Group Lasso for Functional Logistic Regression
Jessica Godwin, Nedret Billor
Department of Mathematics and Statistics, Auburn University

Motivation
Functional Magnetic Resonance Imaging (fMRI) collects information about activation from over 1,000,000 voxels in the brain. As this information is collected over time, the data are a good candidate for functional data analysis (fda). Can we use a variable selection technique to minimize the dimension of the data within the framework of classification?

Functional Data
Functional data are data that have been sampled discretely over a continuum, usually time. There is assumed to be an underlying curve describing the data. The curve can be estimated using a functional eigenequation:

\[ < \xi_i, \xi_j > := \sum_{m=1}^{M} \xi_i^m \xi_j^m dt \]

The functional logistic regression model is defined as follows:

\[ y_i = \pi_i + \varepsilon_i \]

We assume errors \( \varepsilon_i, \varepsilon_j \) are independent for all \( i \neq j \) and \( E[\varepsilon_i] = 0 \). The conditional distribution of \( Y_i | \mathbf{X_i}(t) \) is Bernoulli(\( \pi_i \)), with

\[ \pi_i(t) = E[Y_i | \mathbf{X_i}(t)] = \frac{\exp \left( \sum_{m=1}^{M} x_i^m(t) \beta_i^m \right)}{1 + \exp \left( \sum_{m=1}^{M} x_i^m(t) \beta_i^m \right)} \]

Making the logit transform, a generalized model is formed,

\[ l_i = \ln \left( \frac{\pi_i}{1 - \pi_i} \right) = \alpha + \sum_{m=1}^{M} x_i^m(t) \beta_i^m \]

We can express each \( l_i \) in (5) in terms of the PCs solving equation (1). In matrix notation,

\[ \mathbf{L} = \alpha \mathbf{1}_{M+1} + \sum_{m=1}^{M} \mathbf{A}^m \mathbf{V}^m \beta_i^m = \alpha \mathbf{1}_{M+1} + \sum_{m=1}^{M} \Gamma^m \mathbf{V}^m \beta_i^m \]

where \( \mathbf{L} = (l_1, ..., l_N)^T \), \( \beta_i^m = (\xi_i^1, ..., \xi_i^n)^T \), and \( \Gamma_i^m = (\xi_i^1)^T \mathbf{A}^m \mathbf{V}^m \) are the principal components of the design matrix, \( \mathbf{A}^m \) and \( \mathbf{V}^m \) is the matrix of corresponding eigenvectors. We choose a number of PCs, \( s \leq p \), based on the method of cumulative variance.

Simulation 1
We first simulated data in the manner of [3]. We simulated 3 functional predictors from three different basis systems and defined \( \beta_j(t) \) to be identically zero. The group lasso excluded the third predictor from the model in its first iteration, and went on to exclude predictor 2.

Simulation 2
Using the R package neuRoeim, we simulated an fMRI dataset [5]. We assigned classes by the two different effect sizes used in the experimental design.

Results
<table>
<thead>
<tr>
<th>Initial Variable Count</th>
<th>Final Variable Count</th>
<th>Initial Basis Count</th>
<th>Final Basis Count</th>
<th>Sensitivity</th>
<th>False Positive Rate</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sim 1</td>
<td>3</td>
<td>1</td>
<td>13</td>
<td>3</td>
<td>0.68</td>
<td>0.35</td>
</tr>
<tr>
<td>Sim 2</td>
<td>8000</td>
<td>4</td>
<td>43</td>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Application of Group Lasso
The principal component analysis allows for removal of redundant information on a univariate basis. There is also a need to select only those predictors which provide relevant information to the model. We will select a subset of variables with the use of the group lasso for logistic regression developed by Meier, et al. (2008).

- A hybrid \( L_1 \) and \( L_2 \) variable selection technique that will shrink entire groups of variables to zero
- Each of the \( M \) sets of basis coefficients is a group
- Exclusion from model: a set of basis coefficients is identically zero
- Inclusion in model: a set of basis coefficients is all nonzero

This is done via minimization of the following equation:

\[ S_{\lambda}(\beta) = -l(\beta) + \frac{1}{2} \sum_{m=1}^{M} s(df_m) ||\beta_i^m||_2 \]  

In (7), \( s(df_m) \) is simply a function of the degrees of freedom of each group. \( S(\beta) = df_1^2 \) is suggested.

Looking Ahead
We would like to acquire a real fMRI dataset for future assessment of our methodology. Our plans also include a simulation study in the near future.

References