Linear mixed effects models for zero-training BCI

by
Siamac Fazli, Juerg Schelldorfer
and
Klaus-Robert Müller
Contents

Introduction to Linear mixed-effects models (LMM)

Application to multiple-subject BCI data

Interpretation of the results
Let's assume that Juerg, Klaus and me like to run 100m sprints and each one of us runs \( n = 50 \) sprints a year.
Let's assume that Juerg, Klaus and me like to run 100m sprints and each one of us runs $n=50$ sprints a year.
Introduction to LMM

Lets assume that Juerg, Klaus and me like to run 100m sprints and each one of us runs $n=50$ sprints a year.

Explanatory variables (e.g. subject weight, hours of sleep, etc.) for each subject and every sprint are:

$$X_1, X_2, X_3 \in \mathbb{R}^{n \times p}$$

Response variable is the measured times:

$$Y_1, Y_2, Y_3 \in \mathbb{R}^n$$

$$< Y_1 >= 10 \pm 1 \text{s}$$
$$< Y_2 >= 12 \pm 1 \text{s}$$
$$< Y_3 >= 15 \pm 1 \text{s}$$
Introduction to LMM

Task: find (common) linear projection $b$ for all three subjects

$$Y = Xb + \epsilon$$

$$X = [X_1 X_2 X_3]$$

$$Y = [Y_1 Y_2 Y_3]$$

Problem: A requisite for linear regression is that the measurements are independent from each other. In our case the input signals can have a subject-dependent bias!

Solution: model the bias term as an extra parameter for every group:

$$Y_i = X_i b + \beta_i + \epsilon_i$$
Introduction to LMM

\[ Y_i = X_i b + \beta_i + \epsilon_i \]

- common \( b \) for all observations
- group-specific deviations \( \beta_i \)
- noise term \( \epsilon_i \)

\[ \beta_i \sim \mathcal{N}(0, \tau^2) \]
\[ \epsilon_i \sim \mathcal{N}(0, \sigma^2) \]

between-group variance \( \tau \)
within-group variance \( \sigma \)

assumption: \( \epsilon_i \) is independent of \( \beta_i \)
Generation of the data

ensemble generation

- 18 parallel temporal filters (predefined)
- 80 spatial filters per parallel filter (estimated)
- 80 classifiers per parallel filter (estimated)

$$18 \times 80 = 1400 \text{ classifiers in total}$$

dataset generation

- each dataset has 150 trials
- all trials (80 * 150) are fed into the ensemble
  (this is the group structure)

Thus our data set has **1400 features x 12450 trials**
Leave-one-subject-out-cross-validation

..of course we are not allowed to use all features, but must exclude those, which stem from the subject itself...
## Results - prediction

<table>
<thead>
<tr>
<th>training data</th>
<th>model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LMM</td>
<td>L1-LSR</td>
</tr>
<tr>
<td>10%</td>
<td>29.8</td>
<td>29.52</td>
</tr>
<tr>
<td>20%</td>
<td>29.24</td>
<td>29.3</td>
</tr>
<tr>
<td>30%</td>
<td>29.21</td>
<td>29.44</td>
</tr>
<tr>
<td>40%</td>
<td>28.91</td>
<td>28.81</td>
</tr>
</tbody>
</table>

LMM does not outperform our original approach, percentage-wise...
Results – model interpretation

• surprisingly the estimated between-group variance low $\hat{\tau} \approx 0.1$, as compared to the within-group variance is large $\hat{\sigma} \approx 0.9$

\[
\frac{\tau^2}{\tau^2 + \sigma^2} = 0.012
\]

• i.e. 98.8% of total variability is explained by within-group variance

$\beta_i$ has a very small effect => set $\beta_i = 0$ brings us back to LSR

-> it seems our original approach was viable
Thank you for your time
adding artificial spatial features

‘mexican hat’

simple laplacians
Introduction to LMM

Linear regression

\[ Y = Xb + \epsilon \]

In case inputs are grouped and not independent within groups...

\[ X_i = X_1, X_2, ... X_N \quad i = 1, ..., N \]

..one can consider a Linear mixed-effects model

\[ Y_i = X_ib + \beta_i + \epsilon_i \]