text categorization [J]. *Pattern Recognition Letters*, 2011, **32**(2): 101–106.
- [15] Chen Z, Haykin S. On different facets of regularization theory [J]. *Neural Computation*, 2002, **14**(12): 2791–2846.
- [16] Belkin M, Niyogi P, Sindhwani V. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples [J]. *The Journal of Machine Learning Research*, 2006, **7**(12): 2399–2434.
- [17] He X, Niyogi P. *Locality preserving projections* [D]. Chicago, IL, USA: University of Chicago, 2005.
- [18] Krishnapuram B, Carin L, Figueiredo M A T, et al. Sparse multinomial logistic regression: fast algorithms and generalization bounds [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(6): 957–968.
- [19] Ryali S, Supekar K, Abrams D A, et al. Sparse logistic regression for whole-brain classification of fMRI data [J]. *Neuroimage*, 2010, **51**(2): 752–764.
- [20] Bielza C, Robles V, Larranaga P. Regularized logistic regression without a penalty term: An application to cancer classification with microarray data [J]. *Expert Systems with Applications*, 2011, **38**(5): 5110–5118.
- [21] Poggio T, Girosi F. Regularization algorithms for learning that are equivalent to multilayer networks [J]. *Science*, 1990, **247**(4945): 978–982.
- [22] Cawley G C, Talbot N L C, Girolami M. Sparse multinomial logistic regression via Bayesian l1 regularisation [C]//*Advances in Neural Information Processing Systems*. Vancouver, CA, USA, 2007.
- [23] Meier L, Geer S V D, Buhlmann P. The group lasso for logistic regression [J]. *Journal of the Royal Statistical Society: Series B*, 2008, **70**(1): 53–71.
- [24] Zou H, Hastie T. Regularization and variable selection via the elastic net [J]. *Journal of the Royal Statistical Society: Series B*, 2005, **67**(2): 301–320.
- [25] Tomioka R, Müller K R. A regularized discriminative framework for EEG analysis with application to brain–computer interface [J]. *Neuroimage*, 2010, **49**(1): 415–432.
- [26] He X, Yan S, Hu Y, et al. Face recognition using Laplacianfaces [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(3): 328–340.
- [27] Kurita T, Watanabe K, Otsu N. Logistic discriminant analysis [C]//*IEEE International Conference on Systems, Man and Cybernetics*. San Antonio, USA, 2009: 2167–2172.
- [28] He X, Cai D, Shao Y, et al. Laplacian regularized Gaussian mixture model for data clustering [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2007, **23**(9): 1406–1418.
- [29] Blankertz B, Curio G, Müller K R. Classifying single trial EEG: towards brain computer interfacing [C]//*Advances in Neural Information Processing Systems*. Vancouver, CA, USA, 2002: 157–164.
- [30] Wang H X, Xu J. Local discriminative spatial patterns for movement-related potentials-based EEG classification [J]. *Biomedical Signal Processing and Control*, 2011, **6**(5): 427–431.
- [31] Wang Y, Zhang Z, Li Y, et al. BCI competition 2003-data set IV: an algorithm based on CSSD and FDA for classifying single-trial EEG [J]. *IEEE Trans Biomed Eng*, 2004, **51**(6): 1081–1086.

正则 logistic 回归在 EEG 信号运动相关电位分类中的应用

胡晨晨

王海贤

(东南大学学习科学中心, 南京 210096)

摘要: 单次脑电分类实验中, 采用基于 logistic 回归的正则化方法来提高分类准确率. 首先, 提出一种新算法——局部保持投影稀疏 logistic 回归, 将局部保持投影正则项加入到稀疏 logistic 回归中. 该算法旨在保留原始特征空间邻域信息的同时保证结果的稀疏性. 然后, 利用边界优化法和逐分量迭代算法在训练集上求解权重向量, 克服了牛顿-拉夫森法和迭代重加权最小二乘法的局限性. 最后, 在自步调手指运动数据集上采用十重交叉验证法得到 80% 的分类准确率, 并与稀疏 logistic 回归的实验结果进行对比, 说明局部保持投影正则项有效地保留了对脑电分类有用的信息.

关键词: logistic 回归; 局部保持投影正则化; 脑电图

中图分类号: TP391