

CE043 - GAMLSS

Termos aditivos

Taconeli, C.A.; Silva, J.P.

10 de agosto, 2020

Introdução

- Neste módulo vamos tratar da forma como as covariáveis podem ser inseridas em um GAMLSS.

- Neste módulo vamos tratar da forma como as covariáveis podem ser inseridas em um GAMLSS.
- A incorporação adequada das covariáveis tem por objetivo modelar corretamente os respectivos efeitos nos parâmetros do modelo.

- Neste módulo vamos tratar da forma como as covariáveis podem ser inseridas em um GAMLSS.
- A incorporação adequada das covariáveis tem por objetivo modelar corretamente os respectivos efeitos nos parâmetros do modelo.
- Em modelos lineares, GLMs ou GAMs, este problema se limita, em geral, à especificação do preditor associado à média da distribuição.

- Neste módulo vamos tratar da forma como as covariáveis podem ser inseridas em um GAMLSS.
- A incorporação adequada das covariáveis tem por objetivo modelar corretamente os respectivos efeitos nos parâmetros do modelo.
- Em modelos lineares, GLMs ou GAMs, este problema se limita, em geral, à especificação do preditor associado à média da distribuição.
- Em GAMLSS, entretanto, temos a possibilidade de modelar qualquer parâmetro da distribuição em função de covariáveis.

- Um modelo generalizado aditivo para localização, escala e forma é definido, de forma geral, por:

$$y|\mathbf{x} \stackrel{ind}{\sim} D(\mu_{\mathbf{x}}, \sigma_{\mathbf{x}}, \nu_{\mathbf{x}}, \tau_{\mathbf{x}}),$$

onde

$$\begin{aligned} g_1(\mu_{\mathbf{x}}) &= \beta_{10} + \beta_{11}x_1 + \dots + \beta_{1j_1}x_{j_1} + \dots + s_{j_1+1}(x_{j_1+1}) + \dots + s_{r_1}(x_{r_1}) \\ g_2(\sigma_{\mathbf{x}}) &= \beta_{20} + \beta_{21}x_1 + \dots + \beta_{2j_2}x_{j_2} + \dots + s_{j_2+1}(x_{j_2+1}) + \dots + s_{r_2}(x_{r_2}) \\ g_3(\nu_{\mathbf{x}}) &= \beta_{30} + \beta_{31}x_1 + \dots + \beta_{3j_3}x_{j_3} + \dots + s_{j_3+1}(x_{j_3+1}) + \dots + s_{r_3}(x_{r_3}), \\ g_4(\tau_{\mathbf{x}}) &= \beta_{40} + \beta_{41}x_1 + \dots + \beta_{4j_4}x_{j_4} + \dots + s_{j_4+1}(x_{j_4+1}) + \dots + s_{r_4}(x_{r_4}) \end{aligned}$$

em que μ , σ , ν e τ são os parâmetros da distribuição.

- O termo *aditivo* se refere ao fato de que o preditor linear do modelo é definido pela soma de componentes que expressam os efeitos das covariáveis.

- O termo *aditivo* se refere ao fato de que o preditor linear do modelo é definido pela soma de componentes que expressam os efeitos das covariáveis.
- Importante ressaltar que ao preditor podem ser adicionados efeitos multiplicativos (como interações), de forma que a aditividade não implica ausência de interação.

- O termo *aditivo* se refere ao fato de que o preditor linear do modelo é definido pela soma de componentes que expressam os efeitos das covariáveis.
- Importante ressaltar que ao preditor podem ser adicionados efeitos multiplicativos (como interações), de forma que a aditividade não implica ausência de interação.
- Além disso, termos não lineares ou não paramétricos podem ser incluídos ao preditor por meio de funções suavizadoras.

Tabela 1: Alguns termos aditivos em implementados em `gamlss`

Termo aditivo	Função
Polinômios	<code>poly()</code>
Polinômios fracionários	<code>fp()</code>
Nós livres	<code>fk()</code>
Splines cúbicos	<code>cs()</code> , <code>scs()</code>
Regressão local	<code>lo()</code>
P-splines	<code>pb()</code> , <code>pb0()</code> , <code>ps()</code>
Árvores de decisão	<code>tr()</code>
Efeitos aleatórios	<code>random()</code> , <code>re()</code>

Termos aditivos

Termos aditivos

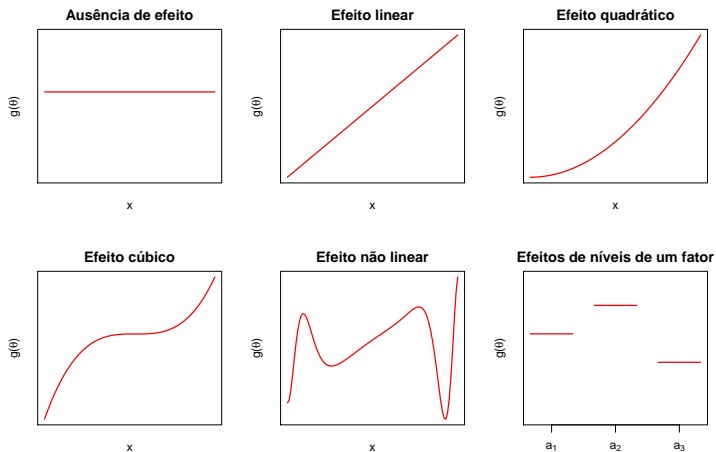


Figura 1: Ilustração de termos aditivos (Parte 1)

Termos aditivos

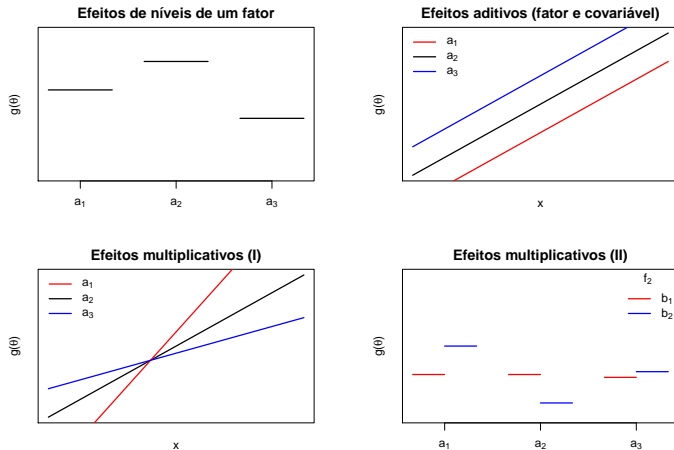


Figura 2: Ilustração de termos aditivos (Parte 2)

Termos aditivos

- Na sequência vamos apresentar diferentes especificações de termos aditivos para um parâmetro θ .

Termos aditivos

- Na sequência vamos apresentar diferentes especificações de termos aditivos para um parâmetro θ .
- Seja x uma covariável numérica. A forma mais simples de inseri-la a um preditor é através de seu efeito linear, na forma:

$$g(\theta) = \beta_0 + \beta_1 x.$$

Termos aditivos

- Na sequência vamos apresentar diferentes especificações de termos aditivos para um parâmetro θ .
- Seja x uma covariável numérica. A forma mais simples de inseri-la a um preditor é através de seu efeito linear, na forma:

$$g(\theta) = \beta_0 + \beta_1 x.$$

- Se f é um fator com m níveis (a_1, a_2, \dots, a_m) , então o mais usual é incorporá-lo ao preditor por:

$$g(\theta) = \beta_0 + \beta_1 \times I(f = a_2) + \beta_2 \times I(f = a_3) + \dots + \beta_{m-1} \times I(f = a_m),$$

em que $I(f = a_k) = 1$ se o indivíduo pertence ao nível k de f , e igual a zero, caso contrário.

- Observe que para incorporar f ao modelo, são criadas $m - 1$ variáveis indicadoras (variáveis *dummy*). Neste caso a primeira categoria (a_1) serve como referência.

- Observe que para incorporar f ao modelo, são criadas $m - 1$ variáveis indicadoras (variáveis *dummy*). Neste caso a primeira categoria (a_1) serve como referência.
- A omissão de uma variável indicadora é necessária pois, caso contrário, a matriz do modelo (matriz \mathbf{X}) apresentará dependência linear entre suas colunas (multicolinearidade).

- Observe que para incorporar f ao modelo, são criadas $m - 1$ variáveis indicadoras (variáveis *dummy*). Neste caso a primeira categoria (a_1) serve como referência.
- A omissão de uma variável indicadora é necessária pois, caso contrário, a matriz do modelo (matriz \mathbf{X}) apresentará dependência linear entre suas colunas (multicolinearidade).
- Um preditor pode acomodar simultaneamente diversas covariáveis numéricas e/ou fatores.

- A título de ilustração, considere x_1 , x_2 e x_3 variáveis explicativas numéricas, f_1 um fator com dois níveis (a_1 , a_2) e f_2 um fator com três níveis (b_1 , b_2 e b_3).

- A título de ilustração, considere x_1 , x_2 e x_3 variáveis explicativas numéricas, f_1 um fator com dois níveis (a_1 , a_2) e f_2 um fator com três níveis (b_1 , b_2 e b_3).
- Assumindo que os efeitos sejam aditivos no parâmetro a ser modelado, então o preditor linear fica definido por:

$$g(\theta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 \times I(f_1 = a_2) \\ + \beta_5 \times I(f_2 = b_2) + \beta_6 \times I(f_2 = b_3)$$

- Efeitos multiplicativos podem ser utilizados como forma de incorporar interações entre covariáveis.

- Efeitos multiplicativos podem ser utilizados como forma de incorporar interações entre covariáveis.
- Interações podem envolver duas ou mais variáveis numéricas, ou dois ou mais fatores, ou ainda variáveis numéricas e fatores.

- Efeitos multiplicativos podem ser utilizados como forma de incorporar interações entre covariáveis.
- Interações podem envolver duas ou mais variáveis numéricas, ou dois ou mais fatores, ou ainda variáveis numéricas e fatores.
- Além disso, podemos considerar como efeitos multiplicativos termos definidos por potências de variáveis, como x^2 , x^3 ,...

- Sejam x_1 e x_2 covariáveis numéricas. O seguinte preditor configura um modelo com termos de segunda ordem:

$$g(\theta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$$

- Sejam x_1 e x_2 covariáveis numéricas. O seguinte preditor configura um modelo com termos de segunda ordem:

$$g(\theta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$$

- Seja f_1 um fator com dois níveis (a_1 e a_2) e f_2 um fator com três níveis (b_1 , b_2 e b_3). Um modelo com efeito de interação fica definido da seguinte forma:

$$g(\theta) = \beta_0 + \beta_1 \times I(f_1 = a_2) + \beta_2 \times I(f_2 = b_2) + \beta_3 \times I(f_2 = b_3) \\ + \beta_4 \times I(f_1 = a_2) \times I(f_2 = b_2) + \beta_5 \times I(f_1 = a_2) \times I(f_2 = b_3)$$

Termos aditivos paramétricos

- **Polinômios** são a forma mais simples de modelar relações não lineares em regressão.

- **Polinômios** são a forma mais simples de modelar relações não lineares em regressão.
- Um preditor baseado num polinômio de grau p para uma covariável x tem a seguinte forma:

$$g(\theta) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \dots + \beta_p x^p.$$

- **Polinômios** são a forma mais simples de modelar relações não lineares em regressão.
- Um preditor baseado num polinômio de grau p para uma covariável x tem a seguinte forma:

$$g(\theta) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \dots + \beta_p x^p.$$

- As colunas da matriz do modelo correspondem aos vetores de 1's, x , x^2, \dots, x^p , formando uma *base polinomial*.

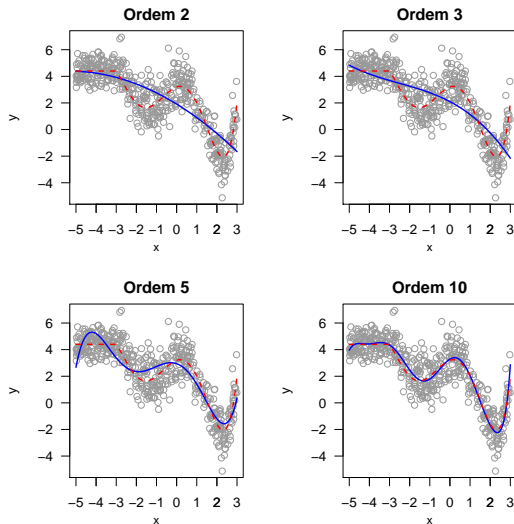


Figura 3: Modelos polinomiais de diferentes ordens.

- Modelos baseados em polinômios podem apresentar os seguintes inconvenientes:

- Modelos baseados em polinômios podem apresentar os seguintes inconvenientes:
 - Os valores de x^p podem aumentar rapidamente (ou diminuir, para $|x| < 1$), gerando problemas numéricos;

- Modelos baseados em polinômios podem apresentar os seguintes inconvenientes:
 - Os valores de x^p podem aumentar rapidamente (ou diminuir, para $|x| < 1$), gerando problemas numéricos;
 - A matriz do modelo pode ser mal condicionada devido à correlação entre as colunas (multicolinearidade).

- Modelos baseados em polinômios podem apresentar os seguintes inconvenientes:
 - Os valores de x^p podem aumentar rapidamente (ou diminuir, para $|x| < 1$), gerando problemas numéricos;
 - A matriz do modelo pode ser mal condicionada devido à correlação entre as colunas (multicolinearidade).
- Uma forma de contornar tais problemas é trocar a base polinomial padrão por uma base polinomial ortogonal (*polinômios ortogonais*).

- Modelos baseados em polinômios podem apresentar os seguintes inconvenientes:
 - Os valores de x^p podem aumentar rapidamente (ou diminuir, para $|x| < 1$), gerando problemas numéricos;
 - A matriz do modelo pode ser mal condicionada devido à correlação entre as colunas (multicolinearidade).
- Uma forma de contornar tais problemas é trocar a base polinomial padrão por uma base polinomial ortogonal (*polinômios ortogonais*).
- Polinômios ortogonais produzem valores ajustados idênticos aos polinômios não ortogonais (se ambos tiverem a mesma ordem).

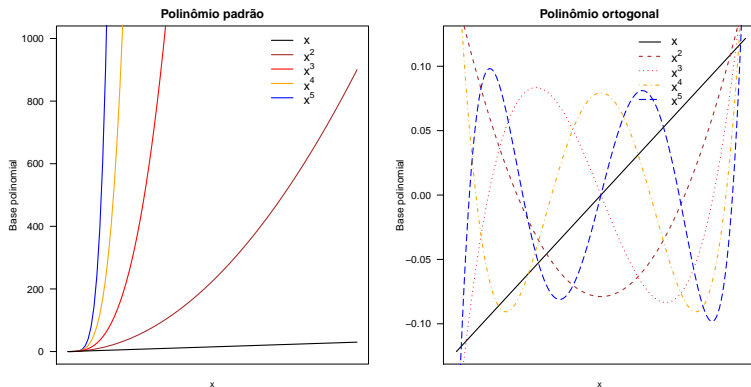


Figura 4: Base polinomial e base polinomial ortogonal.

- **Polinômios fracionários** se caracterizam pelo fato das potências utilizadas (p) não serem, necessariamente, números inteiros.

Polinômios fracionários

- **Polinômios fracionários** se caracterizam pelo fato das potências utilizadas (p) não serem, necessariamente, números inteiros.
- Por não se restringir a valores inteiros, polinômios fracionários produzem uma maior diversidade de bases polinomiais.

- **Polinômios fracionários** se caracterizam pelo fato das potências utilizadas (p) não serem, necessariamente, números inteiros.
- Por não se restringir a valores inteiros, polinômios fracionários produzem uma maior diversidade de bases polinomiais.
- Usando a função `fp` do R, podemos definir o número de termos de um polinômio fracionário pelo argumento `npoly`. Para `npoly=3`, por exemplo, o seguinte modelo fica especificado:

$$g(\theta) = \beta_0 + \beta_1 x^{p_1} + \beta_2 x^{p_2} + \beta_3 x^{p_3},$$

em que $p_1, p_2, p_3 \in \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$, com $p_j = 0$ associado a $\log(x)$, $j = 1, 2, 3$.

- Se $p_1 = p_2 = c$, então os termos $\beta_1 x^c$ e $\beta_2 x^c \log(x)$ são inseridos no modelo.

- Se $p_1 = p_2 = c$, então os termos $\beta_1 x^c$ e $\beta_2 x^c \log(x)$ são inseridos no modelo.
- Se $p_1 = p_2 = p_3 = c$, então os termos $\beta_1 x^c$, $\beta_2 x^c \log(x)$ e $\beta_3 x^c [\log(x)]^2$ são inseridos no modelo.

- Se $p_1 = p_2 = c$, então os termos $\beta_1 x^c$ e $\beta_2 x^c \log(x)$ são inseridos no modelo.
- Se $p_1 = p_2 = p_3 = c$, então os termos $\beta_1 x^c$, $\beta_2 x^c \log(x)$ e $\beta_3 x^c [\log(x)]^2$ são inseridos no modelo.
- A função **fp** ajusta todos os possíveis polinômios fracionários aos dados.

Polinômios fracionários

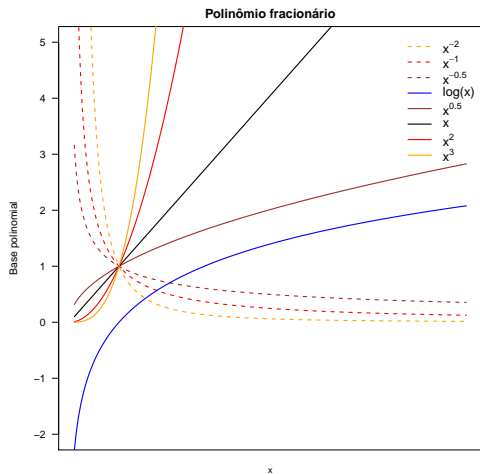


Figura 5: Base polinomial fracionária.

- O ajuste de modelos polinomiais, conforme descrito até o momento, requer a especificação da função de regressão globalmente, para todo o intervalo de valores de x .

Regressão por partes

- O ajuste de modelos polinomiais, conforme descrito até o momento, requer a especificação da função de regressão globalmente, para todo o intervalo de valores de x .
- Entretanto, nem sempre um único polinômio é capaz de descrever a variação dos dados globalmente, para todo x .

Regressão por partes

- O ajuste de modelos polinomiais, conforme descrito até o momento, requer a especificação da função de regressão globalmente, para todo o intervalo de valores de x .
- Entretanto, nem sempre um único polinômio é capaz de descrever a variação dos dados globalmente, para todo x .
- O ajuste de polinômios de elevada ordem, nesses casos, não é recomendável, devido:

- O ajuste de modelos polinomiais, conforme descrito até o momento, requer a especificação da função de regressão globalmente, para todo o intervalo de valores de x .
- Entretanto, nem sempre um único polinômio é capaz de descrever a variação dos dados globalmente, para todo x .
- O ajuste de polinômios de elevada ordem, nesses casos, não é recomendável, devido:
 - Ao elevado número de parâmetros;

- O ajuste de modelos polinomiais, conforme descrito até o momento, requer a especificação da função de regressão globalmente, para todo o intervalo de valores de x .
- Entretanto, nem sempre um único polinômio é capaz de descrever a variação dos dados globalmente, para todo x .
- O ajuste de polinômios de elevada ordem, nesses casos, não é recomendável, devido:
 - Ao elevado número de parâmetros;
 - Risco de *overfitting*;

- O ajuste de modelos polinomiais, conforme descrito até o momento, requer a especificação da função de regressão globalmente, para todo o intervalo de valores de x .
- Entretanto, nem sempre um único polinômio é capaz de descrever a variação dos dados globalmente, para todo x .
- O ajuste de polinômios de elevada ordem, nesses casos, não é recomendável, devido:
 - Ao elevado número de parâmetros;
 - Risco de *overfitting*;
 - Inflação da variância dos estimadores.

Regressão por partes

- Modelos de **regressão por partes** (ou **regressão segmentada**) são mais flexíveis nessas situações.

Regressão por partes

- Modelos de **regressão por partes** (ou **regressão segmentada**) são mais flexíveis nessas situações.
- Tais modelos consistem em dividir o intervalo de x em sub-intervalos e ajustar polinômios mais simples em cada intervalo, usando os dados disponíveis em cada um deles.

Regressão por partes

- Modelos de **regressão por partes** (ou **regressão segmentada**) são mais flexíveis nessas situações.
- Tais modelos consistem em dividir o intervalo de x em sub-intervalos e ajustar polinômios mais simples em cada intervalo, usando os dados disponíveis em cada um deles.
- O modelo mais simples de regressão por partes é aquele em que uma constante é ajustada em cada sub-intervalo.

Regressão por partes

- Modelos de **regressão por partes** (ou **regressão segmentada**) são mais flexíveis nessas situações.
- Tais modelos consistem em dividir o intervalo de x em sub-intervalos e ajustar polinômios mais simples em cada intervalo, usando os dados disponíveis em cada um deles.
- O modelo mais simples de regressão por partes é aquele em que uma constante é ajustada em cada sub-intervalo.
- Modelos mais complexos podem ser definidos aumentando a ordem do polinômio a ser ajustado em cada intervalo.

Regressão por partes

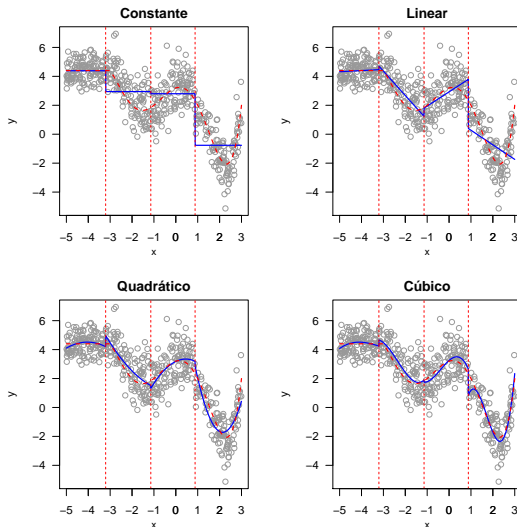


Figura 6: Regressões por partes de diferentes ordens.

Regressão por partes

- Sejam $\epsilon_1, \epsilon_2, \dots, \epsilon_K$ os pontos (*nós*) que dividem o intervalo de valores de x em $K + 1$ intervalos disjuntos, tal que cada uma das n observações pertença exclusivamente a um intervalo.

Regressão por partes

- Sejam $\epsilon_1, \epsilon_2, \dots, \epsilon_K$ os pontos (*nós*) que dividem o intervalo de valores de x em $K + 1$ intervalos disjuntos, tal que cada uma das n observações pertença exclusivamente a um intervalo.
- O modelo de regressão por partes definido por constantes tem a seguinte forma:

$$g(\theta) = \beta_0 + \beta_1 \times I(\epsilon_1 \leq x < \epsilon_2) + \beta_2 \times I(\epsilon_2 \leq x < \epsilon_3) + \dots \\ + \beta_{K-1} \times I(\epsilon_{K-1} \leq x < \epsilon_K)$$

Regressão por partes

- Sejam $\epsilon_1, \epsilon_2, \dots, \epsilon_K$ os pontos (*nós*) que dividem o intervalo de valores de x em $K + 1$ intervalos disjuntos, tal que cada uma das n observações pertença exclusivamente a um intervalo.
- O modelo de regressão por partes definido por constantes tem a seguinte forma:

$$g(\theta) = \beta_0 + \beta_1 \times I(\epsilon_1 \leq x < \epsilon_2) + \beta_2 \times I(\epsilon_2 \leq x < \epsilon_3) + \dots \\ + \beta_{K-1} \times I(\epsilon_{K-1} \leq x < \epsilon_K)$$

- Observe que o modelo fica definido por K parâmetros, de forma a produzir um único valor ajustado para cada intervalo entre nós definido em x .

- Uma segunda opção é o modelo de regressão por partes definido por segmentos de reta com interceptos e inclinações distintos em cada intervalo de x .

Regressão por partes

- Uma segunda opção é o modelo de regressão por partes definido por segmentos de reta com interceptos e inclinações distintos em cada intervalo de x .
- Neste caso, o modelo fica definido da seguinte forma:

$$\begin{aligned}g(\theta) = & \beta_0 + \beta_1 x \\ & + \beta_{01} \times I(\epsilon_1 \leq x < \epsilon_2) + \beta_{11} \times I(\epsilon_1 \leq x < \epsilon_2)x \\ & + \beta_{02} \times I(\epsilon_2 \leq x < \epsilon_3) + \beta_{12} \times I(\epsilon_2 \leq x < \epsilon_3)x + \dots \\ & + \beta_{0K-1} \times I(\epsilon_{K-1} \leq x < \epsilon_K) + \beta_{1K-1} \times I(\epsilon_{K-1} \leq x < \epsilon_K)x,\end{aligned}$$

tal que o modelo fica composto por $2K$ parâmetros, produzindo interceptos e inclinações diferentes para cada segmento de reta.

Regressão por partes

- De maneira semelhante, poderíamos especificar modelos de regressão por partes com polinômios de maior ordem. Para polinômios cúbicos, por exemplo:

$$\begin{aligned}g(\theta) = & \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 \\& + \beta_{01} \times I(I_{1,2}) + \beta_{11} \times I(I_{1,2})x + \beta_{21} \times I(I_{1,2})x^2 + \beta_{31} \times I(I_{1,2})x^3 \\& + \beta_{02} \times I(I_{2,3}) + \beta_{12} \times I(I_{2,3})x + \beta_{22} \times I(I_{2,3})x^2 + \beta_{32} \times I(I_{2,3})x^3 + \dots \\& + \beta_{0K-1} \times I(I_{K-1,K}) + \beta_{1K-1} \times I(I_{K-1,K})x + \beta_{2K-1} \times I(I_{K-1,K})x^2 \\& + \beta_{3K-1} \times I(I_{K-1,K})x^3,\end{aligned}$$

em que $I(I_{i,i+1})$ é uma representação abreviada de $I(\epsilon_i \leq x < \epsilon_{i+1})$, para $i = 1, 2, \dots, K - 1$.

- Modelos de regressão por partes apresenta inconvenientes como:

- Modelos de regressão por partes apresenta inconvenientes como:
 - A função de regressão é descontínua, uma vez que os segmentos de reta ajustados em intervalos vizinhos não se conectam.

- Modelos de regressão por partes apresenta inconvenientes como:
 - A função de regressão é descontínua, uma vez que os segmentos de reta ajustados em intervalos vizinhos não se conectam.
 - O excesso de parâmetros, à medida que se aumenta o número de nós.

- Modelos de regressão por partes apresenta inconvenientes como:
 - A função de regressão é descontínua, uma vez que os segmentos de reta ajustados em intervalos vizinhos não se conectam.
 - O excesso de parâmetros, à medida que se aumenta o número de nós.
- **Regressão por splines** corresponde a modelos de regressão segmentada em que restrições são impostas de forma que a função de regressão seja contínua e suave em todo o intervalo de x .

- Modelos de regressão por partes apresenta inconvenientes como:
 - A função de regressão é descontínua, uma vez que os segmentos de reta ajustados em intervalos vizinhos não se conectam.
 - O excesso de parâmetros, à medida que se aumenta o número de nós.
- **Regressão por splines** corresponde a modelos de regressão segmentada em que restrições são impostas de forma que a função de regressão seja contínua e suave em todo o intervalo de x .
- Impondo tais restrições, o número de parâmetros a serem estimados diminui consideravelmente.

- **Definição:** Uma função $s(x)$ é chamada de spline de grau m definida em uma partição do intervalo $[a, b]$ baseada em K nós $\epsilon_1, \epsilon_2, \dots, \epsilon_K \in [a, b]$, se:

- **Definição:** Uma função $s(x)$ é chamada de spline de grau m definida em uma partição do intervalo $[a, b]$ baseada em K nós $\epsilon_1, \epsilon_2, \dots, \epsilon_K \in [a, b]$, se:
 - $s(x)$ é um polinômio de grau m em cada subintervalo do tipo $[\epsilon_j, \epsilon_{j+1})$;

- **Definição:** Uma função $s(x)$ é chamada de spline de grau m definida em uma partição do intervalo $[a, b]$ baseada em K nós $\epsilon_1, \epsilon_2, \dots, \epsilon_K \in [a, b]$, se:
 - $s(x)$ é um polinômio de grau m em cada subintervalo do tipo $[\epsilon_j, \epsilon_{j+1})$;
 - $s(x)$ tem $m - 1$ derivadas contínuas em cada ϵ_j e, portanto, em $[a, b]$.

Regressão por splines

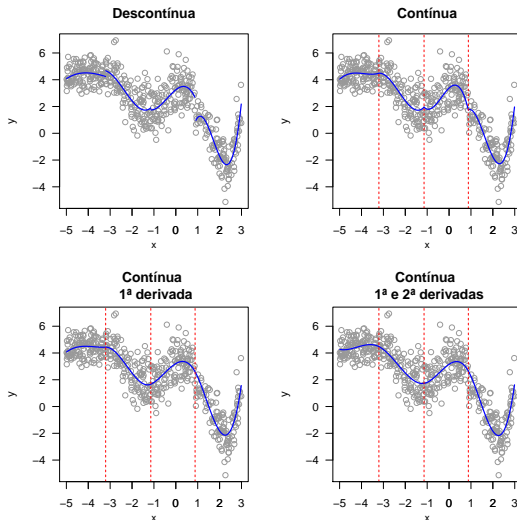


Figura 7: Modelos de regressão segmentada com polinômios cúbicos e diferentes níveis de continuidade.

- Um spline de grau m é definido da seguinte forma:

$$g(\theta) = \sum_{j=0}^m \beta_{0j} x^j + \sum_{j=1}^K \beta_j (x - \epsilon_j)_+^m,$$

em que

$$(x - \epsilon_j)_+^m = \begin{cases} (x - \epsilon_j)^m, & \text{se } x > \epsilon_j \\ 0, & \text{se } x \leq \epsilon_j \end{cases}.$$

Regressão por splines

- Assim, como casos particulares de splines, temos:

Regressão por splines

- Assim, como casos particulares de splines, temos:

- 1 Spline linear:

$$g(\theta) = \beta_{00} + \beta_{01}x + \sum_{j=1}^K \beta_j(x - \epsilon_j)_+$$

Regressão por splines

- Assim, como casos particulares de splines, temos:

- 1 Spline linear:

$$g(\theta) = \beta_{00} + \beta_{01}x + \sum_{j=1}^K \beta_j(x - \epsilon_j)_+$$

- 2 Spline quadrático:

$$g(\theta) = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \sum_{j=1}^K \beta_j(x - \epsilon_j)_+^2$$

Regressão por splines

- Assim, como casos particulares de splines, temos:

- 1 Spline linear:

$$g(\theta) = \beta_{00} + \beta_{01}x + \sum_{j=1}^K \beta_j(x - \epsilon_j)_+$$

- 2 Spline quadrático:

$$g(\theta) = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \sum_{j=1}^K \beta_j(x - \epsilon_j)_+^2$$

- 3 Spline cúbico:

$$g(\theta) = \beta_{00} + \beta_{01}x + \beta_{02}x^2 + \beta_{03}x^3 + \sum_{j=1}^K \beta_j(x - \epsilon_j)_+^3$$

Regressão por splines

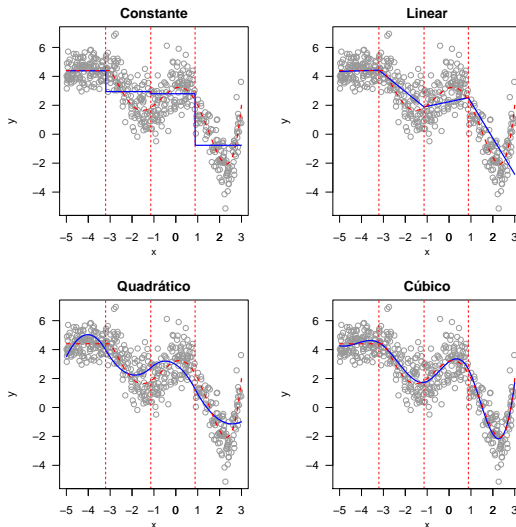


Figura 8: Splines de diferentes ordens.

- Splines cúbicos são muito populares, produzindo uma função de regressão contínua e bastante suave.

Regressão por splines

- Splines cúbicos são muito populares, produzindo uma função de regressão contínua e bastante suave.
- O ajuste de um spline cúbico com K nós baseia-se em $K + 3$ preditores, de maneira que a base correspondente é definida por $1, x, x^2, x^3, (x - \epsilon_1)_+^3, (x - \epsilon_2)_+^3, \dots, (x - \epsilon_K)_+^3$.

Regressão por splines

- Splines cúbicos são muito populares, produzindo uma função de regressão contínua e bastante suave.
- O ajuste de um spline cúbico com K nós baseia-se em $K + 3$ preditores, de maneira que a base correspondente é definida por $1, x, x^2, x^3, (x - \epsilon_1)_+^3, (x - \epsilon_2)_+^3, \dots, (x - \epsilon_K)_+^3$.
- Desta forma, um spline cúbico com K nós está associado a $K + 4$ graus de liberdade (um grau de liberdade adicional para o intercepto).

Regressão por splines

- Splines costumam apresentar elevada variância nos extremos ($x < \epsilon_1$ e $x > \epsilon_K$).

Regressão por splines

- Splines costumam apresentar elevada variância nos extremos ($x < \epsilon_1$ e $x > \epsilon_K$).
- Uma maneira de reduzir esse problema é usando **splines naturais**.

- Splines costumam apresentar elevada variância nos extremos ($x < \epsilon_1$ e $x > \epsilon_K$).
- Uma maneira de reduzir esse problema é usando **splines naturais**.
- Nos splines naturais, splines lineares são ajustados nos dois extremos, isto é, em $[a, \epsilon_1)$ e em $[\epsilon_K, b)$, e splines cúbicos nos intervalos intermediários.

Regressão por splines

- Splines costumam apresentar elevada variância nos extremos ($x < \epsilon_1$ e $x > \epsilon_K$).
- Uma maneira de reduzir esse problema é usando **splines naturais**.
- Nos splines naturais, splines lineares são ajustados nos dois extremos, isto é, em $[a, \epsilon_1)$ e em $[\epsilon_K, b)$, e splines cúbicos nos intervalos intermediários.
- Essa restrição adicional garante maior estabilidade no ajuste dos splines nas extremidades.

Splines e regressão segmentada

- A especificação de um modelo de regressão segmentada ou por splines requer determinar:

Splines e regressão segmentada

- A especificação de um modelo de regressão segmentada ou por splines requer determinar:
 - O número de nós (K);

Splines e regressão segmentada

- A especificação de um modelo de regressão segmentada ou por splines requer determinar:
 - O número de nós (K);
 - A locação dos nós ($\epsilon_1, \epsilon_2, \dots, \epsilon_k$);

Splines e regressão segmentada

- A especificação de um modelo de regressão segmentada ou por splines requer determinar:
 - O número de nós (K);
 - A locação dos nós ($\epsilon_1, \epsilon_2, \dots, \epsilon_k$);
 - O grau do polinômio (m) a ser ajustado em cada intervalo.

Splines e regressão segmentada

- A especificação de um modelo de regressão segmentada ou por splines requer determinar:
 - O número de nós (K);
 - A locação dos nós ($\epsilon_1, \epsilon_2, \dots, \epsilon_k$);
 - O grau do polinômio (m) a ser ajustado em cada intervalo.
- Quanto maiores K e m , melhor o ajuste do modelo aos dados, mas maior o risco de *overfitting*.

Splines e regressão segmentada

- A especificação de um modelo de regressão segmentada ou por splines requer determinar:
 - O número de nós (K);
 - A locação dos nós ($\epsilon_1, \epsilon_2, \dots, \epsilon_k$);
 - O grau do polinômio (m) a ser ajustado em cada intervalo.
- Quanto maiores K e m , melhor o ajuste do modelo aos dados, mas maior o risco de *overfitting*.
- Um spline de ordem m definido em uma partição do intervalo $[a, b]$ em K nós requer $K + m + 1$ parâmetros.

- Como estratégias para definir a alocação dos nós podemos considerar:

- Como estratégias para definir a alocação dos nós podemos considerar:
 - Nós uniformemente espalhados no intervalo $[a, b]$;

- Como estratégias para definir a alocação dos nós podemos considerar:
 - Nós uniformemente espalhados no intervalo $[a, b]$;
 - Nós espalhados com maior frequência nas regiões de x em $[a, b]$ em que a relação varia mais rapidamente e menor frequência em regiões onde a relação é mais estável.

- Como estratégias para definir a alocação dos nós podemos considerar:
 - Nós uniformemente espalhados no intervalo $[a, b]$;
 - Nós espalhados com maior frequência nas regiões de x em $[a, b]$ em que a relação varia mais rapidamente e menor frequência em regiões onde a relação é mais estável.
- Para definir o número de nós (K), podemos ajustar modelos com diferentes valores de K e comparar os ajustes resultantes via AIC.

Regressão por splines

- B-splines se diferenciam de splines por serem definidos localmente nos intervalo de valores de x em $[a, b]$.

Regressão por splines

- B-splines se diferenciam de splines por serem definidos localmente nos intervalo de valores de x em $[a, b]$.
- O conjunto de funções de base usados na construção de splines podem ser representadas por combinações de B-splines de mesmo grau (m).

Regressão por splines

- B-splines se diferenciam de splines por serem definidos localmente nos intervalo de valores de x em $[a, b]$.
- O conjunto de funções de base usados na construção de splines podem ser representadas por combinações de B-splines de mesmo grau (m).
- B-splines são colunas de uma matriz base \mathbf{B} , usada como matriz do modelo em problemas de regressão.

Regressão por splines

- B-splines se diferenciam de splines por serem definidos localmente nos intervalo de valores de x em $[a, b]$.
- O conjunto de funções de base usados na construção de splines podem ser representadas por combinações de B-splines de mesmo grau (m).
- B-splines são colunas de uma matriz base \mathbf{B} , usada como matriz do modelo em problemas de regressão.
- O número de graus de liberdade para o modelo ajustado é $m + K + 1$ (1 para o intercepto, m para o grau do polinômio e K para o número de nós em $[a, b]$).

Ilustração das funções de base para splines

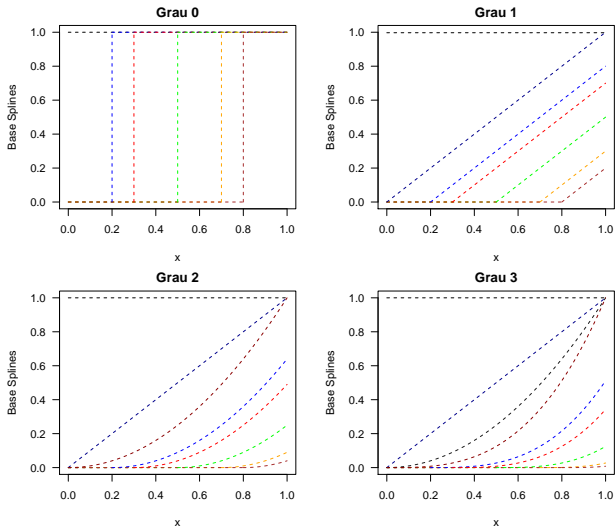


Figura 9: Funções de base (splines).

Ilustração das funções de base para B-splines

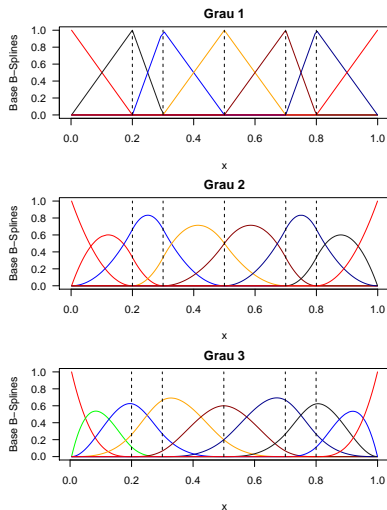


Figura 10: Funções de base (B-splines).

- Modelos com nós livres permitem que o número e a localização dos nós sejam estimados diretamente com base nos dados.

- Modelos com nós livres permitem que o número e a locação dos nós sejam estimados diretamente com base nos dados.
- O ajuste de modelos com nós livres configura um problema altamente não linear, e a função de verossimilhança pode apresentar problemas quanto a máximos locais.

- Modelos com nós livres permitem que o número e a locação dos nós sejam estimados diretamente com base nos dados.
- O ajuste de modelos com nós livres configura um problema altamente não linear, e a função de verossimilhança pode apresentar problemas quanto a máximos locais.
- O pacote `gamlss` dispõe de implementação para modelos com nós livres (função `fk`, pacote `gamlss.add`).

Termos não paramétricos

- **Smoothing splines** permitem modelar a relação entre uma ou mais covariáveis e um parâmetro da distribuição sem assumir uma forma paramétrica (sem a inclusão de parâmetros).

- **Smoothing splines** permitem modelar a relação entre uma ou mais covariáveis e um parâmetro da distribuição sem assumir uma forma paramétrica (sem a inclusão de parâmetros).
- Além dos smoothing splines, outros métodos podem ser usados para incorporar covariáveis através de termos não paramétricos, como aqueles baseadas em regressão local.

- **Smoothing splines** permitem modelar a relação entre uma ou mais covariáveis e um parâmetro da distribuição sem assumir uma forma paramétrica (sem a inclusão de parâmetros).
- Além dos smoothing splines, outros métodos podem ser usados para incorporar covariáveis através de termos não paramétricos, como aqueles baseadas em regressão local.
- Importante ter em mente que regressão por splines e smoothing splines são métodos distintos. A diferença, e o que têm em comum, ficarão mais claros adiante.

Smoothing splines

- Smoothing splines baseiam-se no seguinte problema de minimização.

Smoothing splines

- Smoothing splines baseiam-se no seguinte problema de minimização.
- Seja $g(x)$ uma função com primeira e segunda derivadas contínuas no intervalo $[a, b]$ e λ um parâmetro de penalização.

Smoothing splines

- Smoothing splines baseiam-se no seguinte problema de minimização.
- Seja $g(x)$ uma função com primeira e segunda derivadas contínuas no intervalo $[a, b]$ e λ um parâmetro de penalização.
- A função $g(x)$ que minimiza:

$$Q_2(g) = \sum_{i=1}^n (y_i - g(x_i))^2 + \lambda \int_a^b \{g''(t)\}^2 dt \quad (1)$$

é conhecida como smoothing spline.

Smoothing splines

- Algumas observações importantes sobre *smoothing splines*:

- Algumas observações importantes sobre *smoothing splines*:
 - O primeiro termo na equação (1) é a soma de quadrados de resíduos;

- Algumas observações importantes sobre *smoothing splines*:
 - O primeiro termo na equação (1) é a soma de quadrados de resíduos;
 - O segundo termo na equação (1) é um termo de penalização, que será maior quanto maior $g''(x)$ (função ruidosa) e menor quanto menor $g''(x)$ (função suave);

- Algumas observações importantes sobre *smoothing splines*:
 - O primeiro termo na equação (1) é a soma de quadrados de resíduos;
 - O segundo termo na equação (1) é um termo de penalização, que será maior quanto maior $g''(x)$ (função ruidosa) e menor quanto menor $g''(x)$ (função suave);
 - Num primeiro extremo, se adotarmos $\lambda = 0$ para a constante de penalização, então $g(x)$ poderá simplesmente interpolar os dados;

- Algumas observações importantes sobre *smoothing splines*:
 - O primeiro termo na equação (1) é a soma de quadrados de resíduos;
 - O segundo termo na equação (1) é um termo de penalização, que será maior quanto maior $g''(x)$ (função ruidosa) e menor quanto menor $g''(x)$ (função suave);
 - Num primeiro extremo, se adotarmos $\lambda = 0$ para a constante de penalização, então $g(x)$ poderá simplesmente interpolar os dados;
 - No outro extremo, para λ suficientemente grande, $g(x)$ será selecionada de maneira que $g''(x) = 0$ para todo $x \in [a, b]$, resultando no ajuste de uma reta de mínimos quadrados em todo o intervalo de x .

Smoothing splines

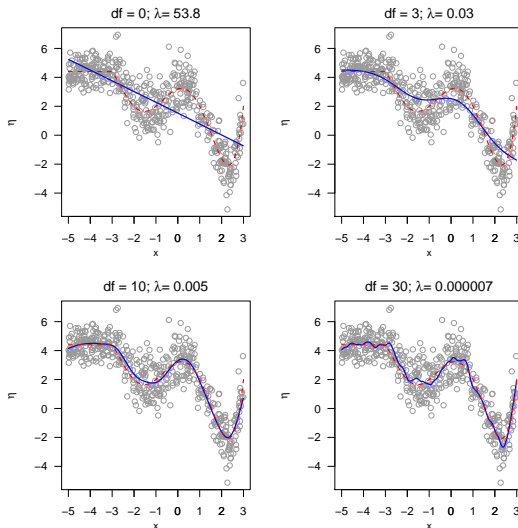


Figura 11: Smoothing splines para diferentes valores de λ .

- **Teorema:** A função g que minimiza a equação (1) é um spline cúbico natural com nós nos valores observados distintos de x .

- **Teorema:** A função g que minimiza a equação (1) é um spline cúbico natural com nós nos valores observados distintos de x .
- No entanto, o modelo obtido usando smoothing splines não é idêntico ao produzido usando splines cúbicos numa regressão por partes, devido ao efeito da penalização.

- **Teorema:** A função g que minimiza a equação (1) é um spline cúbico natural com nós nos valores observados distintos de x .
- No entanto, o modelo obtido usando smoothing splines não é idêntico ao produzido usando splines cúbicos numa regressão por partes, devido ao efeito da penalização.
- Adicionalmente, o número de graus de liberdade usados no modelo diminui substancialmente devido às restrições impostas pela penalização, prevenindo a simples interpolação dos dados.

Smoothing splines

- **Teorema:** A função g que minimiza a equação (1) é um spline cúbico natural com nós nos valores observados distintos de x .
- No entanto, o modelo obtido usando smoothing splines não é idêntico ao produzido usando splines cúbicos numa regressão por partes, devido ao efeito da penalização.
- Adicionalmente, o número de graus de liberdade usados no modelo diminui substancialmente devido às restrições impostas pela penalização, prevenindo a simples interpolação dos dados.
- Como o número de graus de liberdade está associado ao nível de penalização (e de suavização, consequentemente), é uma prática comum definir λ indiretamente, fixando o número de graus de liberdade.

Suavizador para variáveis categóricas (fatores)

- O pacote `gamlss` permite aplicar suavizadores também aos níveis de fatores, “encolhendo” seus efeitos em direção a zero.

Suavizador para variáveis categóricas (fatores)

- O pacote `gamlss` permite aplicar suavizadores também aos níveis de fatores, “encolhendo” seus efeitos em direção a zero.
- Uma das funções que podem ser aplicadas com essa finalidade é a `pcat`, que possibilita ainda agrupar níveis dos fatores com efeitos semelhantes.

Suavizador para variáveis categóricas (fatores)

- O pacote `gamlss` permite aplicar suavizadores também aos níveis de fatores, “encolhendo” seus efeitos em direção a zero.
- Uma das funções que podem ser aplicadas com essa finalidade é a `pcat`, que possibilita ainda agrupar níveis dos fatores com efeitos semelhantes.
- A intensidade do encolhimento é controlada por um parâmetro de suavização, denotado por λ .

Suavizador para variáveis categóricas (fatores)

- O pacote `gamlss` permite aplicar suavizadores também aos níveis de fatores, “encolhendo” seus efeitos em direção a zero.
- Uma das funções que podem ser aplicadas com essa finalidade é a `pcat`, que possibilita ainda agrupar níveis dos fatores com efeitos semelhantes.
- A intensidade do encolhimento é controlada por um parâmetro de suavização, denotado por λ .
- O ajuste resultante vai gastar menos graus de liberdade que o número de níveis resultante do agrupamento dos níveis originais (devido à penalização).

Suavizador para variáveis categóricas (fatores)

- A função de penalização pode ser vista como a solução de um problema de mínimos quadrados penalizados:

$$(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma})'(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma}) + \lambda \|\mathbf{D}\boldsymbol{\gamma}\|_p^p,$$

em que \mathbf{Z} é uma matriz de incidências (matriz de variáveis indicadoras), $\|\mathbf{D}\boldsymbol{\gamma}\|_p^p$ uma matriz de penalidades.

Suavizador para variáveis categóricas (fatores)

- A função de penalização pode ser vista como a solução de um problema de mínimos quadrados penalizados:

$$(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma})'(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma}) + \lambda \|\mathbf{D}\boldsymbol{\gamma}\|_p^p,$$

em que \mathbf{Z} é uma matriz de incidências (matriz de variáveis indicadoras), $\|\mathbf{D}\boldsymbol{\gamma}\|_p^p$ uma matriz de penalidades.

- Para $p = 1$, a penalização aplicada é de primeira ordem, sendo resultante da soma dos valores absolutos das penalidades; para $p = 2$, a penalização é quadrática, e assim por diante.

Suavizador para variáveis categóricas (fatores)

- A função de penalização pode ser vista como a solução de um problema de mínimos quadrados penalizados:

$$(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma})'(\mathbf{y} - \mathbf{Z}\boldsymbol{\gamma}) + \lambda \|\mathbf{D}\boldsymbol{\gamma}\|_p^p,$$

em que \mathbf{Z} é uma matriz de incidências (matriz de variáveis indicadoras), $\|\mathbf{D}\boldsymbol{\gamma}\|_p^p$ uma matriz de penalidades.

- Para $p = 1$, a penalização aplicada é de primeira ordem, sendo resultante da soma dos valores absolutos das penalidades; para $p = 2$, a penalização é quadrática, e assim por diante.
- Quanto maior o valor de λ , maior a penalidade aplicada, e maior o encolhimento (tendência de agrupar os níveis originais em um menor número de novos níveis).

Resumindo

- Os efeitos de covariáveis e fatores são inseridos aos preditores em GAMLSS através de termos aditivos;

- Os efeitos de covariáveis e fatores são inseridos aos preditores em GAMLSS através de termos aditivos;
- Termos aditivos podem ser paramétricos, por meio de polinômios e regressões segmentadas, ou não paramétricos, baseados em smoothing splines ou regressões locais, dentre outros;

- Os efeitos de covariáveis e fatores são inseridos aos preditores em GAMLSS através de termos aditivos;
- Termos aditivos podem ser paramétricos, por meio de polinômios e regressões segmentadas, ou não paramétricos, baseados em smoothing splines ou regressões locais, dentre outros;
- Problemas como overfitting e multicolinearidade devem ser considerados na especificação dos preditores;

- Os efeitos de covariáveis e fatores são inseridos aos preditores em GAMLSS através de termos aditivos;
- Termos aditivos podem ser paramétricos, por meio de polinômios e regressões segmentadas, ou não paramétricos, baseados em smoothing splines ou regressões locais, dentre outros;
- Problemas como overfitting e multicolinearidade devem ser considerados na especificação dos preditores;
- O diagnóstico do ajuste (análise de resíduos) e medidas de qualidade de ajuste (como AIC) são fundamentais nesta etapa da modelagem.

Próximos passos

- Família GAMLSS

- Família GAMLSS
 - Distribuições contínuas;

- Família GAMLSS
 - Distribuições contínuas;
 - Distribuições discretas;

- Família GAMLSS
 - Distribuições contínuas;
 - Distribuições discretas;
 - Distribuições contínuas em $(0,1)$;

- Família GAMLSS
 - Distribuições contínuas;
 - Distribuições discretas;
 - Distribuições contínuas em $(0,1)$;
 - Distribuições produzidas por misturas.