

# An empirical comparison of different LDA methods in fMRI-based brain states decoding

Maogeng Xia<sup>a,b</sup>, Sutao Song<sup>c</sup>, Li Yao<sup>a,b,d</sup> and Zhiying Long<sup>a,b,\*</sup>

<sup>a</sup>State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, 100875, China

<sup>b</sup>Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing, 100875, China

<sup>c</sup>School of Education and Psychology, Jinan University, Jinan, 250022, China

<sup>d</sup>School of Information Science & Technology, Beijing Normal University, Beijing, 100875, China

**Abstract.** Decoding brain states from response patterns with multivariate pattern recognition techniques is a popular method for detecting multivoxel patterns of brain activation. These patterns are informative with respect to a subject's perceptual or cognitive states. Linear discriminant analysis (LDA) cannot be directly applied to fMRI data analysis because of the "few samples and large features" nature of functional magnetic resonance imaging (fMRI) data. Although several improved LDA methods have been used in fMRI-based decoding, little is known regarding the relative performance of different LDA classifiers on fMRI data. In this study, we compared five LDA classifiers using both simulated data with varied noise levels and real fMRI data. The compared LDA classifiers include LDA combined with PCA (LDA-PCA), LDA with three types of regularizations (identity matrix, diagonal matrix and scaled identity matrix) and LDA with optimal-shrinkage covariance estimator using Ledoit and Wolf lemma (LDA-LW). The results indicated that LDA-LW was the most robust to noises. Moreover, LDA-LW and LDA with scaled identity matrix showed better stability and classification accuracy than the other methods. LDA-LW demonstrated the best overall performance.

Keywords: LDA, classification, fMRI, decoding, regularized

## 1. Introduction

Recently, decoding brain states using functional Magnetic Resonance Imaging (fMRI) has attracted increasing attention in the field of cognitive neuroscience. Multivariate pattern recognition techniques have been widely applied to decode a person's brain activity. Unlike the popular univariate methods that employ a voxel-by-voxel basis to reveal brain activation patterns, multivariate pattern recognition methods consider brain voxels as an integration. They examine the entire pattern of activity observed across multiple voxels to predict which stimulus or task condition the subjects are in [1, 2].

---

\* Corresponding author: Zhiying Long, State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University. Tel.: 86-10-5880-7727; Fax: 86-10-5880-9444; E-mail: zlong@bnu.edu.cn.

Various multivariate pattern recognition techniques including Support Vector Machine (SVM), Logistic regression, Linear Discriminant Analysis, Naïve Bayesian and Neural Networks (NN) have been applied to fMRI data analysis [1-3]. Because of its low computation cost and satisfactory performance, LDA is a popular brain states classifier [2-4]. In fMRI-based decoding, voxels of fMRI data are regarded as features and are input for classification. Because fMRI data contain many voxels compared with the number of training samples, LDA encounters the dimension curse when it is directly applied to fMRI data with “high dimensional, low sample size (HDLSS)” [5, 6]. Specifically, the within class scatter matrix ( $\mathbf{S}_w$ ) becomes singular for data with high dimension. Therefore, LDA cannot be directly applied to fMRI data analysis.

Two improved LDA approaches were proposed to address the HDLSS problem [5]. One type is a subspace method which rejects the null space and pre-projects the original data to a subspace with small features. LDA was then implemented in the subspace. LDA combined with Principle Component Analysis (LDA-PCA) is based on this idea. However, LDA-PCA may discard the discriminative information contained in the removed components [5]. Another improved LDA approach employs a regularization method to solve this problem. Regularization techniques are successful in the solution of ill- and poorly-posed inverse problems. Regularization applies an additional term as a perturbation to diminish larger eigenvalues and enlarge small eigenvalues of  $\mathbf{S}_w$ . The regularization can be added by using the shrinkage estimator. Shrinkage estimators have three common forms including the identity matrix, scaled identity matrix and the diagonal matrix. The optimal shrinkage intensity was commonly determined by cross validation with expensive computation. A novel shrinkage covariance estimator that exploits the Ledoit-Wolf lemma for analytic calculation of the optimal shrinkage intensity was proposed by Schäfer [7, 8], and was examined in this study.

Recently, many studies have applied the two types of LDA methods to decode brain states on fMRI [2, 4, 9-12]. LDA-PCA is the most popular LDA method in fMRI-based decoding. Many fMRI studies apply LDA-PCA to classify visual objects and to acquire good classification performance [2, 9, 11]. Misaki, et al. compared the performance of LDA with the optimal-shrinkage covariance estimate and five other multivariate classifiers at decoding the category of visual objects from response patterns in human early visual and inferior temporal cortex. Results indicated that LDA (with an optimal-shrinkage covariance estimator) and the linear SVM performed best [12]. All of the previous studies in decoding brain states suggest that LDA is a simple method that can achieve effective and powerful performance in classification.

Despite widespread use of LDA in fMRI data, investigating the performance of different LDA methods in fMRI-based brain state decoding can provide an insight into the selection of suitable LDA methods for fMRI data analysis. Although our previous study compared the regularized LDA and LDA-PCA using real fMRI data and found that LDA using scaled identity matrix (LDA-Scaled) performed best [13], it did not consider LDA using optimal-shrinkage covariance estimator described by Ledoit and Wolf (LDA-LW) and investigate the robustness to noise of the various LDA methods. The purpose of the current study was to empirically evaluate and compare the performance of various LDA methods in fMRI-based brain state decoding. Both simulated and real fMRI data were used to compare the performance of five LDA methods. The five LDA methods include LDA-PCA, LDA using identity matrix (LDA-Identity) as regularization, LDA-Scaled, LDA using diagonal matrix (LDA-Diagonal) as regularization and LDA-LW. Results indicated that LDA-LW and LDA-Scaled resulted in more accurate classification in fMRI decoding compared with LDA-PCA, LDA-Identity and LDA-Diagonal.

## 2. Theory

### 2.1. Linear discriminant analysis

The basic idea of LDA is to find a linear projection to maximize the scatter between classes and minimize the scatter within classes [2, 3, 14]. The linear equation can be represented as follows:

$$y = w^T x \quad (1)$$

in which  $x$  denotes the samples,  $y$  is the projected samples and  $w$  is the projection matrix that projects the raw data  $x$  into a discriminant projection. LDA attempts to identify the projection matrix that maximizes the margin among different classes. The optimization problem is formulated as Eq. (2).

$$w = \arg \max_w J(w) = \frac{w^T S_b w}{w^T S_w w} \quad (2)$$

When the data dimension is larger than the number of samples, LDA encounters a “dimension curse” problem [1]. For HDLSS data, the scatter matrix within classes  $S_w$  is irreversible. To solve the “dimension curse” problem, the matrix  $S_w$  must be reversible. There are two approaches to solve this problem. One method is to reduce the feature dimension of data using PCA and the other one is to apply the regularized LDA methods that can obtain a nonsingular matrix by appending a regularized term.

### 2.2. Regularized LDA

In this section, some common regularized LDA methods were introduced. Regularized LDA methods use shrinkage estimators to shrink the sample covariance matrix toward target structures

$$\hat{R} = \lambda D + (1 - \lambda)S \quad (3)$$

in which  $D$  is a positive definite matrix and  $S$  is the sample covariance matrix. Different regularized LDA methods set  $D$  to different forms. The  $\hat{R}$  matrix in the LDA-Identity, LDA-Diagonal and LDA-Scaled methods can be expressed as Eqs. (4)-(6).

$$\text{LDA-Identity: } \hat{R} = \lambda I + (1 - \lambda)S \quad (4)$$

$$\text{LDA-Diagonal: } \hat{R} = \lambda \text{diag}(S) + (1 - \lambda)S \quad (5)$$

$$\text{LDA-Scaled: } \hat{R} = \lambda \left( \frac{\text{tr}(S)}{N} \right) I + (1 - \lambda)S \quad (6)$$

In these equations,  $\lambda$  is the regularized parameter that represents the shrinkage intensity. All three methods are regularized using an additional term weighted by  $\lambda$ , and the additional terms can convert  $S$  into non-singularity. The optimal regularized parameter is commonly determined by cross

validation with expensive computation.

Schäfer proposed a novel shrinkage covariance estimator that exploits the Ledoit-Wolf lemma for analytic calculation of the optimal shrinkage intensity [7, 8]. Its superiority is to have the covariance matrix be reversible and maintain positive definite and well-defined properties. The  $\hat{R}$  matrix in the LDA-LW method can be expressed as follows:

$$\text{LDA-LW: } \hat{R} = \lambda S + (1 - \lambda)U \quad (7)$$

in which  $U$  is the correlation coefficient matrix, and  $\lambda$  is a shrinkage parameter. This method has been discussed previously [7, 8].

### 3. Experiments

#### 3.1. Simulations

In this section, a simulated experiment was performed to illustrate the robustness to noises of the five LDA methods. The classification performance of the five LDA methods was compared under different noise levels.

The simulated experiment was designed in a block fashion in which each dataset consisted of two runs. For each run, a two-dimensional  $36 \times 36$  matrix with each pixel's intensity of 100, was duplicated 160 times, one for each time point, to simulate real fMRI time series data. Gaussian noise, with zero mean and a specific standard deviation ( $SD$ ), was added to all pixels at each time point to simulate system noise. Three rectangular regions of interests (ROIs) (the binary masks) over this matrix were constructed for the introduction of brain activations by tasks. Each run was composed of eight task blocks that alternated with eight control blocks. The two tasks occurred in an ABBABAAB sequence for each run. The simulated fMRI response of each task was derived from convolution of the stimulus paradigm of each task with a hemodynamic response function (HRF). It was assumed that task A activated regions S1 and S2, and task B activated regions S2 and S3. The signal change of each task randomly ranged from 0.5–2% relative to the mean intensity value of the individual pixel. The  $SD$  of the Gaussian noise varied from 0.5 to 3 with a step size of 0.5. Twenty datasets were generated at each noise level.

For each dataset, the first run was used as the training data and the second run was used as the testing data. For the training data, the Pearson correlation between the time series of each pixel and the reference function of each task was calculated. The reference function of each task was derived from the convolution of the stimulus paradigm of each task with the hemodynamic response function (HRF). 150 pixels with the largest correlation coefficient were selected for each task. The selected pixels were used to estimate the performance of LDA-Identity, LDA-Diagonal, LDA-Scaled and LDA-LW classifiers. The parameter  $\lambda$  was set from 0.1 to 1 with a step size of 0.1 and was determined using a two-fold cross validation for the LDA-Identity, LDA-Diagonal and LDA-Scaled methods. For LDA-PCA, PCA was applied to retain 95% variance. The classification accuracy of each dataset was obtained by calculating the ratio between the number of time points that were correctly classified and the number of total time points. For each LDA method, the mean classification accuracy across the 20 datasets at each noise level was obtained.

### 3.2. Real fMRI experiment

In this section, real fMRI data were used to examine the classification performance of the five LDA methods for brain state decoding. The real fMRI data used in this study was the same as a previous study that examined the feature selection methods [15].

#### 3.2.1 Participants and scanner parameters

Fifteen right-handed college participants with normal vision were recruited. All participants gave written consent. Brain scans were performed using a 3.0-T Siemens whole-body MRI scanner. A single-shot T2\*-weighted gradient-echo, echo planar imaging sequence was used for functional imaging acquisition with the following parameters: repeated time (TR) = 2000 ms, echo time (TE) = 30 ms, flip angle (FA) = 90°, field of view (FOV) = 190×200 mm, matrix = 64×64 and slice thickness = 1.33 mm.

#### 3.2.2 Feature selection

We selected the first four runs as the training datasets and subsequent four runs as the test datasets. Generalized linear model (GLM) analysis in SPM8 was applied to each subject's training datasets. Four regions of interest (ROI) were taken into consideration, including the fusiform gyrus, the inferior temporal gyrus, the inferior occipital gyrus and the middle occipital gyrus were defined in the WFU Pickatlas (<http://www.fmri.wfubmc.edu>). Voxels that are significantly activated by house/face/car/cat ( $p < 0.001$ , uncorrected) within the ROIs were selected as the features of the training datasets.

#### 3.2.3 Classification

The LDA-Identity, LDA-Diagonal, LDA-Scaled and LDA-LW were applied to the selected features of each participant to decode brain states. The regularization parameter  $\lambda$  was chosen same as the simulated experiment. For each LDA method, six two-class LDA classifiers and one four-class LDA classifier were trained. After the LDA classifiers were trained, the classifiers were applied to the test data to decode the brain states. The test data used identical features as the training data.

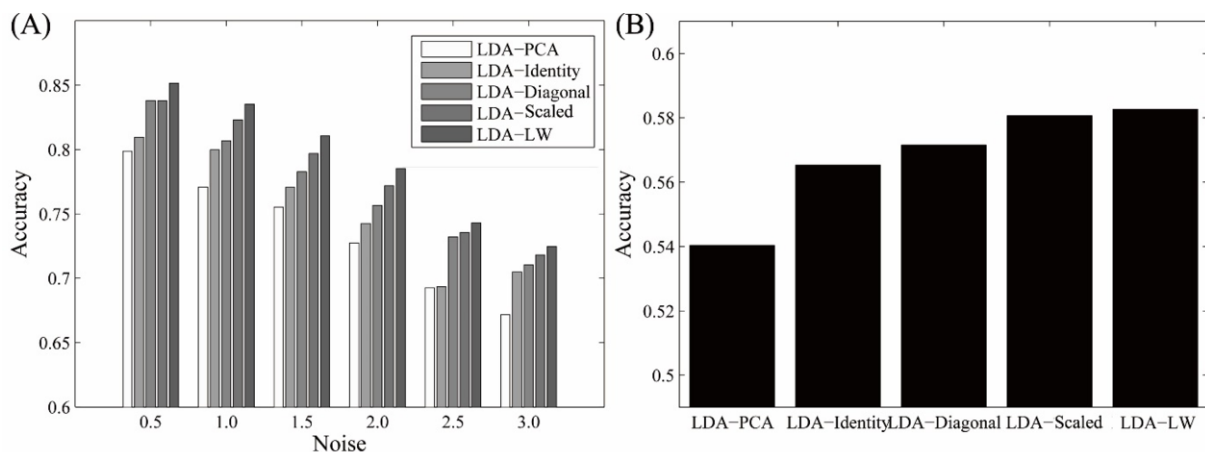


Fig. 1. Results of simulated and real fMRI data. (A) The mean classification accuracy of the simulated data at different noise levels for the five LDA methods. (B) The mean accuracy of the four-class classification for real fMRI data using the five LDA approaches.

#### 4. Results

Figure 1(A) shows the classification performance of the simulated data using LDA-PCA, LDA-Identity, LDA-Diagonal, LDA-Scaled and LDA-LW. The classification accuracy of all five LDA methods decreased with increasing noise. LDA-PCA had the lowest classification accuracy and LDA-LW demonstrated the highest accuracy. The three regularized LDA methods (LDA-Identity, LDA-Diagonal and LDA-Scaled) exhibited better performance than LDA-PCA and poor performance compared with LDA-LW at all noise levels.

The average four-class classification accuracies over all subjects estimated by the five LDA approaches are presented in Figure 1(B). LDA-LW exhibited the highest average classification accuracy. LDA-Identity, LDA-Diagonal, LDA-Scaled and LDA-LW displayed better classification performance than LDA-PCA. The mean accuracies of the six two-class classifications are presented in Figure 2. For house vs. face, face vs. cat, face vs. car and cat vs. car, both LDA-Scaled and LDA-LW displayed higher classification accuracy than LDA-PCA. However, for house vs. cat and house vs. car, the accuracy of LDA-PCA was higher than LDA-Scaled and LDA-LW. The performance of LDA-Identity and LDA-Diagonal demonstrated poor performance in most two-class classifications.

#### 5. Discussion

To gain a better understanding of the behavior of LDA, the performance of LDA-PCA, LDA-Identity, LDA-Diagonal, LDA-Scaled and LDA-LW were compared empirically through the use of simulation with different noise levels and real fMRI data. Results of both simulated and real fMRI data demonstrated that the overall classification performance of LDA-LW is better than the other LDA methods.

Results of the simulated data indicated that LDA-LW resulted in the most accurate classification, followed by LDA-Scaled, LDA-Diagonal, LDA-Identity and LDA-PCA at all noise levels (see Figure 2). LDA-LW was the most robust with respect to noise, whereas LDA-PCA was the least robust. For LDA-PCA, After PCA reduction, LDA can be applied to the data using the empirical covariance

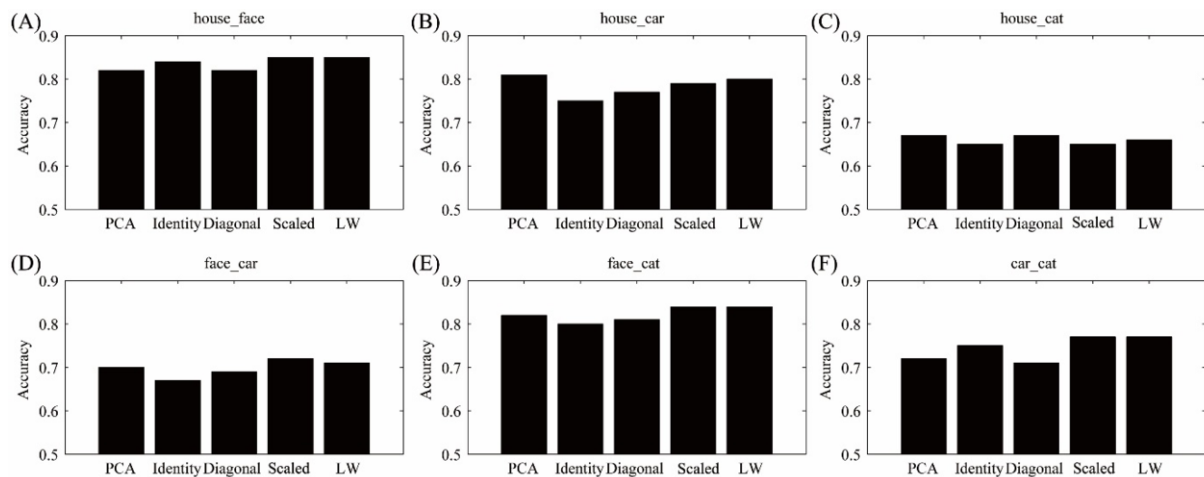


Fig. 2. The mean accuracy of the two-class classification of real fMRI data using the five LDA methods for house vs. face (A), house vs. cat (B), house vs. car (C), face vs. cat (D), face vs. car (E) and cat vs. car (F).

matrix. Although PCA maximizes the data structure information in the principal space and is optimal for data reconstruction [3, 5], it is the discriminative information that is important for pattern recognition. Because regularization can reduce the variance associated with the sample-based estimate of covariance matrix [3, 5], the four regularized LDA methods outperformed LDA-PCA at all noise levels for the simulated data.

The results of the four-class classification for the real fMRI data are consistent with the simulated results (see Figure 1). The classification accuracies of LDA-LW, LDA-Scaled, LDA-Diagonal, LDA-Identity and LDA-PCA are in descending order. However, the results of two-class classifications are not consistent with the simulated results (see Figure 2). For real fMRI data, the noise is complex and includes system noise, thermal noise and physiological noise. More variations existed in the results of the real fMRI data than the simulated results. Although there were some variations across the results of the six two-class classifications, some trends can still be identified. The classification accuracy of LDA-LW and LDA-Scaled is higher than that of LDA-PCA in most cases. Results of both the simulated and the real fMRI data suggest that LDA-LW and LDA-Scaled were more robust to noise and were more stable than LDA-Identity, LDA-Diagonal and LDA-PCA. Because the improved covariance estimator can provide more accurate covariance estimate and is always a positive definite even for small sample size, LDA-LW performs better compared with the other LDA methods. LDA-Scaled scales the identity matrix by the average of the eigenvalues. It decreases the larger eigenvalues and increases the smaller ones, counteracting the inherent bias in sample-based estimation of eigenvalues [3, 5]. LDA-Scaled displayed superior performance compared with LDA-Identity and LDA-Diagonal.

There is a regularization parameter  $\lambda$  that denotes the shrinkage intensity for the four regularized LDA methods. It is critical to select an optimal value for the shrinkage parameter  $\lambda$ . In this study, the parameter  $\lambda$  in LDA-Identity, LDA-Diagonal and LDA-Scaled was estimated by two-fold cross validation. Ten-fold cross validation was also examined in the experiment, similar results obtained while five times more time costs. As a result, we choose the 2-fold cross validation. For LDA-LW, the parameter  $\lambda$  is chosen by using a simple theorem that was derived by Ledoit and Wolf [7, 8].

In summary, the present study revealed both LDA-LW and LDA-Scaled had better robustness to noise, stability and classification accuracy than LDA-Identity, LDA-Diagonal and LDA-PCA. Moreover, LDA-LW outperformed LDA-Scaled in both classification accuracy and computational time cost.

## Acknowledgment

This work is supported by Key Program of National Natural Science Foundation of China (91320201), the National Natural Science Foundation of China (61271111 and 61473044).

## References

- [1] K.A. Norman, S.M. Polyn, G.J. Detre and J.V. Haxby, Beyond mind-reading: multi-voxel pattern analysis of fMRI data, *Trends in Cognitive Sciences* **10** (2006), 424–430.
- [2] D.D. Cox and R.L. Savoy, Functional magnetic resonance imaging (fMRI) “brain reading”: Detecting and classifying distributed patterns of fMRI activity in human visual cortex, *Neuroimage* **19** (2003), 261–270.
- [3] T. Li, S. Zhu and M. Ogihara, Using discriminant analysis for multi-class classification: An experimental investigation, *Knowledge and Information Systems* **10** (2006), 453–472.

- [4] T. Carlson, P. Schrater and S. He, Patterns of activity in the categorical representations of objects, *Journal of Cognitive Neuroscience* **15** (2003), 704–717.
- [5] J. Lu, K.N. Plataniotis and A.N. Venetsanopoulos, Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition, *Pattern Recognition Letters* **26** (2005), 181–191.
- [6] G. Cao and C. Bouman, Covariance estimation for high dimensional data vectors using the sparse matrix transform, *Advances in Neural Information Processing Systems* **21** (2009), 225–232.
- [7] O. Ledoit and M. Wolf, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, *Journal of Empirical Finance* **10** (2003), 603–621.
- [8] J. Schäfer and K. Strimmer, A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics, *Statistical Applications in Genetics and Molecular Biology* **4** (2005), 1175–1189.
- [9] J. Mourão-Miranda, A.L. Bokde, C. Born, H. Hampel and M. Stetter, Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data, *Neuroimage* **28** (2005), 980–995.
- [10] S. LaConte, S. Strother, V. Cherkassky, J. Anderson and X. Hu, Support vector machines for temporal classification of block design fMRI data, *Neuroimage* **26** (2005), 317–329.
- [11] S. Ku, A. Gretton, J. Macke and N.K. Logothetis, Comparison of pattern recognition methods in classifying high-resolution BOLD signals obtained at high magnetic field in monkeys, *Magnetic Resonance Imaging* **26** (2008), 1007–1014.
- [12] M. Misaki, Y. Kim, P.A. Bandettini and N. Kriegeskorte, Comparison of multivariate classifiers and response normalizations for pattern-information fMRI, *Neuroimage* **53** (2010), 103–118.
- [13] M. Xia, S. Song, X. Wu, L. Yao and Z. Long, Comparison of different LDA methods in decoding brain states on fMRI data, *The Organization for Human Brain Mapping (OHBM), Hamburg, Germany, 2014.*
- [14] R.A. Fisher, The use of multiple measurements in taxonomic problems, *Annals of Eugenics* **7** (1936), 179–188.
- [15] C. Zhang, S. Song, X. Wen, L. Yao and Z. Long, Improved sparse decomposition based on a smoothed L0 norm using a Laplacian kernel to select features from fMRI data, *Journal of neuroscience methods* **245** (2015), 15–24.