<table>
<thead>
<tr>
<th>Study</th>
<th>Data Type</th>
<th>Number of p-values</th>
<th>Contributors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeast</td>
<td>understanding basic life functions</td>
<td>11,904</td>
<td>Blomberg et al. 2003, 2010</td>
</tr>
<tr>
<td>Arabidopsis Thaliana</td>
<td>association mapping</td>
<td>3,745</td>
<td>Zhao et al. 2007</td>
</tr>
<tr>
<td>fMRI brain scans</td>
<td>function of brain language network</td>
<td>appr. 3 mill.</td>
<td>Taylor et al. 2006</td>
</tr>
</tbody>
</table>
Exercises: Problem 4.1
Example:

Factor 1: white, red, green, blue, coded as \{0,1,2,3\} gives covariate \(X^{(1)}\)

Factor 2: round, square, triangle, upside down triangle, coded as \{0,1,2,3\} gives covariate \(X^{(2)}\)

\(Y_i\) the distance when driver reads road sign

\[
Y_i = \mu + \sum_{k=0}^{3} \gamma_k 1 \{X^{(1)}_i = k\} + \sum_{k=0}^{3} \delta_k 1 \{X^{(2)}_i = k\} \\
+ \sum_{k,\ell=0}^{3} \kappa_{k,\ell} 1 \{X^{(1)}_i = k, X^{(2)}_i = \ell\} + \epsilon_i
\]
The model is of the form

\[ Y_i = \mu + \sum_{k=0}^{3} \gamma_k 1\{X_i^{(1)} = k\} + \sum_{k=0}^{3} \delta_k 1\{X_i^{(2)} = k\} \]
\[ + \sum_{k,\ell=0}^{3} \kappa_{k,\ell} 1\{X_i^{(1)} = k, X_i^{(2)} = \ell\} + \epsilon_i \]
\[ = \mu + \sum_{j=1}^{p} X_i^{(j)} \beta_j + \epsilon_i, \]

with \( \beta_1 = \gamma_0, \cdots \beta_4 = \gamma_3, \beta_5 = \delta_0, \cdots \) and (with somewhat conflicting notation), \( X_i^{(1)} = 1 \{X_i^{(1)} = 0\}, \cdots X_i^{(4)} = 1 \{X_i^{(1)} = 3\}, X_i^{(5)} = 1 \{X_i^{(2)} = 0\}, \cdots \)

\( Y_i = \) main effects 1 + main effects 2 + 2-way interactions

Often one would believe that either all main effects 1 parameters are nonzero, or all are zero, etc. This gives group structure which can be used to make the Lasso correspond to this belief – and additionally make computations faster
\[ Y_i = \mu + \sum_{k=0}^{3} \gamma_k 1 \left\{ X_i^{(1)} = k \right\} + \sum_{k=0}^{3} \delta_k 1 \left\{ X_i^{(2)} = k \right\} + \sum_{k, \ell=0}^{3} \kappa_{k, \ell} 1 \left\{ X_i^{(1)} = k, X_i^{(2)} = \ell \right\} + \epsilon_i \]

is overparametrized: to get an identifiable parametrization one assumes that
\[ \sum_{k=0}^{3} \gamma_k = 0, \sum \delta_k = 0, \text{ and } \sum_{k=0}^{3} \kappa_{k, \ell} = 0, \text{ all } \ell, \text{ and } \sum_{\ell=0}^{3} \kappa_{k, \ell} = 0, \text{ all } k. \]

all in all 24 "\( \beta \)-parameters", and 2+8 constraints, and thus 16 "free" parameters. Reparametrize to get a \( n \times 16 \) design matrix \( X \), orthonormal within groups, such that
\[ Y = X\beta + \epsilon \]

Homework: problem 4.1
B&vdG 4.2.1: Group Lasso

\( G_1, \ldots G_q \) partition of \( \{1, \ldots p\} \)

\[
\hat{\beta}(\lambda) = \arg\min_{\mu, \beta} \left( n^{-1} \sum_{i=1}^{n} \rho_{\beta}(x_i, y_i) + \lambda \sum_{j=1}^{q} m_j \left\| \beta_{G_j} \right\|_2 \right)
\]

\( m_j \) can be chosen freely, but typically one uses \( m_j = \sqrt{|G_j|} \), and

\[
\left\| \beta_{G_j} \right\|_2 = \sqrt{\sum_{k \in G_j} \beta_k^2}
\]

Ensures that either all parameters in a group is non-zero, or all are zero.
B&vdG  4.2.1: Group Lasso

$G_1, \ldots G_q$ partition of $\{1, \ldots p\}$

$$\hat{\beta}(\lambda) = \arg\min_{\mu, \beta} \left( n^{-1} \sum_{i=1}^{n} \rho_{\beta}(x_i, y_i) + \lambda \sum_{j=1}^{q} m_j \| \beta_{G_j} \|_2 \right)$$

$m_j$ can be chosen freely, but typically one uses $m_j = \sqrt{|G_j|}$, and

$$\| \beta_{G_j} \|_2 = \sqrt{\sum_{k \in G_j} \beta_k^2}$$

Ensures that either all parameters in a group is non-zero, or all are zero
$G_1$: color  $G_2$: shape  $G_3$: interaction

\[
X\beta = \begin{pmatrix}
1 & 2 & \ldots & \rho = 16 \\
1 & \times & \times & \times \\
\vdots & \vdots & \ddots & \times \\
1 & \times & \times & \times \\
\times & \times & \times & \times \\
\times & \times & \times & \times \\
\times & \times & \times & \times \\
\times & \times & \times & \times \\
\times & \times & \times & \times \\
\end{pmatrix}
\]

\[
\beta G_1 = \begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\end{pmatrix}
\]

and so on

\[
X'X = X'G_1\beta G_1 + X'G_2\beta G_2 + X'G_3\beta G_3
\]

Group Lasso: minimize $\|X - X\beta\|_2^2$ + penalty

penalty = $\lambda \left[ \sqrt{3} \sqrt{\beta_1^2 + \beta_2^2 + \beta_3^2} + \sqrt{3} \sqrt{\beta_4^2 + \beta_5^2 + \beta_6^2} \\
\vdots \right] + \sqrt{10} \sqrt{\beta_7^2 + \ldots + \beta_{16}^2}$
**B&vdG 4.3.1: Splice sites in DNA**

Worth reading. Logistic model, gives similar prediction skills as the "maximum entropy" machine learning method, but more understanding and a more understandable model.

**B&vdG 4.4: Asymptotics for group Lasso**

More complicated, but similar to the Lasso. Not prediction consistent if groups are too large. Useful also because it encourages sparsity in the selection of groups.

**B&vdG 4.6: Adaptive group Lasso**

"Same" as for adaptive Lasso, but rescales group penalties instead of penalties for individual parameters.
B&vdG 4.5: General penalty

Group Lasso penalty:
\[ \lambda \sum_{j=1}^{q} m_j \| \beta_{G_j} \|_2 = \lambda \sum_{j=1}^{q} m_j \sqrt{\beta_{G_j}^t \beta_{G_j}} \]

More general penalty:
\[ \lambda \sum_{j=1}^{q} m_j \sqrt{\beta_{G_j}^t A_j \beta_{G_j}} , \quad \text{with} \quad A_j = R_j^t R_j \]

\[ X_{G_j} \] = the matrix containing the group \( G_j \) columns in \( X \), \( \beta_{G_j} \) the parameters in group \( G_j \)

\[ X\beta = \sum_{j=1}^{q} X_{G_j} \beta_{G_j} \]
\[ A_j = R_j^t R_j \text{ with } R_j \text{ invertible} \]

\[
\tilde{X}_{G_j} = X_{G_j} R_j^{-1}, \quad \tilde{\beta}_{G_j} = R_j \beta_{G_j}
\]

\[
X\beta = \sum_{j=1}^{q} X_{G_j} \beta_{G_j} = \sum_{j=1}^{q} \tilde{X}_{G_j} \tilde{\beta}_{G_j}
\]

**Generalized group Lasso**

\[
\hat{\beta} = \arg\min_{\beta} (\|Y - X\beta\|_2^2 + \lambda \sum_{j=1}^{q} m_j \sqrt{\beta_{G_j}^t A_j \beta_{G_j}} )
\]

then same as

\[
\hat{\beta} = \arg\min_{\tilde{\beta}} (\|Y - \sum_{j=1}^{q} \tilde{X}_{G_j} \tilde{\beta}_{G_j}\|_2^2 + \lambda \sum_{j=1}^{q} m_j \|\tilde{\beta}_{G_j}\|_2 )
\]

\[
\hat{\beta}_{G_j} = R_j^{-1} \tilde{\beta}_{G_j}
\]
B&vdG 4.5.1: Prediction invariance

Assume $X_{Gj}^t X_{Gj}$ is invertible, choose $A_j = X_{Gj}^t X_{Gj}$. Makes

$$
\|Y - X\beta\|_2^2 + \lambda \sum_{j=1}^q m_j \sqrt{\beta_{Gj}^t A_j \beta_{Gj}}
$$

$$
= \|Y - \sum_{j=1}^q X_{Gj} \beta_{Gj}\|_2^2 + \lambda \sum_{j=1}^q m_j \sqrt{\beta_{Gj}^t X_{Gj}^t X_{Gj} \beta_{Gj}}
$$

invariant under 1-1 reparametrizations within groups: if $\beta_{Gj}$ is replaced by $B_j \beta_{Gj}$ and $X_{Gj}$ is replaced by $X_{Gj} B_j^{-1}$ for an invertible matrix $B_j$ then the expression above doesn’t change, and in particular the estimated set of active groups doesn’t change.