4
Algoritmo e Comentários sobre o Método

Nesse capítulo será apresentado o passo a passo do algoritmo recursivo que constrói a árvore do método RCRI e em seguida alguns comentários importantes sobre ele.

4.1
Algoritmo: Regressão Construtiva por Regiões Implícitas

Antes da descrição do algoritmo serão mostrados quais são os dados de entrada fornecidos pelo usuário. Como já foi comentado, a amostra inicial é formada pelos conjuntos $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} \subset \mathbb{R}^d$ e $Y = \{y_1, y_2, \ldots, y_N\} \subset \mathbb{R}$. Além disso, os parâmetros que devem ser definidos antes do algoritmo começar são: g, grau escolhido para os polinômios no nós internos; d_{lim}, profundidade limite da árvore; $n_{lim} = qN > n(g, d)$, número mínimo de pontos admitidos em um certo nó; $0 < \alpha < 1$, parâmetro do teste de hipótese do 3º critério de parada; e λ, parâmetro da função $\mathbf{G}$.

Algoritmo:

Passo 1) Para X e Y definidos, encontre os coeficientes de $\mathbf{p}$ de acordo com a equação (2-2).

Passo 2) Defina,

$$\begin{aligned} X_1 &= \{\mathbf{x} \in X \mid \mathbf{p}(\mathbf{x}) \leq 0\} \\ X_2 &= \{\mathbf{x} \in X \mid \mathbf{p}(\mathbf{x}) > 0\} \\ Y_1 &= \{y_i \in Y \mid \mathbf{x}_i \in X_1\} \\ Y_2 &= \{y_i \in Y \mid \mathbf{x}_i \in X_2\} \\ n_1 &= \text{número de elementos em } X_1 \\ n_2 &= \text{número de elementos em } X_2 \end{aligned}$$

Passo 3) Se $n_1 > n_{lim}$ e $n_2 > n_{lim}$, então siga para o passo 4.
Se $n_1 \leq n_{lim}$ ou $n_2 \leq n_{lim}$, então esse nó será uma folha, FIM.

Passo 4) Seja d a profundidade desse nó da árvore.
Se $d < d_{lim}$, então siga para o passo 5.
Se $d \geq d_{lim}$, então esse nó será uma folha, FIM.

Passo 5) Seja t_0 é dado pela equação 2-3.
Se $-t_{\frac{\alpha}{2},n_1+n_2-2} < t_o < t_{\frac{\alpha}{2},n_1+n_2-2}$, siga para o passo 6.
Se $t_o < -t_{\frac{\alpha}{2},n_1+n_2-2}$ ou $t_o > t_{\frac{\alpha}{2},n_1+n_2-2}$, então esse nó será uma folha, FIM.

Passo 6) É criado um nó interno e os coeficientes de $\mathbf{p}$ são armazenados por ele.

Passo 7) Para determinar o filho à esquerda desse nó interno, volte ao passo 1 com $X = X_1$ e $Y = Y_1$.

Passo 8) Para determinar o filho à direita desse nó interno, volte ao passo 1 com $X = X_2$ e $Y = Y_2$.

Passo 9) Para estimar os intervalos nas folhas escolha entre os métodos local e global, apresentados em (3-7) e (3-8), respectivamente.

4.2 Alguns Comentários sobre o Método

O processo de avaliação também é feito de forma recursiva. Por exemplo, suponha uma árvore τ e um ponto $\mathbf{x} \in \mathbb{R}^d$ qualquer, para o qual se deseja avaliar sua estimativa para essa árvore. Se esta estimativa for chamada de $\widehat{\mathbf{F}_\tau}(\mathbf{x})$, seu valor pode ser determinado da seguinte maneira:

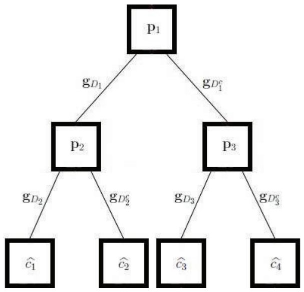
$$\widehat{\mathbf{F}_\tau}(\mathbf{x}) = \begin{cases} \widehat{c} & \text{, se a árvore } \tau \text{ for uma folha;} \\ \mathbf{g}_D(\mathbf{x})\widehat{\mathbf{F}}_{\tau_E}(\mathbf{x}) + \mathbf{g}_{D^c}(\mathbf{x})\widehat{\mathbf{F}}_{\tau_D}(\mathbf{x}) & \text{, caso contrário.} \end{cases} \tag{4-1}$$

onde $\widehat{c}$ é o coeficiente guardado pela folha, caso a árvore seja formada por uma única folha; $\widehat{\mathbf{F}}_{\tau_E}(\mathbf{x})$ representa a estimativa fornecida pela sub-árvore à esquerda; $\widehat{\mathbf{F}}_{\tau_D}(\mathbf{x})$ representa a estimativa pela sub-árvore à direita; e $D \cup D^c$ é a partição do domínio definida pelo nó interno raiz, no caso da árvore conter algum nó interno.

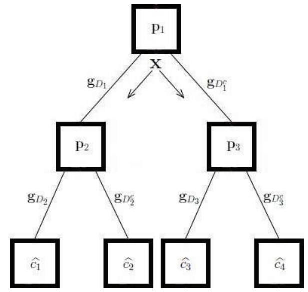
Vantagens na escolha da função degrau

A função degrau $\mathbf{s}$ apresentada pela equação (3-3) é a escolhida para compor a função de transição com relação às regiões definidas por cada nó interno. A escolha dessa função se deve principalmente ao fato de que com ela o conjunto $\{t \in \mathbb{R} \mid \mathbf{s}(t) \neq 0 \text{ e } \mathbf{s}(t) \neq 1\}$ é limitado, e assim o processo de avaliação é bem simplificado quando se tem $\mathbf{g}_D(\mathbf{x}) = \mathbf{s}(\mathbf{d}_D(\mathbf{x})) = 0$ ou

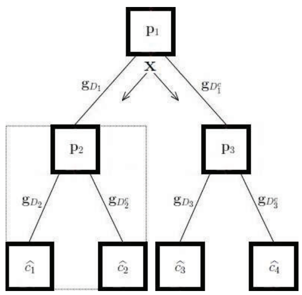
$\mathbf{g}_{D^c}(\mathbf{x}) = \mathbf{s}(\mathbf{d}_D^c(\mathbf{x})) = 0$. Esta simplificação pode ser melhor entendida na ilustração a seguir.



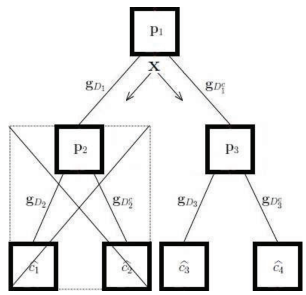
4.1(a): Suponha uma árvore já estruturada, para a qual se deseja avaliar a estimativa em um certo ponto $\mathbf{x}$.



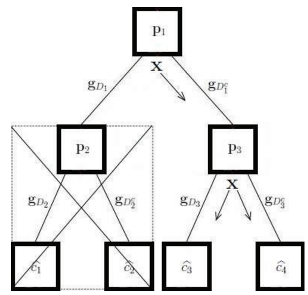
4.1(b): De acordo com a equação (4-1), $\widehat{\mathbf{F}_\tau}(\mathbf{x}) = \mathbf{g}_{D_1}(\mathbf{x})\widehat{\mathbf{F}}_{\tau_E}(\mathbf{x}) + \mathbf{g}_{D_1^c}(\mathbf{x})\widehat{\mathbf{F}}_{\tau_D}(\mathbf{x})$.



4.1(c): Se, por exemplo, $\mathbf{g}_{D_1}(\mathbf{x}) = 0$ ...



4.1(d): ... toda a sub-árvore à esquerda será desconsiderada na avaliação.



4.1(e): Nesse caso, só serão percorridos os nós da sub-árvore à direita.

Figura 4.1: Ilustração para a avaliação do método RCRI em um ponto qualquer.

No exemplo exposto na Figura 4.1, $\mathbf{g}_{D_1}(\mathbf{x}) = 0$ quando a função degrau $\mathbf{s}$ é aquela apresentada na equação (3-3). Então, a estimativa final é dada por:

$$\widehat{\mathbf{F}_\tau}(\mathbf{x}) = \mathbf{g}_{D_1^c}(\mathbf{x})\mathbf{g}_{D_3}(\mathbf{x})\widehat{c}_3 + \mathbf{g}_{D_1^c}(\mathbf{x})\mathbf{g}_{D_3^c}(\mathbf{x})\widehat{c}_4.$$

Se fosse usada outra função degrau $\mathbf{s}$, tal que $\mathbf{s}(t) \neq 0$ e $\mathbf{s}(t) \neq 1 \ \forall \ t \in \mathbb{R}$, como é o caso da função logística apresentada na equação (3-1), a estimativa final seria:

$$\begin{aligned}\widehat{\mathbf{F}_\tau}(\mathbf{x}) = \ & \mathbf{g}_{D_1}(\mathbf{x})\mathbf{g}_{D_2}(\mathbf{x})\widehat{c}_1 + \mathbf{g}_{D_1}(\mathbf{x})\mathbf{g}_{D_2^c}(\mathbf{x})\widehat{c}_2 + \\ & \mathbf{g}_{D_1^c}(\mathbf{x})\mathbf{g}_{D_3}(\mathbf{x})\widehat{c}_3 + \mathbf{g}_{D_1^c}(\mathbf{x})\mathbf{g}_{D_3^c}(\mathbf{x})\widehat{c}_4.\end{aligned}$$

Comparando as duas equações acima, é nítida a diminuição no número de operações feitas para realizar a avaliação.