Applied Informatics

CrossMark

A spatial-constrained multi-target regression model for human brain activity prediction

Zhenfu Wen[1,2] and Yuanqing Li[1,2]*

*Correspondence:
auyqli@scut.edu.cn
[1] Center for Brain
Computer Interfaces
and Brain Information
Processing, South China
University of Technology,
Guangzhou 510640, China
Full list of author information
is available at the end of the
article

Abstract

Analyzing functional magnetic resonance imaging (fMRI) data from the encoding perspective provides a powerful tool to explore human vision. Using voxel-wise encoding models, previous studies predicted the brain activity evoked by external stimuli successfully. However, these models constructed a regularized regression model for each single voxel separately, which overlooked the intrinsic spatial property of fMRI data. In this work, we proposed a multi-target regression model that predicts the activities of adjacent voxels simultaneously. Different from the previous models, the spatial constraint is considered in our model. The effectiveness of the proposed model is demonstrated by comparing it with two state-of-the-art voxel-wise models on a publicly available dataset. Results indicate that the proposed method can predict voxel responses more accurately than the competing methods.

Keywords: fMRI, Encoding, Spatial constraint, Multi-target regression

Background

One important goal of neuroscience is to understand the relationship between external visual stimulus and human brain activity. We can gain the understanding by analyzing fMRI data from the mirror perspectives of neural decoding and neural encoding (Naselaris et al. 2011). In the view of neural decoding, we often attempt to predict information of stimuli from measured brain activity. Numerous studies have explored human vision using decoding models (Haxby et al. 2001, 2014; Norman et al. 2006). Conversely, in the view of neural encoding, we try to model how brain activity varies corresponding to external stimulus and attempt to predict brain activity from stimuli features. Previous studies have indicated that encoding models are more efficient in describing the function of brain areas than decoding models (Naselaris et al. 2011), suggesting the advantages of analyzing fMRI in the encoding view.

In recent years, voxel-based encoding models were proposed and caught much attention (Kay et al. 2008). A typical encoding model can be divided into two parts. The first part tries to find a feature space to describe the external stimulus. The second part corresponds to the construction of regression models, which uses the stimulus features to predict corresponding brain activity. Lots of effort were taken to find ways to represent the stimulus images. Previous studies used Gabor wavelet pyramid model (Kay et al.

2008; Vu et al. 2011), two-layer sparse coding model (Güçlü and van Gerven 2014), and convolutional neural networks (Agrawal et al. 2014) to extract features that can represent natural images effectively. However, fewer studies focused on efficient regression model construction.

In the regression part of encoding, regularized linear regression models such as lasso (Kay et al. 2008), ridge regression (Güçlü and van Gerven 2014) and graph-constrained elastic net (Kay et al. 2008; Schoenmakers et al. 2013) were most commonly used. Recently, a more advanced sparse nonparametric regression model was proposed (Vu et al. 2011). In spite of the successful prediction of brain activity using these models, one drawback of these voxel-wise models in previous studies is that the response of each voxel is modeled separately; thus, the estimated parameters of different voxels are independent. As a result, these regression models cannot fully employ the correlations between voxels and brain regions. Numerous studies have indicated the benefits of taking the spatial smoothness of fMRI data into account. For example, in the decoding models, when the spatial structure of the data is considered, higher decoding accuracies and more informative and interpretable results can be obtained (Michel et al. 2011; de Brecht and Yamagishi 2012). In functional brain mapping, combining local brain activity often results in more consistent patterns across subjects (Kriegeskorte et al. 2006). All these results suggest that spatial structure of fMRI data should also be considered in encoding models.

In this paper, we focus on the part of regression models construction in the encoding models, i.e., given the features of external stimuli images, we try to construct a regression model that can predict internal brain activity efficiently. We employ the spatial smoothness property of fMRI data and construct a multi-target linear regression model (Evgeniou and Pontil 2004, Argyriou et al. 2008) in which the activities of local adjacent voxels will be predicted simultaneously, and a spatial constraint is proposed to restrict the model parameters. To demonstrate the effectiveness of this model, we compare the brain activity prediction performances of the proposed method with two state-of-the-art voxel-wise models on a public fMRI dataset.

Methods

Data description

The publicly available fMRI data (Kay et al. 2011) were used for model validation; this dataset is widely used in comparing models (Güçlü and van Gerven 2014; Naselaris et al. 2009; Agrawal et al. 2014), and detailed experiment information is available in the original papers (Kay et al. 2008; Naselaris et al. 2009). The fMRI responses were recorded when human subjects viewing grayscale natural images while fixating on a central white square. Two subjects took part in the experiments. They viewed 1750 training images (for encoding model training), each presented twice; and 120 validation images (for encoding model testing), each presented ten times. For each subject, the data were acquired in five scanner sessions on five different days. Each scan session consisted of five training runs, each lasted 11 min, and two validation runs, each lasted 12 min.

The brain activity from the occipital cortex were recorded at a spatial resolution of 2 mm × 2 mm × 2.5 mm and a temporal resolution of 1 s using a 4T INOVA MR scanner (Varian, Inc.). Brain volumes were co-registered to correct head movements, and the

time-series data were deconvolved from the data to account for the delay in the hemodynamic response (Friston et al. 1994). Thus after the preprocessing, each stimulus image corresponds to one brain volume. The voxels in early visual areas were further divided into visual area one (V1), visual area two (V2,) and visual area three (V3). We only considered brain activity prediction in these areas in this study.

Problem formulation

In a standard regression framework, the design matrix $X \in \mathfrak{R}^{N \times M}$ is formed by $1 \times M$ feature vectors $x_s, s = 1, 2, \ldots, N$ of N samples. The goal is to predict the value of a $N \times 1$ target vector y, which contains corresponding target values of x_s. In this work, the design matrix comprises the features of N stimuli images, and the target vector is composed of intensities of a voxel, with each intensity corresponding to a image feature vector. Thus the problem here is to find a model that can predict voxel activity in response to stimuli accurately.

To evaluate the encoding performance of the prediction models, we calculate the coefficient of determination (R^2) between the observed and predicted voxel responses across the samples in the validation set. The R^2 is defined as

$$R^2 = 1 - \frac{\|y - \hat{y}\|^2}{\|y - \bar{y}\|^2},$$ (1)

where $\| \cdot \|$ is the Euclidean norm in $\mathfrak{R}^n$, y is the recorded true response vector, $\hat{y}$ is the predicted response vector, and $\bar{y}$ is the mean response vector. A higher R^2 means the model performs better in the prediction.

Voxel-wise models

Most voxel-wise models proposed in previous studies assume that voxel response is a weighted sum of the transformed image features. The regression model for each voxel is constructed separately, i.e., the model of voxel v is

$$y_v = Xb_v + \varepsilon_v, v = 1, 2, \ldots, V,$$ (2)

where $X \in R^{N \times M}$ is the design matrix that contains features of stimuli images, $b_v \in R^M$ is the parameter of the model, M is the number of features of each stimuli image, V is the total number of voxels and ε_v is zero mean Gaussian random vector.

A common problem that often occurs in regression is the so-called over-fitting, which may result in models with good performance in training data, but poor generalization performance in testing data. To estimate the model and control over-fitting, the common method is to find parameters that minimize sum-of-squares error function with an additional regularization term added:

$$L(b_v) = \|y_v - Xb_v\|^2 + \lambda_v J(b_v),$$ (3)

where X is the known design matrix, b_v is the parameter to estimate. The first term in the right side is the usual sum of squared errors, and $J(b_v)$ is a function of b_v as a penalty term, λ_v is the regularization coefficient that controls the relative importance of the error term and penalty term $J(b_v)$. One widely used $J(b_v)$ is the sum of squares of the weight vector elements:

$$J(b_v) = \frac{1}{2}\|b_v\|^2. \tag{4}$$

This is often termed ridge regularizer. Minimizing $L(b_v)$ with ridge regularizer controls over-fitting and yields a closed-form solution.

Another popular regularizer is the $\ell 1$ norm of the weight vector elements:

$$J(b_v) = \|b_v\|_1, \tag{5}$$

where $\|\cdot\|_1$ is the $\ell 1$ norm in $\mathfrak{R}^n$. This regularizer is often termed Lasso (Tibshirani 1996). The Lasso regularizer often results sparse parameter estimation with many parameters shrunk to zero.

To determine the optimal λ_v in the models, we conduct a nested threefold cross-validation and choose for each voxel v that model which maximizes the correlation between Xb_v and y_v on hold-out data. As done in the previous study (Schoenmakers et al. 2013), we sample lambda in the range $(10^{-5}, 10^5)$ on a log scale. For the convenience of discussion, we refer the voxel-wise model with ridge regularizer as Ridge and the model with Lasso regularizer as Lasso.

Proposed model

The voxel-wise models proposed in previous studies constructed regression model for each voxel separately, but ignored the dependents between voxels. However, fMRI data often possess the specific spatial smoothness property, and voxels from the same local brain area often exhibit similar properties. To elevate the performance of brain activity prediction, we employ the spatial smoothness property of fMRI data and construct a multi-target regression model.

For each voxel v, we construct the response matrix

$$Y_v = [y_v, y_{v1}, \ldots, y_{v(q-1)}]., \tag{6}$$

where q is the total number of voxel v's neighbors, and $y_{vj}, j = 1, 2, \ldots, q-1$ are the response vectors of voxel v's neighbors. The neighbors of v are defined as voxels contained in a sphere that centered on voxel v. In this work, we set the radius of the sphere to 3 voxel size, results in 33 voxels as each voxel v's neighbors, i.e., q equals 33. We try to minimize the total error function for voxel v:

$$L(B_v) = Tr[(XB_v - Y_v)^T(XB_v - Y_v) + \lambda_1(B_v R_v)^T(B_v R_v) + \lambda_2 B_v^T B_v]., \tag{7}$$

where X is the same as in voxel-wise models, $B_v \in \mathfrak{R}^{M \times q}$ is the parameter to determine, $Tr[X]$ means the trace of matrix X, and R_v is a $q \times q$ matrix with the (i, j) element being

$$R_{i,j} = \begin{cases} q-1 & if \ i = j \\ -1 & otherwise \end{cases} \tag{8}$$

The first element in the trace operator is the sum-of-squares error function to make sure the predicted response matrix $\hat{Y}_v$ is similar to the true response matrix. The second element is a regularizer that controls the parameter matrix B_v and trends to set the estimated parameter of voxel v similar to its neighbors. Here, we hypothesize that a voxel responds to external stimuli in a similar way as those voxels that locate around it; thus,

these voxels may possess similar parameters in the regression model. The third element is a regularizer similar with the ridge penalty to control over-fitting.

Model estimation

To estimate the model, we consider the gradient of the total error function:

$$\nabla L(B_v) = 2 \times [X^T X B_v - X^T Y_v + B_v(\lambda_1 R R^T + \lambda_2 I)]., \tag{9}$$

where I is a $q \times q$ identity matrix. There are two regularization coefficients (λ_1 and λ_2) to be determined using nested cross-validation. For description convenient, the gradient is expressed as

$$\nabla L(B_v) = 2 \times [X^T X B_v - X^T Y_v + B_v \hat{R}]., \tag{10}$$

where $\hat{R} = \lambda_1 R R^T + \lambda_2 I$. Setting this gradient to zero gives

$$X^T X B_v + B_v \hat{R} = X^T Y_v. \tag{11}$$

Note that this equation is different from the traditional equation of penalized least square regression, the unknown parameter B_v is in the left hand of the second term in this equation, which means the equation cannot be formed into a formation like $Ax = b$.

Actually, this is the Sylvester equation, with B_v the unknown parameter matrix to be determined. The equation can be solved efficiently (Bartels and Stewart 1972). Similar to the estimation of Ridge and Lasso regression, we used a nested threefold cross-validation to determine λ_1, λ_2 in the range $(10^{-5}, 10^5)$ on a log scale.

Prediction

Similar with the widely used searchlight strategy (Kriegeskorte et al. 2006) in brain mapping, we move a spherical searchlight through the brain volume. For each center voxel v, we can obtain the estimated parameter matrix $\hat{B}_v$ by solving Eq. (11). Thus the prediction response matrix $\hat{Y}_v$ is calculated as

$$\hat{Y}_v = X \hat{B}_v. \tag{12}$$

The brain activity prediction of voxel v and its neighbors are thus the columns of $\hat{Y}_v$. Note that in this strategy, the brain activity of voxel v will be predicted in several models, i.e., it will appear as the center voxel for one model and will also be as neighbor of other voxels for several other models. To obtain a smooth prediction, we set the response of voxel v as the mean of these responses.

Implementation details

In this work, we used the Gabor wavelet pyramid model (Jones and Palmer 1987) with six frequencies and eight possible orientations to extract stimulus features. To address the residual nonlinearity in the model, we applied an additional nonlinear transformation

$$f(x) = \log(1 + \sqrt{x}). \tag{13}$$

for each stimuli feature as done in previous studies (Kay et al. 2008). This resulted a $1 \times 10,920$ feature vector for each stimuli image. It is time consuming to optimize the regression models when X is so large; so, for computational reasons, we reduced the
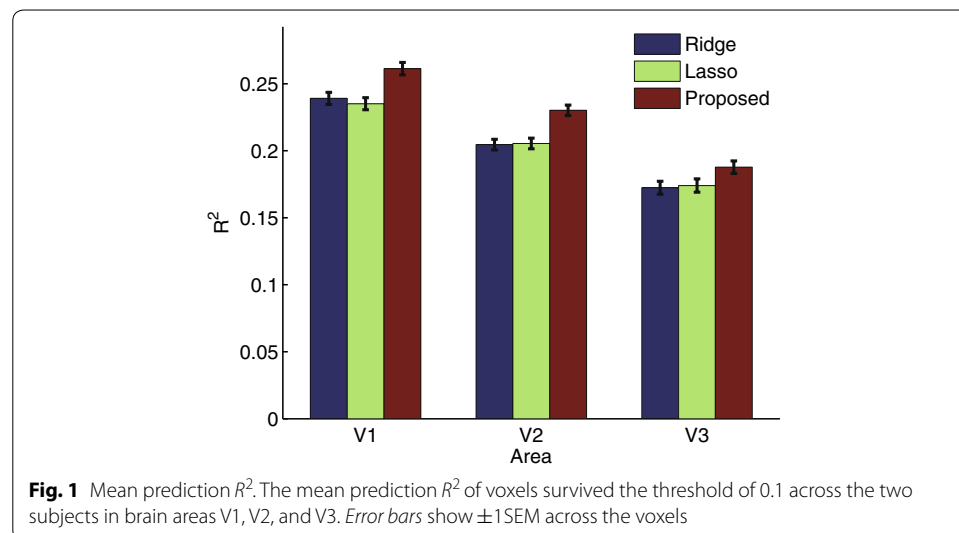
features by performing a principal component analysis (PCA) (Bishop 2006) first, which is a common strategy in machine learning field for dimension reduction. Only the largest 500 components were retained; these components capture over 80% of the variance, and so the transformed feature vector is 1×500 for each stimuli.
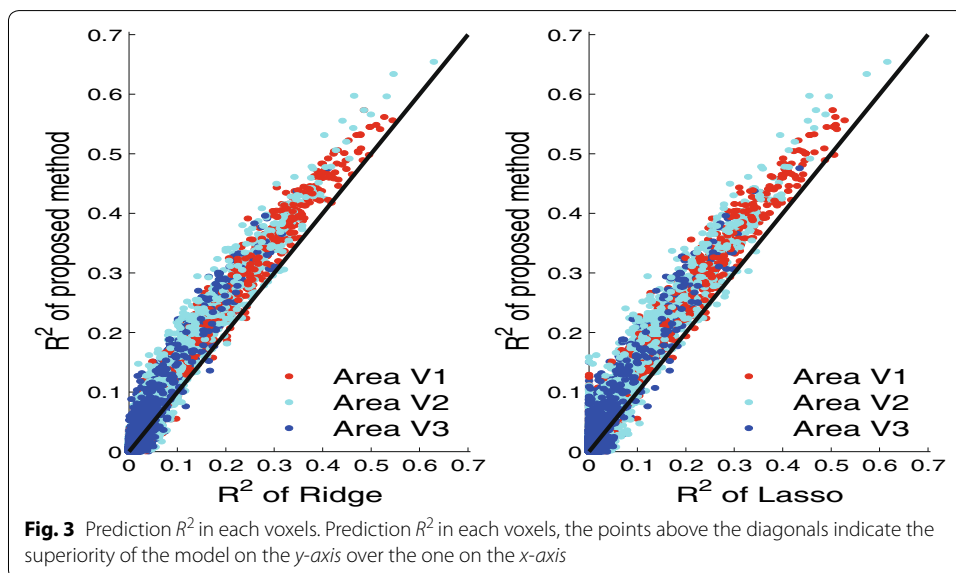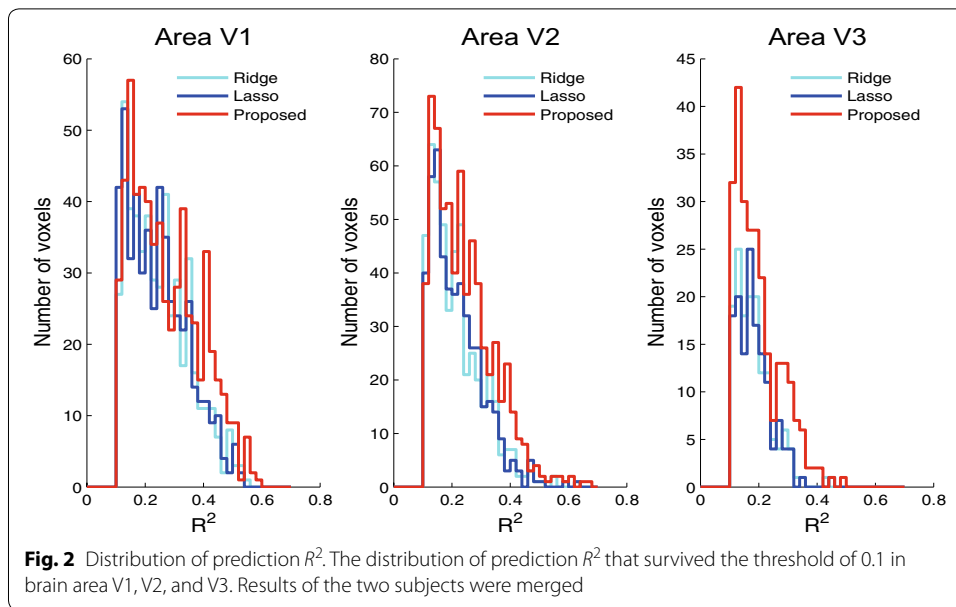
Results and discussion

Here, we present results obtained by different models on the dataset. We compare our proposed multi-target model with two state-of-the-art voxel-wise models (Ridge and Lasso); these two models were widely employed in fMRI encoding models (Agrawal et al. 2014; Kay et al. 2008; Schoenmakers et al. 2013; Güçlü and van Gerven 2014). Only data from training sessions were used to construct models and select regularization coefficients λ_1, λ_2; data from validation sessions were used to validate the model performances.

Table 1 lists how many voxels (in percentage) survived a R^2 threshold of 0.1 for different models in brain area V1, V2, and V3; these voxels are thought as activity well predicted. In all models, the performance in V1 is better than in V2 and V3. For subject 1, the percent of survived voxels systematically decreased from 29% in V1 to 10% in V3 when proposed method is used. While for the voxel-based models (ridge and lasso), the percent of survived voxels systematically decreased from about 25% in V1 to 6% in V3. Similar trend is observed for subject 2, though the performance is not as better as for subject 1.

Fig.1 compares the mean R^2 of different models across the survived voxels. The mean R^2 of the proposed method is about 0.26 in V1, and it systematically decreases to 0.19 in V3. In contrast, the mean R^2 of voxel-based ridge and lasso models are similar, systematically decreasing from 0.24 in V1 to 0.17 in V3.

Fig. 1 Mean prediction R^2. The mean prediction R^2 of voxels survived the threshold of 0.1 across the two subjects in brain areas V1, V2, and V3. *Error bars* show ± 1SEM across the voxels

Fig. 2 Distribution of prediction R^2. The distribution of prediction R^2 that survived the threshold of 0.1 in brain area V1, V2, and V3. Results of the two subjects were merged

Fig. 3 Prediction R^2 in each voxels. Prediction R^2 in each voxels, the points above the diagonals indicate the superiority of the model on the *y-axis* over the one on the *x-axis*

Figures 2 and 3 compare the performance of different models across voxels and brain areas. Figure 2 represents the distribution of prediction R^2 for survived voxels. In most values of R^2, the proposed method obtained more voxels than ridge and lasso models. The prediction R^2 for all voxels are displayed in Fig. 3, where the points above the diagonals indicate the superiority of the model on the y-axis over the one on the x-axis. Obviously, most voxels in each brain area are better predicted by the proposed model than the traditional voxel-wise models.

Table 1 Percentage of voxels survived an R^2 threshold of 0.1 for different models

Models	Subject 1			Subject 2		
	V1 (%)	V2 (%)	V3 (%)	V1 (%)	V2 (%)	V3 (%)
Ridge	25.50	16.61	6.09	12.29	7.78	2.20
Lasso	25.73	16.13	5.75	12.29	7.25	2.09
Proposed	29.44	21.84	10.22	16.30	11.11	4.29

Conclusions

In this paper, we proposed a multi-target regression model to predict brain activity when subjects view grayscale images. Based on the hypothesis that the property of a voxel is similar to its local neighbors, we constructed a spatial constraint on model parameters. The parameters can be estimated in an efficient way. We illustrated that the proposed method achieves better prediction performance on a public dataset fMRI data than voxel-wise ridge and lasso models did. The prediction R^2 of proposed model was higher than those acquired by voxel-wise models and more voxels survived an R^2 threshold of 0.1. These results suggest the benefits of considering essential spatial property of fMRI data in encoding models.

Author details
[1] Center for Brain Computer Interfaces and Brain Information Processing, South China University of Technology, Guangzhou 510640, China. [2] Guangzhou Key Laboratory of Brain Computer Interaction and Applications, South China University of Technology, Guangzhou 510640, China.

References
Agrawal P, Stansbury D, Malik J, Gallant JL (2014) Pixels to voxels: modeling visual representation in the human brain. arXiv preprint arXiv:1407.5104
Argyriou A, Evgeniou T, Pontil M (2008) Convex multi-task feature learning. Mach Learn 73(3):243–272
Bartels RH, Stewart G (1972) Solution of the matrix equation ax+ xb= c [f4]. Commun ACM 15(9):820–826
Bishop CM (2006) Pattern recognition and machine learning. Springer, New York
de Brecht M, Yamagishi N (2012) Combining sparseness and smoothness improves classification accuracy and interpretability. Neuroimage 60(2):1550–1561
Evgeniou T, Pontil M (2004) Regularized multi-task learning. In: Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining, ACM, New York, pp 109–117
Friston KJ, Holmes AP, Worsley KJ, Poline J, Frith CD, Frackowiak RS (1994) Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp 2(4):189–210
Güçlü U, van Gerven MA (2014) Unsupervised feature learning improves prediction of human brain activity in response to natural images. PLoS Comput Biol 10:e1003724

Haxby JV, Connolly AC, Guntupalli JS (2014) Decoding neural representational spaces using multivariate pattern analysis. Annu Rev Neurosci 37:435–456

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293(5539):2425–2430

Jones JP, Palmer LA (1987) An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. J Neurophysiol 58(6):1233–1258

Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. Nature 452(7185):352–355

Kay K, Naselaris T, Gallant J (2011) fmri of human visual areas in response to natural images. http://CRCNS.org/. Accessed 18 June 2015

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci USA 103(10):3863–3868

Michel V, Gramfort A, Varoquaux G, Eger E, Thirion B (2011) Total variation regularization for fmri-based prediction of behavior. IEEE Trans Med Imaging 30(7):1328–1340

Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fmri. Neuroimage 56(2):400–410

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian reconstruction of natural images from human brain activity. Neuron 63(6):902–915

Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fmri data. Trends Cogn Sci 10(9):424–430

Schoenmakers S, Barth M, Heskes T, van Gerven M (2013) Linear reconstruction of perceived images from human brain activity. Neuroimage 83:951–961

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc Ser B 58:267–288

Vu VQ, Ravikumar P, Naselaris T, Kay KN, Gallant JL, Yu B (2011) Encoding and decoding v1 fmri responses to natural images with sparse nonparametric models. Ann Appl Stat 5(2B):1159