\textbf{Introduction}

- use large set of BCI data to obtain a subject-independent classifier \cite{J. Schelldorfer and P. Buhlmann. Estimation for high-dimensional linear mixed-effects models using \ell_1-penalization. arXiv preprint 1002.3784.}
- novel statistical approach differentiates within-subject and between-subject variability \cite{José C. Pinheiro and Douglas M. Bates.}
- find a unifying model that (inherently) takes care of possible shifts in the input space

\textbf{1 Statistical Model}

\textbf{1.1 Model Setup}

Let $i = 1, \ldots, N$ be the number of subjects, $j = 1, \ldots, n_i$ the number of observations per subject and $N_j = \sum_j n_i$ the total number of observations. For each subject we observe an $n_i$-dimensional response vector $y_i$. Moreover, let $X_i$ and $Z_i$ be $n_i \times p$ and $n_i \times q$ covariate matrices, where $X_i$ contains the fixed-effects covariates and $Z_i$ the corresponding random-effects covariates. Denote by $h \in \mathbb{R}^p$ the $p$-dimensional fixed-effects vector and by $\beta_j, j = 1, \ldots, N$ the $q$-dimensional random-effects vectors. Then the linear mixed-effects model can be written as (1):

\begin{equation}
    y_i = X_i h + Z_i \beta_j^i + \epsilon_i, \quad i = 1, \ldots, N,
\end{equation}

where we assume that (i) $\beta_j \sim \mathcal{N}(0, \widehat{\Gamma})$, (ii) $c_j \sim \mathcal{N}(0, \sigma^2_c)$ and (iii) that the errors $\epsilon$ are mutually independent of the random effects $\beta_j$.

\textbf{1.2 Available Data and Experiments}

- 83 BCI datasets (45 EEG channels), each consisting of 150 trials ($t = 3s, f = 100Hz$)
- preprocess each dataset by 18 predefined temporal filters in parallel (see Figure 2)
- calculate a corresponding spatial filter and linear classifier for every band-pass filtered dataset to obtain a large number of subject-dependent BCI filters/classifiers (see Figure 1)
- process every dataset by this large set of basis functions
- perform an $\ell_1$-regularized logistic regression LMM (and classic $\ell_1$ logistic regression) on each classifier’s output to obtain an optimal combination of basis functions
- our method is validated by leave-one-subject-out cross-validation

Figure 3: selected features in white, inactive features in black. top: L1 logreg , bottom: LMM logreg, both at ideal L1 regularization constant

\begin{equation}
    g_{ij} = x_i^T h + \beta_i + \epsilon_{ij} \quad i = 1, \ldots, N, \quad j = 1, \ldots, n_i
\end{equation}

\textbf{2 Results}

Preliminary analysis of the data indicates that a so called random-intercept is appropriate for this data:

\begin{figure}
\centering
\includegraphics[width=\textwidth]{image3.png}
\caption{four figures show loss, averages over 83 subjects for L1-LSR, and LMM-LSR as well as L1-logistic regression and LMM logistic regression. two top figures are not bias corrected, while two lower ones are.}
\end{figure}

\textbf{3 Conclusion}

- solution is sparser, as compared to classical L1
- LMM helps in achieving lower overall error
- chosen features have a lower self-prediction error
- method is suitable for finding features, common to multiple subject data

\textbf{References}