

4

Modelo de Regressão Logística com Transição Suave Estruturado por Árvore (STLR-Tree)

Na busca por uma melhor explicação de fenômenos complexos por modelos de regressão e de séries temporais, a utilização da modelagem não-linear vem crescendo ao longo dos anos amparada pelo avanço de recursos computacionais e a modernização dos pacotes estatísticos.

A modelagem não-linear vem para superar os métodos lineares de previsão de maior recorrência, indo além dos modelos estacionários Gaussianos e abrangendo as situações onde os dados apresentam um comportamento que não equivale ao linear tendo algumas características, citadas por (8), por exemplo: não-normalidade, ciclos assimétricos, bimodalidade, não-normalidade entre variáveis defasadas, variação do desempenho de previsão sobre o espaço de estado, irreversibilidade temporal etc. Podemos acrescentar a isso outras características como: ciclos-limite, salto de ressonância, amplitude dependente da frequência e caos. Estes últimos inseridos no contexto da análise de séries temporais. No Apêndice A encontram-se alguns dos modelos não-lineares, em sua maioria utilizados na análise de séries temporais, que, de alguma forma, estão relacionados com o STLR-Tree, seja em sua forma estrutura, estimação e/ou previsão.

A idéia central deste trabalho é adaptar o modelo proposto em (7), usando-o como um método de classificação, para o caso em que nossa variável dependente, y_i , assuma apenas dois valores, 0 ou 1, caindo assim no contexto de uma Regressão Logística.

4.1

Revisão do STR-Tree

Proposto em (7) o modelo STR-Tree tem como idéia principal a substituição das transições abruptas nas árvores de regressão feitas através da função indicadora $I(\mathbf{x}_i; c)$ pela função logística definida por

$$G(x_i; \gamma, c) = \frac{1}{1 + e^{-\gamma(x_i - c)}}. \quad (4-1)$$

Definição 4.1 *Seja $\mathbf{z}_i \subseteq \mathbf{x}_i$ tal que $\mathbf{x}_i \in \mathbb{X} \subseteq \mathbb{R}^q$ e $\mathbf{z}_i \in \mathbb{R}^p$ onde $p \leq q$. Considere $\tilde{\mathbf{z}}_i = (1, \mathbf{z}_i)'$. Um modelo paramétrico \mathcal{M} definido pela função $H_{\mathbb{T}}(\mathbf{x}_i; \psi) : \mathbb{R}^{q+1} \rightarrow \mathbb{R}$, indexado pelo vetor de parâmetros $\psi \in \Psi$, um subconjunto compacto do espaço Euclidiano, é chamado modelo Smooth Transition Regression Tree (STR-Tree) se*

$$y_i = H_{\mathbb{T}}(\mathbf{x}_i; \psi) = \sum_{k \in \mathbb{T}} \beta'_k \tilde{\mathbf{z}}_i B_{\mathbb{J}_k}(\mathbf{x}_i; \boldsymbol{\theta}_k) + \epsilon_i \quad (4-2)$$

onde

$$B_{\mathbb{J}_k}(\mathbf{x}_i; \boldsymbol{\theta}_k) = \prod_{j \in \mathbb{J}} G(x_{s_j, i}; \gamma_j, c_j)^{\frac{n_{k,j}(1+n_{k,j})}{2}} [1 - G(x_{s_j, i}; \gamma_j, c_j)]^{(1-n_{k,j})(1+n_{k,j})},$$

e $n_{i,j}$ assume três valores:

- -1 , se o caminho até o nó i não incluir o nó gerador j ;
- 0 , se caminho para o nó i incluir o nó filho da direita do nó gerador j ;
- 1 , se o caminho para o nó terminal i incluir o nó filho da esquerda do nó gerador j .

Sendo \mathbb{J}_k o subconjunto de \mathbb{J} que contém os índices dos nós geradores (ou nós pais) do caminho para o nó terminal k . Então, $\boldsymbol{\theta}_k$ é o vetor que contém todos os parâmetros não-lineares (γ_t, c_t) tal que $t \in \mathbb{J}_k, k \in \mathbb{T}$ (índice dos nós terminais).

As funções $B_{\mathbb{J}_k}$ são tais que, $0 < B_{\mathbb{J}_k} < 1$ e $\sum_{j \in \mathbb{J}} B_{\mathbb{J}_k}(\mathbf{x}_i; \boldsymbol{\theta}_j) = 1, \forall \mathbf{x}_i \in \mathbb{R}^{q+1}$.

Vamos considerar agora um STR-Tree em uma árvore completamente crescida, ou cheio, com profundidade d , $K = 2d$, nós terminais (folhas) e $N = \sum_{k=1}^d 2^k$ nós geradores definido como

$$y_i = \sum_{k=1}^K \beta'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k) + \epsilon_i, \quad (4-3)$$

onde $y_i \in \mathbb{R}$, $\mathbf{x}_i = (x_{1i}, \dots, x_{qi})' \in \mathbb{X} \subseteq \mathbb{R}^q$ e considerando que o vetor \mathbf{z}_i receba valores defasados das variáveis explicativas, temos que $\mathbf{z}_i = (x_{1,i}, \dots, x_{1,i-p_1}, \dots, x_{k,i}, \dots, x_{k,i-p_k})' \in \mathbb{R}^m$, onde $m = p + \sum_{j=1}^k (p_j + 1)$, sem esquecer que $\tilde{\mathbf{z}}_i = (1, \mathbf{z}_i)'$.

O vetor de parâmetros $\boldsymbol{\psi} = (\beta'_{K-1}, \dots, \beta'_{2K-2}, \boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_K)' \in \mathbb{R}^r$ possui $r = (p+1)K + 2N$.

Dois exemplos de estruturas de árvore são dados a seguir. Primeiramente o caso mais simples, figura 4.1, com uma profundidade, ou seja, quando temos o nó raiz se dividindo em apenas dois nós terminais, assim $d = 1, K = 2$ e $N = 1$. Este modelo sofrerá uma reparametrização para que possamos, mais adiante, fazer as hipóteses em relação a divisão dos nós. O segundo exemplo, figura 4.2, é uma estrutura de árvore maior com profundidade $d = 2$, quatro nós terminais, $K = 4$, e $N = 3$. Abaixo se encontram as equações de cada um deles bem como suas figuras ilustrativas.

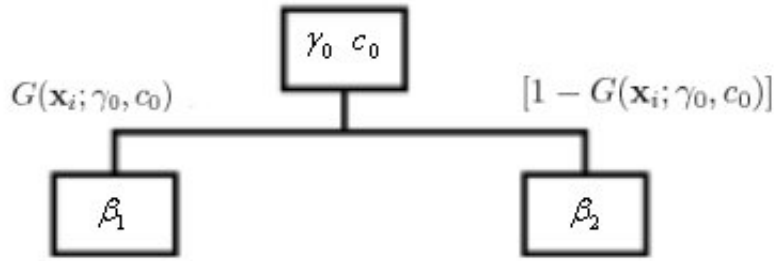


Figura 4.1: Exemplo Árvore 1

$$y_i = \beta_1' \tilde{\mathbf{z}}_i G(\mathbf{x}_i; \gamma_0, c_0) + \beta_2' \tilde{\mathbf{z}}_i [1 - G(\mathbf{x}_i; \gamma_0, c_0)] + \epsilon_i$$

Reparametrizando a fim de obter uma representação mais parcimoniosa,

$$y_i = \phi_0' \tilde{\mathbf{z}}_i + \lambda_0' \tilde{\mathbf{z}}_i G(\mathbf{x}_i; \gamma_0, c_0) + \epsilon_i \quad (4-4)$$

onde $\phi_0 = \beta_2$ e $\lambda_0 = \beta_1 - \beta_2$

O outro exemplo é

$$y_i = \{\beta_3' \tilde{\mathbf{z}}_i G(\mathbf{x}_i; \gamma_1, c_1) + \beta_4' \tilde{\mathbf{z}}_i [1 - G(\mathbf{x}_i; \gamma_1, c_1)]\} G(\mathbf{x}_i; \gamma_0, c_0) + \{\beta_5' \tilde{\mathbf{z}}_i G(\mathbf{x}_i; \gamma_2, c_2) + \beta_6' \tilde{\mathbf{z}}_i [1 - G(\mathbf{x}_i; \gamma_2, c_2)]\} [1 - G(\mathbf{x}_i; \gamma_0, c_0)] + \epsilon_i$$

$$y_i = \beta_3' \tilde{\mathbf{z}}_i G(\mathbf{x}_i; \gamma_0, c_0) G(\mathbf{x}_i; \gamma_1, c_1) + \beta_4' \tilde{\mathbf{z}}_i G(\mathbf{x}_i; \gamma_0, c_0) [1 - G(\mathbf{x}_i; \gamma_1, c_1)] + \beta_5' \tilde{\mathbf{z}}_i [1 - G(\mathbf{x}_i; \gamma_0, c_0)] G(\mathbf{x}_i; \gamma_2, c_2) + \beta_6' \tilde{\mathbf{z}}_i [1 - G(\mathbf{x}_i; \gamma_0, c_0)] [1 - G(\mathbf{x}_i; \gamma_2, c_2)] + \epsilon_i$$

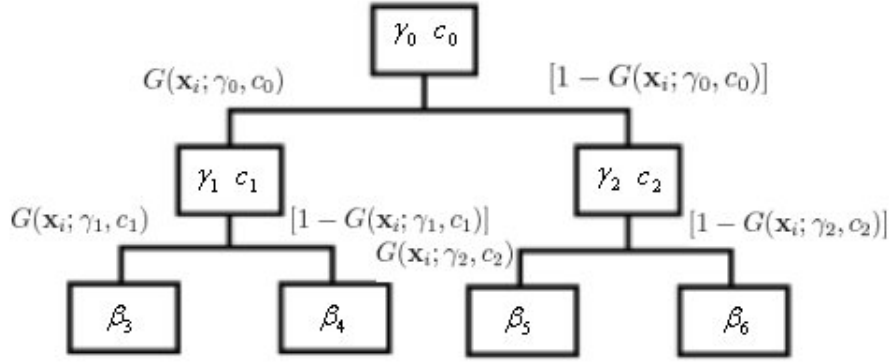


Figura 4.2: Exemplo Árvore 2

com isso podemos deduzir que

$$\begin{aligned}
 B_1(\mathbf{x}_i; \boldsymbol{\theta}_1) &= G(\mathbf{x}_i; \gamma_0, c_0)G(\mathbf{x}_i; \gamma_1, c_1) \\
 B_2(\mathbf{x}_i; \boldsymbol{\theta}_2) &= G(\mathbf{x}_i; \gamma_0, c_0)[1 - G(\mathbf{x}_i; \gamma_1, c_1)] \\
 B_3(\mathbf{x}_i; \boldsymbol{\theta}_3) &= [1 - G(\mathbf{x}_i; \gamma_0, c_0)]G(\mathbf{x}_i; \gamma_2, c_2) \\
 B_4(\mathbf{x}_i; \boldsymbol{\theta}_4) &= [1 - G(\mathbf{x}_i; \gamma_0, c_0)][1 - G(\mathbf{x}_i; \gamma_2, c_2)]
 \end{aligned}$$

assim

$$\begin{aligned}
 y_i &= \boldsymbol{\beta}'_3 \tilde{\mathbf{z}}_i B_1(\mathbf{x}_i; \boldsymbol{\theta}_1) + \boldsymbol{\beta}'_4 \tilde{\mathbf{z}}_i B_2(\mathbf{x}_i; \boldsymbol{\theta}_2) \\
 &\quad + \boldsymbol{\beta}'_5 \tilde{\mathbf{z}}_i B_3(\mathbf{x}_i; \boldsymbol{\theta}_3) + \boldsymbol{\beta}'_6 \tilde{\mathbf{z}}_i B_4(\mathbf{x}_i; \boldsymbol{\theta}_4) + \epsilon_i
 \end{aligned}$$

$$y_i = \sum_{k=1}^4 \boldsymbol{\beta}'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k) + \epsilon_i.$$

4.2 Especificação do STLR-Tree

Ao especificar o modelo devemos primeiramente selecionar as variáveis relevantes, elementos de \mathbf{z}_i . A segunda etapa na especificação do modelo é a busca pelo nó a ser dividido e, na sequência, a escolha da a variável de transição, elementos de \mathbf{x}_i .

Antes de descrever as etapas de especificação vamos reescrever o modelo STR-Tree levando em consideração que nossa variável dependente seja dicotômica, onde cada elemento tenha uma distribuição Bernoulli de parâmetro

π , e que tanto \mathbf{z}_i quanto \mathbf{x}_i sejam variáveis contínuas, independentes entre si, respeitando que $\mathbf{z}_i \subseteq \mathbf{x}_i$. Além disso, vamos considerar o modelo em uma árvore completamente crescida, com profundidade d e $K = 2d$, nós terminais. Desta maneira, podemos escrever o modelo de modo similar a uma Regressão Logística, utilizando como função de ligação a logito, denominado STLR-Tree, como se segue

$$\log \left[\frac{\pi(\tilde{\mathbf{z}}_i)}{1 - \pi(\tilde{\mathbf{z}}_i)} \right] = \sum_{k=1}^K \boldsymbol{\beta}'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k), \quad (4-5)$$

como descrito anteriormente, após algumas contas, encontra-se o valor de $\pi(\tilde{\mathbf{z}}_i)$ igual a

$$\pi(\tilde{\mathbf{z}}_i) = \frac{e^{\sum_{k=1}^K \boldsymbol{\beta}'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k)}}{1 + e^{\sum_{k=1}^K \boldsymbol{\beta}'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k)}}. \quad (4-6)$$

Por fim, temos a função de log-verossimilhança do STLR-Tree, semelhante à encontrada em (2-14), que será utilizada em seções posteriores.

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n \left[y_i \sum_{k=1}^K \boldsymbol{\beta}'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k) - \log \left(1 + e^{\sum_{k=1}^K \boldsymbol{\beta}'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k)} \right) \right]$$

4.2.1

Escolha das variáveis relevantes

Para a escolha dos elementos de \mathbf{z}_i destacam-se três métodos: por critérios de informação, AIC e BIC, os quais já foram descritos em (2-12) e (2-13) respectivamente, sendo o melhor modelo aquele que minimiza tais critérios; aproximação polinomial do modelo, proposto em (26); outra opção é dada através de técnicas não paramétricas, porém esta classe é computacionalmente dispendiosa principalmente para um grande número de observações.

No caso o STLR-Tree será restrito para variáveis contínuas independentes. Se fossem consideradas variáveis categóricas, a extensão do modelo que contempla tal mudança é feita através da inclusão de um vetor contendo variáveis indicadoras, $\mathbf{D}_i(\mathbf{w}_i)$, que representa o vetor categórico denominado

por \mathbf{w}_i . Assim o modelo em (4-5) terá a forma:

$$\log \left[\frac{\pi(\tilde{\mathbf{z}}_i)}{1 - \pi(\tilde{\mathbf{z}}_i)} \right] = \sum_{k=1}^K \beta'_{K+k-2} \tilde{\mathbf{z}}_i \mathbf{D}_i(\mathbf{w}_i) B_k(\mathbf{x}_i; \boldsymbol{\theta}_k), \quad (4-7)$$

4.2.2

Escolha do nó a ser dividido

Estudo da significância da divisão dos nós através de testes de hipótese, fundamentados em inferência estatística. A descrição a seguir é similar ao procedimento de estimação do modelo STAR seguindo a abordagem *específico-geral* partindo de um modelo simples para um modelo mais complexo à medida que testes de diagnóstico assim o permita. Maiores detalhes do ciclo de modelagem STAR em (32).

Estamos testando a linearidade do modelo, o que implica dizer se devemos ou não dividir um nó.

Neste ponto iremos voltar a escrever o modelo no caso clássico, considerando que $\epsilon_i \sim NID(0, \sigma^2)$, para justificar algumas manipulações algébricas entendidas com mais facilidade nesta situação. Assim, considerando K nós terminais e testando se o nó $k^* \in \mathbb{T}$ será dividido temos o modelo escrito como:

$$y_i = \sum_{k \in \mathbb{T} - \{k^*\}} \beta'_k \tilde{\mathbf{z}}_i B_{\mathbb{J}_k}(\mathbf{x}_i; \boldsymbol{\theta}_k) + \beta'_{2k^*+1} \tilde{\mathbf{z}}_i B_{\mathbb{J}_{2k^*+1}}(\mathbf{x}_i; \boldsymbol{\theta}_{2k^*+1}) + \beta'_{2k^*+2} \tilde{\mathbf{z}}_i B_{\mathbb{J}_{2k^*+2}}(\mathbf{x}_i; \boldsymbol{\theta}_{2k^*+2}) + \epsilon_i$$

onde

$$B_{\mathbb{J}_{2k^*+1}}(\mathbf{x}_i; \boldsymbol{\theta}_{2k^*+1}) = B_{\mathbb{J}_{k^*}}(\mathbf{x}_i; \boldsymbol{\theta}_{k^*}) G(x_{k^*i}; \gamma_{k^*}, c_{k^*})$$

$$B_{\mathbb{J}_{2k^*+2}}(\mathbf{x}_i; \boldsymbol{\theta}_{2k^*+2}) = B_{\mathbb{J}_{k^*}}(\mathbf{x}_i; \boldsymbol{\theta}_{k^*}) [1 - G(x_{k^*i}; \gamma_{k^*}, c_{k^*})]$$

Como feito em (4-4) reparametrizamos o modelo de tal forma que

$$y_i = \sum_{k \in \mathbb{T} - \{k^*\}} \beta'_k \tilde{\mathbf{z}}_i B_{\mathbb{J}_k}(\mathbf{x}_i; \boldsymbol{\theta}_k) + \phi' \tilde{\mathbf{z}}_i B_{\mathbb{J}_{k^*}}(\mathbf{x}_i; \boldsymbol{\theta}_{k^*}) + \lambda' \tilde{\mathbf{z}}_i B_{\mathbb{J}_{k^*}}(\mathbf{x}_i; \boldsymbol{\theta}_{k^*}) G(x_{k^*i}; \gamma_{k^*}, c_{k^*}) + \epsilon_i$$

onde $\phi = \beta_{2k^*+1}$ e $\lambda = \beta_{2k^*+1} - \beta_{2k^*+2}$.

Deve-se testar a hipótese de significância dessa divisão (inclusão de um

novo regime), que é equivalente a testar as hipóteses:

$$\begin{cases} H_0 : \gamma_{k^*} = 0 \\ H_1 : \gamma_{k^*} > 0. \end{cases} \quad (4-8)$$

Porém, sob H_0 temos que enfrentar um problema de especificação do modelo, pois os parâmetros γ_{k^*} e c_{k^*} podem assumir diferentes valores sem alterar a função de verossimilhança. Para solucionar tal problema foi proposto em (18) uma aproximação da função de transição por uma expansão de Taylor de terceira ordem em torno de $\gamma_{k^*} = 0$ e assim, após manipulações algébricas, podemos reescrever o modelo como

$$\begin{aligned} y_i = & \sum_{k \in \mathbb{T} - \{k^*\}} \beta'_k \tilde{\mathbf{z}}_i B_{\mathbb{J}_k}(\mathbf{x}_i; \boldsymbol{\theta}_k) + \alpha'_0 \tilde{\mathbf{z}}_i B_{\mathbb{J}_{k^*}}(\mathbf{x}_i; \boldsymbol{\theta}_{k^*}) + \\ & + \alpha'_1 \tilde{\mathbf{z}}_i B_{\mathbb{J}_{k^*}}(\mathbf{x}_i; \boldsymbol{\theta}_{k^*}) + \alpha'_2 \tilde{\mathbf{z}}_i B_{\mathbb{J}_{k^*}}(\mathbf{x}_i; \boldsymbol{\theta}_{k^*}) + \alpha'_3 \tilde{\mathbf{z}}_i B_{\mathbb{J}_{k^*}}(\mathbf{x}_i; \boldsymbol{\theta}_{k^*}) + e_i \end{aligned}$$

onde $e_i = \epsilon_i + \boldsymbol{\lambda}' \tilde{\mathbf{z}}_i R(x_{k^*i}; \gamma_{k^*}, c_{k^*})$ e $R(x_{k^*i}; \gamma_{k^*}, c_{k^*})$ é o termo restante da expansão de Taylor.

Assim podemos reescrever a hipótese de nulidade dos parâmetros como:

$$H_0 : \alpha_i = 0, \quad i = 1, 2 \text{ e } 3$$

e sob H_0 temos que $e_i = \epsilon_i$.

Com isso, pegando o último modelo escrito anteriormente e considerando novamente que a variável dependente é binária, voltamos para o modelo em que

$$\begin{aligned} \log \left[\frac{\pi(\tilde{\mathbf{z}}_i)}{1 - \pi(\tilde{\mathbf{z}}_i)} \right] = & \sum_{k \in \mathbb{T} - \{k^*\}} \beta'_k \tilde{\mathbf{z}}_i B_{\mathbb{J}_k}(\mathbf{x}_i; \boldsymbol{\theta}_k) + \alpha'_0 \tilde{\mathbf{z}}_i B_{\mathbb{J}_{k^*}}(\mathbf{x}_i; \boldsymbol{\theta}_{k^*}) + \\ & + \alpha'_1 \tilde{\mathbf{z}}_i B_{\mathbb{J}_{k^*}}(\mathbf{x}_i; \boldsymbol{\theta}_{k^*}) + \alpha'_2 \tilde{\mathbf{z}}_i B_{\mathbb{J}_{k^*}}(\mathbf{x}_i; \boldsymbol{\theta}_{k^*}) + \alpha'_3 \tilde{\mathbf{z}}_i B_{\mathbb{J}_{k^*}}(\mathbf{x}_i; \boldsymbol{\theta}_{k^*}). \end{aligned}$$

Em seguida é feita uma seqüência de teste de *Razão de Verossimilhança* comparando o modelo sob H_0 contra o modelo irrestrito através da estatística de teste, que foram mostradas em (2-14) e (2-15). A primeira compara os valores do logaritmo da função de verossimilhança maximizada e a segunda está em termos da diferença entre as *deviances* dos modelos conforme mostrado a seguir

$$\xi_{RV} = 2 \left[l(\hat{\boldsymbol{\beta}}; \mathbf{y}) - l(\boldsymbol{\beta}^0; \mathbf{y}) \right] = \phi^{-1} \left[D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - D(\mathbf{y}; \hat{\boldsymbol{\mu}}) \right] \xrightarrow{a} \chi_{3(p+1)}^2, \quad (4-9)$$

em que $\hat{\boldsymbol{\mu}}^0 = \mathbf{g}^{-1}(\hat{\boldsymbol{\eta}})^0$ e $\hat{\boldsymbol{\eta}}^0 = \mathbf{X}\boldsymbol{\beta}^0$.

Se desconhecido o parâmetro de dispersão ϕ deve ser substituído por uma estimativa consistente, $\hat{\phi}$. Contudo para o caso binomial temos que $\phi = 1$, desta forma

$$\xi_{RV} = [D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - D(\mathbf{y}; \hat{\boldsymbol{\mu}})] \xrightarrow{a} \chi_{3(p+1)}^2, \quad (4-10)$$

A inferência também pode ser baseada na estatística F da seguinte maneira:

$$F = \frac{[D(\mathbf{y}; \hat{\boldsymbol{\mu}}^0) - D(\mathbf{y}; \hat{\boldsymbol{\mu}})]}{3(p+1)} \xrightarrow{a} F_{3(p+1), N-4(p+1)}. \quad (4-11)$$

A idéia de utilizar a razão entre *deviances* para determinar o crescimento da árvore também é utilizada em (5) e (14).

4.2.3

Escolha das variáveis de transição

Aplicar os testes de RV para cada uma das variáveis explicativas e selecionar a variável x_{s_0t} que gere o menor *p-valor*, sob um nível de significância α . Sabe-se que $s_0 \in \mathbb{S} = \{1, 2, \dots, m\}$, o conjunto dos índices dos elementos em \mathbf{x}_i .

4.3

Estimação do STLR-Tree

A estimação da parte linear é feita por Máxima Verossimilhança, utilizando o método iterativo de Newton-Raphson. Para isso é necessário o cálculo de $\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ e $\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n \left[y_i \sum_{k=1}^K \boldsymbol{\beta}'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k) - \log \left(1 + e^{\sum_{k=1}^K \boldsymbol{\beta}'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k)} \right) \right]$$

$$\begin{aligned} \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = & \sum_{t=1}^n \left\{ \left[y_i \sum_{k=1}^K \boldsymbol{\beta}'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k) \right] - \right. \\ & \left. - \left[\frac{e^{\sum_{k=1}^K \boldsymbol{\beta}'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k)}}{\left(1 + e^{\sum_{k=1}^K \boldsymbol{\beta}'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k)} \right)} \right] \times \right. \end{aligned}$$

$$\begin{aligned} & \times \left[\frac{\sum_{k=1}^K \beta'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k)}{\left(1 + e^{\sum_{k=1}^K \beta'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k)}\right)} \right] \Bigg\} \\ \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n \left\{ \left[\sum_{k=1}^K \beta'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k) \right] \times \right. \\ & \times \left. \left[y_i - \frac{e^{\sum_{k=1}^K \beta'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k)}}{\left(1 + e^{\sum_{k=1}^K \beta'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k)}\right)} \right] \right\} \\ \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n \left\{ \left[\sum_{k=1}^K \beta'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k) \right] [y_i - \pi(\mathbf{z}_i)] \right\}. \end{aligned}$$

Da mesma forma deduziremos o Hessiano

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} &= - \sum_{i=1}^n \left\{ \pi(\mathbf{z}_i) \left[\sum_{k=1}^K \beta'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k) \right]^2 - \right. \\ & \left. - [\pi(\mathbf{z}_i)]^2 \left[\sum_{k=1}^K \beta'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k) \right]^2 \right\}, \end{aligned}$$

simplificando

$$\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = - \sum_{i=1}^n \left\{ \left[\sum_{k=1}^K \beta'_{K+k-2} \tilde{\mathbf{z}}_i B_k(\mathbf{x}_i; \boldsymbol{\theta}_k) \right]^2 \pi(\mathbf{z}_i) [1 - \pi(\mathbf{z}_i)] \right\}.$$

Neste ponto fixamos os parâmetros não-lineares, γ e c , e obtemos seus valores iniciais através de uma procura em *grid*. Supondo conhecidos tais parâmetros o modelo pode ser encarado como uma Regressão Logística onde o vetor de parâmetros, $\boldsymbol{\beta}$, é estimado por Máxima Verossimilhança, que necessita da utilização do processo iterativo de Newton-Raphson, como mostrado em (2-20). Desta forma para um STLR-Tree completo temos que

$$\boldsymbol{\beta}^{(m+1)} = [\mathbf{B}(\boldsymbol{\theta})' \mathbf{W}^{(m)} \mathbf{B}(\boldsymbol{\theta})]^{-1} \mathbf{B}(\boldsymbol{\theta})' \mathbf{W}^{(m)} \mathbf{z}^{(m)}$$

onde $\mathbf{z} = \mathbf{B}(\boldsymbol{\theta}) \boldsymbol{\beta}^{(m)} + \mathbf{W}^{-1}(\mathbf{y} - \boldsymbol{\pi})$ e

$$\mathbf{B}(\boldsymbol{\theta}) = \begin{pmatrix} B_1(\mathbf{x}_1; \boldsymbol{\theta}_1) & \dots & B_K(\mathbf{x}_1; \boldsymbol{\theta}_K) \\ \vdots & \ddots & \vdots \\ B_1(\mathbf{x}_N; \boldsymbol{\theta}_1) & \dots & B_K(\mathbf{x}_N; \boldsymbol{\theta}_K) \end{pmatrix}$$

Após estimados os parâmetros da parte linear, faz-se a estimação de γ e c também por Máxima Verossimilhança usando as estimativas de $\boldsymbol{\beta}$ nos cálculos.

A estimação dos parâmetros lineares e não-lineares segue basicamente o seguinte processo iterativo:

1. encontra-se os valores iniciais dos parâmetros não-lineares, γ e c , através de uma busca em *grid*, ao maximizar a log-verossimilhança concentrada;
2. aplica-se os valores encontrados no item anterior e estima-se por Máxima Verossimilhança os parâmetros lineares, $\boldsymbol{\beta}$;
3. as estimativas de $\boldsymbol{\beta}$ são usadas na estimação de γ e c também por Máxima Verossimilhança;
4. alternar os dois passos anteriores.

4.4

Avaliação do STLR-Tree

Feita após terem se encerrado as divisões e a árvore não possa mais crescer, assim nenhum nó terminal poderá ser dividido. A avaliação do ajuste é feita através dos métodos expostos em 2.2.3 onde são apresentadas as técnicas de avaliação do ajuste de uma Regressão Logística, que é basicamente a análise da Tabela de Classificação e da área abaixo da curva ROC.

Neste trabalho, entretanto, foi utilizada apenas a Tabela de Classificação (ou Matriz de Confusão) como medida da qualidade do ajuste.

4.5

Ciclo de Modelagem

Para minimizar a possibilidade de superestimar a quantidade de resultados significativos, evitando assim que sejam feitas mais divisões que o necessário, o nível de significância, α , sofre um desconto, ou defasagem, a medida que a árvore cresce da seguinte maneira

$$\alpha(d, n) = \frac{\alpha}{n^d} \tag{4-12}$$

onde n indica o n -ésimo teste aplicado e d a profundidade.

Assim, após a profundidade $d=0$, onde é aplicado somente um teste ($n = 1$) temos o nível de significância igual a α , a partir daí, na primeira profundidade ($d = 1$), são aplicados dois testes ($n = 2$) e o nível de significância passa a valer $\frac{\alpha}{2}$. A evolução de α de acordo com a profundidade e o número de testes se dá a partir do nó raiz como: $\alpha, \frac{\alpha}{2}, \frac{\alpha}{3}, \frac{\alpha}{4^2}, \frac{\alpha}{5^2}, \frac{\alpha}{6^2}, \frac{\alpha}{7^2}, \frac{\alpha}{8^3}, \frac{\alpha}{9^3}, \dots$

Tal método, que aumenta o rigor da significância com o aumento de d , evita que se venha a utilizar técnicas de podagem da árvore a posteriori (*post prunning*).

Todas as etapas para obtenção do STLR-Tree devem ser aplicadas a cada nova profundidade da árvore. A seguir é apresentada a modelagem a partir do nó raiz, depois da primeira profundidade e de forma generalizada, para a k -ésima profundidade.

Criação da Primeira profundidade (a partir de $d = 0$)

Para cada variável explicativa aplicar os testes de RV comparando cada um dos modelos que, sob H_0 , mantenha apenas a parte linear. Escolher as variável x_{s_0i} que gere o menor p -valor. Dado $s_0 \in \mathbb{S} = \{1, 2, \dots, m\}$ é feita a estimação do vetor de parâmetros, $\boldsymbol{\psi} = (\gamma_0, c_0, \beta_1, \beta_2)'$ conforme os métodos de estimação especificados anteriormente e testa-se as hipóteses

$$\begin{cases} H_{01} : \beta_1 = 0 \\ H_{02} : \beta_2 = 0 \\ H_{03} : \beta_1 - \beta_2 = 0 | \beta_1, \beta_2 \neq 0. \end{cases} \quad (4-13)$$

Se pelo menos uma das hipóteses não for rejeitada, busca-se a próxima variável de transição que gerou o segundo menor p -valor e reestima-se os parâmetros. Se para todos os $s_0 \in \mathbb{S}$ não forem produzidas divisões estatisticamente significativas, ou seja, se as hipóteses de linearidade não forem rejeitadas, a raiz é declarada como nó terminal e o modelo apenas com a parte linear é estimado. Caso seja gerada uma divisão estatisticamente significativa, rejeitando-se as hipóteses nulas, dois nós filhotes são gerados.

Criação da Segunda profundidade (a partir de $d = 1$)

Os dois nós filhotes gerados a partir do nó raiz compõem a primeira profundidade. A partir desta, deve-se escolher além da variável de transição, um dos nós a ser dividido e, assim continuar a aplicação dos testes RV para verificar a significância da inclusão de um novo regime, utilizando-se ainda o critério de seleção do par de combinações entre o índice da variável de transição em $\mathbb{S} = \{1, 2, \dots, m\}$ e o número do nó em $\mathbb{D} = \{1, 2\}$ que minimize

o p -valor. Sendo assim, estima-se os parâmetros e testa-se a significância da divisão através das hipóteses

$$\begin{cases} H_{01} : \beta_{2j_1+1} = 0 \\ H_{02} : \beta_{2j_1+2} = 0 \\ H_{03} : \beta_{2j_1+1} - \beta_{2j_1+2} = 0 | \beta_{2j_1+1}, \beta_{2j_1+2} \neq 0. \end{cases} \quad (4-14)$$

Se pelo menos uma das hipóteses não for rejeitada, busca-se o próximo par {nó; variável de transição}, que possua o segundo menor p -valor. Caso o contrário, a divisão seja aceita, testa-se a significância da divisão do nó $j_2 \in \mathbb{D} - \{j_1\}$, hipótese alternativa ao modelo com 3 nós terminais. Caso os dois nós da primeira profundidade gerem mais dois nós filhotes teremos, na segunda profundidade, 4 nós que serão: $2j_1 + 1, 2j_1 + 2, 2j_2 + 1$ e $2j_2 + 2$. Por outro lado, se nenhum dos dois nós gerarem divisões significativas paramos o crescimento da árvore e fazemos a avaliação do ajuste.

Criação da k -ésima profundidade

A generalização do processo de crescimento da árvore para uma profundidade $k > 0$, assumindo que foram criados N nós terminais anteriormente.

Para cada combinação $\{j_k; s_{j_k}\}$, de nó e variável de transição, é aplicado o teste de RV confrontando-os com o modelo apenas linear. As variáveis de transição pertencem ao conjunto $\mathbb{S} = \{1, 2, \dots, m\}$ enquanto que os nós estão em $\mathbb{D}_k = \{2^k - 1, 2^k, \dots, 2^{k+1} - 2\}$ e seleciona-se $j_k \in \mathbb{D}_k$ e $s_{j_k} \in \mathbb{S}$ que gere o menor p -valor. Com isso em mãos estima-se os parâmetros do modelo.

Como feito nas outras profundidades são aplicados os testes da significância das divisões iterativamente segundo a seqüência de nós: $j_2 \in \mathbb{D}_1 - \{j_1\}$, $j_3 \in \mathbb{D}_1 - \{j_1, j_2\}$, $j_4 \in \mathbb{D}_1 - \{j_1, j_2, j_3\}, \dots$

O ciclo é encerrado quando a profundidade em que se está aplicando os testes não gerem mais filhotes.

Simulação de dados através do STLR-Tree

Os dados foram gerados através de modelos STLR-Tree em sua forma mais simples, com apenas uma transição, alternando o parâmetro de suavidade entre os dois valores, $\gamma = 0.5$ e $\gamma = 50$. As covariáveis geradas foram: $x_1 \sim N(30, 8)$ e $x_2 \sim N(8.5, 6)$ e devido ao fato do parâmetro de suavidade ser dependente da escala, tais variáveis foram divididas por seus desvios padrão. Os valores dos parâmetros lineares para ambos os modelos são os seguintes: $\beta = \{2.5, -1.9, -0.2, 1.6, -0.17, 0.45\}$.

As figuras 4.3 e 4.4 correspondem ao modelo onde $c_0 = 4.3$ e $s_0 = 1$, para $\gamma = 0.5$ e $\gamma = 50$, respectivamente. Podemos notar a diferença entre as funções por meio da quebra que ocorre de acordo com a alteração do parâmetro de suavidade.

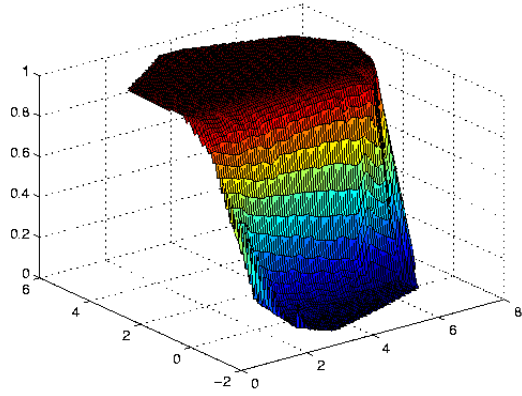


Figura 4.3: Dados gerados ($c_0 = 4.3$ e $s_0 = 1$ e $\gamma = 0.5$)

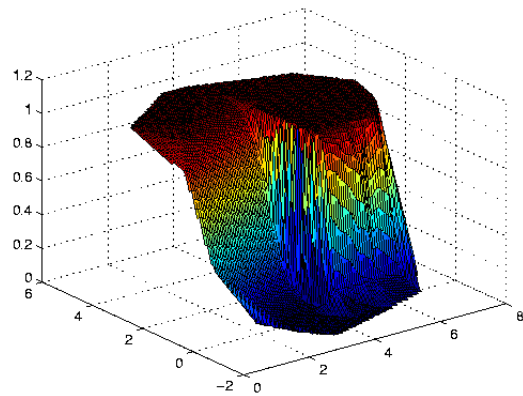


Figura 4.4: Dados gerados ($c_0 = 4.3$ e $s_0 = 1$ e $\gamma = 50$)