The Women’s Health and Aging Study

[With Karen Bandeen-Roche]

The Women’s Health and Aging Study (WHAS) began in 1992 to study the causes and the course of disability in moderately to severely disabled older women living in the community.

The WHAS is a population-based longitudinal study of women with at least mild disability, 65 years of age or older, living at home in eastern Baltimore city or county.

There is evidence that disability results from chronic diseases, and that interactions between diseases (comorbidities) are of importance in causing disability.

In this presentation we are concerned about relating chronic diseases and their interactions to death.
Study subjects:

- 32538 women were identified by searching medicare enrollment files,
- 6521 women were sampled (age-stratified),
- 5316 women were alive and living at home,
- 4137 women participated in the home-based screening,
- 1409 women were eligible,
- 1002 women agreed to participate and provided written informed consent.

The major chronic diseases at baseline were ascertained by using complex algorithms. Follow-up evaluations were conducted every 6 months for 3 years.
The Women’s Health and Aging Study

\[ p = \Pr(\text{death in round } j \mid \text{survival to round } j-1, X, \text{age}) \]

\[ \text{logit}(p) = -9.01 + 0.06 \cdot \text{age} + 1.07 \cdot L(X) \]

or

or

and

cancer

diab

stroke

angina

chf
Comparison to Decision Trees

A Decision Tree (CART) is something different!

Logic Regression

[With Charles Kooperberg and Michael LeBlanc]

$X_1, \ldots, X_k$ are 0/1 (False/True) predictors.

$Y$ is a response variable.

Fit a model $g(E(Y)) = b_0 + \sum_{j=1}^{t} b_j \cdot L_j$, where $L_j$ is a Boolean combination of the covariates, e.g. $L_j = (X_1 \lor X_2) \land X_4^c$.

Determine the logic terms $L_j$ and estimate the $b_j$ simultaneously.
The Move Set for Logic Regression

Possible Moves

- **Alternate Leaf**

  ![Diagram](image)

- **Alternate Operator**

  ![Diagram](image)

- **Grow Branch**

  ![Diagram](image)

- **Initial Tree**

  ![Diagram](image)

- **Prune Branch**

  ![Diagram](image)

- **Split Leaf**

  ![Diagram](image)

- **Delete Leaf**

  ![Diagram](image)

Simulated Annealing for Logic Regression

We try to fit the model \( g(E(Y)) = b_0 + \sum_{j=1}^{t} b_j \cdot L_j \).

- Select a scoring function (RSS, log-likelihood, \ldots).
- Pick the maximum number of Logic Trees.
- Pick the maximum number of leaves in a tree.
- Initialize the model with \( L_j = 0 \) for all \( j \).
- Carry out the Simulated Annealing Algorithm:
  - Propose a move.
  - Accept or reject the move, depending on the scores and the temperature.
and so on...
\[
X \quad T \quad Y \quad \pi(Y)
\]

Permutation

Histograms for each category of X and Y.
Another Example

The Cardiovascular Health Study
(Fried et. al, Annals of Epidemiology, 1991).

- The Cardiovascular Health Study (CHS) is a study of coronary heart disease and stroke in elderly people.
- Between 1989 and 1993, 5888 subjects over the age of 65 were recruited in four communities in the United States.
- During 1992 and 1994, a subset of these patients underwent an MRI scan.
- For 3647 CHS participants, MRI detected strokes (infarcts bigger than 3mm that led to deficits in functioning) were recorded as entries into a 23 region atlas of the brain.
- The mini-mental state examination is a brief screening test for dementia. The response $Y$ is a variable derived by transforming the mini-mental score.

We investigated models of the form $Y = \beta_0 + \beta_1 \times L_1 + \cdots + \beta_p \times L_p + \epsilon$.

Cross-Validation Cont.
The model we found was $Y = 1.96 + 0.36 \times L$, with the following Logic Tree:

```
<table>
<thead>
<tr>
<th>12</th>
<th>or</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>or</td>
</tr>
<tr>
<td></td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>or</td>
</tr>
<tr>
<td>19</td>
<td>4</td>
</tr>
</tbody>
</table>
```

**Logic Model vs Linear Model vs MARS**

**Linear model:**

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\beta}$</th>
<th>s. e.</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.961</td>
<td>0.015</td>
<td>133.98</td>
</tr>
<tr>
<td>Region 4</td>
<td>0.524</td>
<td>0.129</td>
<td>4.06</td>
</tr>
<tr>
<td>Region 12</td>
<td>0.460</td>
<td>0.112</td>
<td>4.09</td>
</tr>
<tr>
<td>Region 17</td>
<td>0.236</td>
<td>0.057</td>
<td>4.17</td>
</tr>
<tr>
<td>Region 19</td>
<td>0.611</td>
<td>0.157</td>
<td>3.89</td>
</tr>
</tbody>
</table>

**Logic model:**

$$Y = 1.96 + 0.36 \times \mathbb{I}_{\{X_4 \land X_{12} \land X_{17} \land X_{19} \text{ is true}\}}$$

**MARS:**

$$Y = 1.96 + 0.53 \times X_4 + 0.37 \times X_{12} + 0.24 \times X_{17} + 0.61 \times X_{19} + 1.05 \times (X_{12} \times X_{15})$$
Logic Model vs CART

Logic Tree

CART Tree

References


Software and manuscripts available at: [http://biostat.jhsph.edu/~iruczins/](http://biostat.jhsph.edu/~iruczins/)