



# APLICAÇÕES DE ONDALETAS EM PROBLEMAS DE APRENDIZADO ESTATÍSTICO SUPERVISIONADO

**Palavras-Chave:** Ondaletas; Regressão Não Paramétrica; Classificação.

**Autores:**

João Victor Siqueira Rodrigues, IMECC - UNICAMP  
Prof. Dr. Alex Rodrigo dos Santos Sousa (orientador), IMECC - UNICAMP

## 1 Introdução

É comum, dentro de diversas áreas da ciência, a busca em relacionar determinadas variáveis, denominadas de variáveis explicativas ou preditoras, com uma variável de interesse. Essa relação faz-se interessante por possibilitar, por exemplo, fazer previsões sobre o comportamento da variável resposta com base nas suas variáveis preditoras. Tal problemática pode ser explorada por problemas de aprendizado estatístico supervisionado. Para mais detalhes, ver Hastie *et al.* (2008), James *et al.* (2014) e Hastie *et al.* (2015).

Ademais, métodos baseados em ondaletas estão sendo utilizados em diversas áreas da matemática e podem ser aplicadas à problemas de aprendizado estatístico, de forma a possibilitar, por meio da Transformada Discreta de Ondaletas, uma análise no domínio das ondaletas por níveis de resolução, entre outras características, como a representação esparsa dos dados. Desta forma, métodos baseados em ondaletas tem sido extensivamente aplicados em problemas de aprendizado estatístico supervisionado, ver Vidakovic (1999) e Nason (2008).

Por conseguinte, devido às características das ondaletas e da essência dos problemas de aprendizado estatístico supervisionado, muitas aplicações vêm sendo desenvolvidas. Além disso, foi feito o estudo em como os métodos baseados em ondaletas se inserem nos problemas de aprendizado estatístico supervisionado relativos à classificação.

## 2 Principais Conceitos

### 2.1 Ondaletas

Considere um vetor  $\mathbf{y} = (y_1, \dots, y_n)^T$ , com  $n = 2^J$ ,  $J \in \mathbb{N}$ , então dizemos que  $\mathbf{y}$  é um vetor diádico. Partindo de um vetor diádico, pode-se extrair os coeficientes da ondaleta  $d_{jk}$  por meio de subtrações normalizadas e os coeficientes da função escala  $c_{jk}$  por meio de somas normalizadas. O subíndice  $j$  representa o nível de resolução ou escala dos coeficientes, em que, no nível mais fino de resolução ( $j = J - 1$ ), os coeficientes são dados pelas equações (1) e (2), enquanto que para os demais níveis calcula-se a partir das equações (3) e (4):

$$d_{jk} = \frac{1}{\sqrt{2}}(y_{2k+1} - y_{2k}) \quad (1) \quad d_{jk} = \frac{1}{\sqrt{2}}(c_{j+1,2k+1} - c_{j+1,2k}) \quad (3)$$

$$c_{jk} = \frac{1}{\sqrt{2}}(y_{2k+1} + y_{2k}) \quad (2) \quad c_{jk} = \frac{1}{\sqrt{2}}(c_{j+1,2k+1} + c_{j+1,2k}) \quad (4)$$

onde  $j = 0, \dots, J - 1$  e  $k = 0, \dots, 2^j - 1$ . O processo de obter estes coeficientes é chamado de Transformada Discreta de Ondaletas (DWT, sigla em inglês).

**Exemplo 1 (Nason, 2008).** Considere o vetor de dados  $\mathbf{y} = (1, 1, 7, 9, 2, 8, 8, 6)^T$ , com  $2^3$  elementos. Os coeficientes da função escala para o nível de resolução  $j = 2$  podem ser obtidos pela equação (2), a qual retorna  $c_{20} = \frac{1}{\sqrt{2}}(y_1 + y_0) = \sqrt{2}$ ,  $c_{21} = \frac{1}{\sqrt{2}}(y_3 + y_2) = 8\sqrt{2}$ ,  $c_{22} = \frac{1}{\sqrt{2}}(y_5 + y_4) = 5\sqrt{2}$  e  $c_{23} = \frac{1}{\sqrt{2}}(y_7 + y_6) = 7\sqrt{2}$ . Os próximos níveis podem ser obtidos aplicando a equação (4). De forma análoga, pode-se obter os coeficientes da ondaleta aplicando (1) e (3). A Figura 1 apresenta uma descrição gráfica das operações que devem ser feitas, em que as setas tracejadas correspondem aos  $d_{jk}$  e as setas sólidas correspondem aos  $c_{jk}$ .

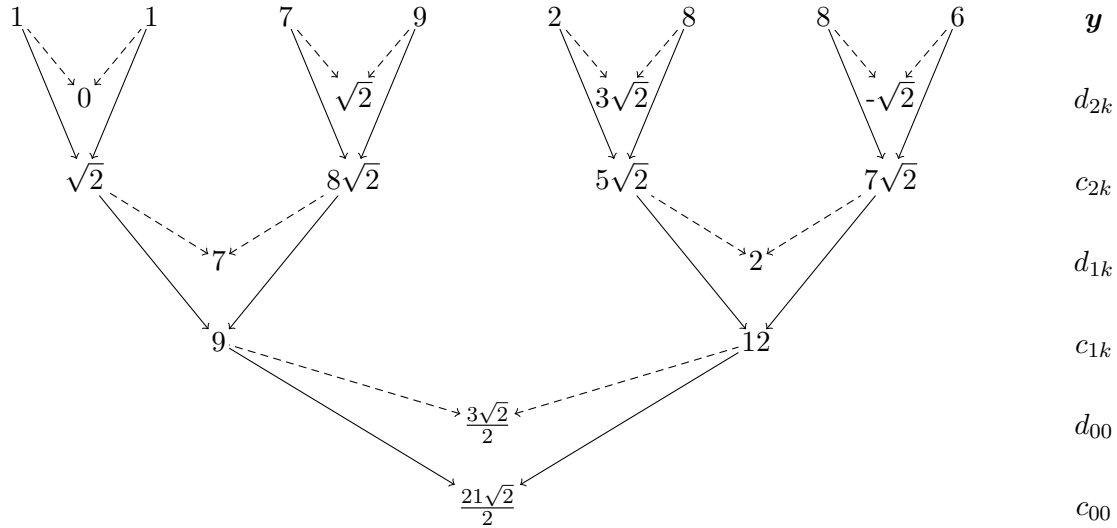


Figura 1: Descrição gráfica das operações para obtenção dos coeficientes  $d_{jk}$  e  $c_{jk}$  por nível de resolução

Além disso, é possível reconstruir a sequência de dados originais conhecendo apenas os valores dos coeficientes de ondaleta  $d_{jk}$  e  $c_{00}$ , este processo é chamado de Transformada Discreta de Ondaletas Inversa (IDWT, sigla em inglês), dada pelas equações:

$$c_{j,2k} = \frac{1}{\sqrt{2}}(c_{j-1,k} + d_{j-1,k}) \quad (5)$$

$$c_{j,2k-1} = \frac{1}{\sqrt{2}}(c_{j-1,k} - d_{j-1,k}) \quad (6)$$

onde  $j = 0, \dots, J - 1$  e  $k = 0, \dots, 2^j - 1$ .

Ademais, considere o modelo no formato vetorial dado por:

$$\mathbf{y} = \mathbf{f} + \mathbf{e}$$

onde  $\mathbf{y} = (y_1, \dots, y_n)^T$  é o vetor de valores observados,  $\mathbf{f} = (f(x_1), \dots, f(x_n))^T$  é a função verdadeira e  $\mathbf{e} = (e_1, \dots, e_n)^T$  é o erro independente e identicamente distribuído (*i.i.d.*) com  $e_i \sim \mathcal{N}(0, \sigma^2)$ ,  $i = 1, \dots, n$ . Tem-se por objetivo estimar a função  $f$  e, para isso, é aplicado a DWT, representada por sua matriz  $W$ , de forma que:

$$\mathbf{d}^* = \mathbf{d} + \boldsymbol{\varepsilon}$$

com  $\mathbf{d}^* = W\mathbf{y}$ ,  $\mathbf{d} = W\mathbf{f}$  e  $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T = W\mathbf{e}$ . Além disso, as propriedades do erro  $e_i$  são mantidas, logo,  $\varepsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$ . Portanto, o problema de estimar  $f$  foi substituído por estimar os coeficientes de ondaletas  $\mathbf{d}$  utilizando os coeficientes de ondaletas empíricos  $\mathbf{d}^*$ . Para isso, utiliza-se a propriedade de esparsidade dos coeficientes de ondaletas, já que os coeficientes nulos estão associados à parte suave da função e os coeficientes não nulos associam-se às principais características da função, como descontinuidades, mudanças de comportamento, picos, entre outras. Logo, tem-se por objetivo encolher (ou zerar) os coeficientes de forma a reduzir o ruído aleatório  $\boldsymbol{\varepsilon}$ .

Escolhido um limiar (*threshold*)  $\lambda$ , é necessário determinar sua regra. Serão ressaltadas duas regras, a de limiar duro (*hard threshold*)  $\delta_\lambda^H$  e a de limiar suave (*soft threshold*)  $\delta_\lambda^S$ , dadas por:

$$\delta_\lambda^H(d^*) = d^* \mathcal{I}\{|d^*| > \lambda\}$$

$$\delta_\lambda^S(d^*) = \text{sgn}(d^*)(|d^*| - \lambda) \mathcal{I}\{|d^*| > \lambda\}$$

onde  $\mathcal{I}$  é a função indicadora. A Figura 2 apresenta a regra do limiar duro e suave. Após a aplicação da regra de limiar aos coeficientes empíricos  $d^*$  basta recuperar a função estimada  $\hat{f}$  por meio da IDWT.

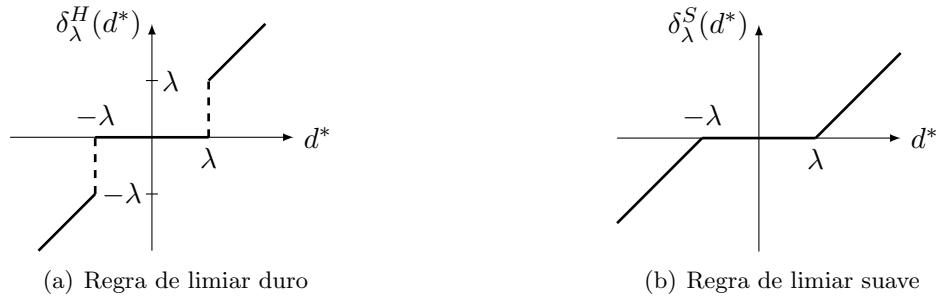


Figura 2: Regra de limiar duro e suave

Ademais, existem diversas propostas para a escolha de  $\lambda$ , como *universal threshold*, dado por:

$$\lambda = \sigma \sqrt{2 \ln(n)}$$

onde  $\sigma$  é estimado pelos desvios absolutos da mediana (MAD, sigla em inglês) dos coeficientes empíricos da escala mais fina. Além de outras, como validação cruzada, *false discovery rate*, SURE.

**Exemplo 2.** *Spike Sorting* é um processo utilizado na classificação de dados eletrofisiológicos, o qual utiliza as diferentes formas e amplitudes captadas por eletrodos de forma extracelular para sua classificação. Contudo, é necessário filtrar esses dados para redução de ruído facilitando a detecção de *spikes* e evitando erros de classificação. A Figura 3 apresenta a recuperação de 1024 medições, dados retirados de Kenneth Harris do *Institute of Neurology, Faculty of Brain Sciences*, disponível em [International Federation of Classification Societies](#).

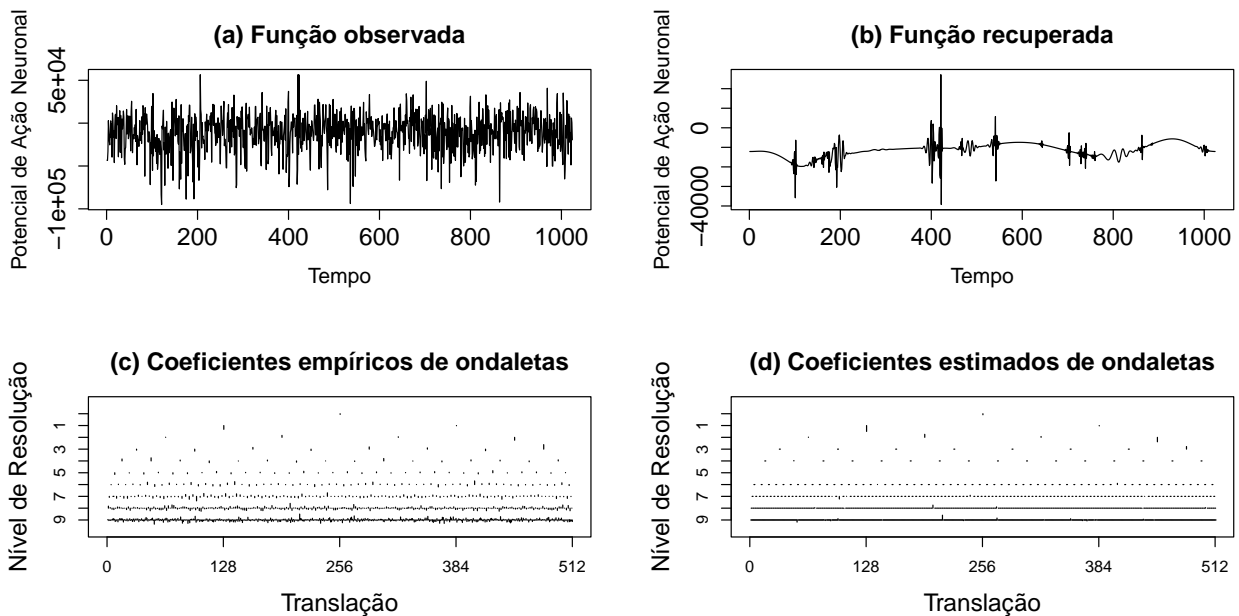


Figura 3: Recuperação da função de disparos neuronais

## 2.2 Classificação de Curvas com Ondaletas

Dentro da área de classificação existem diversas técnicas, dentre elas a regressão logística. Sumariamente, este tipo de regressão busca modelar  $p(X) = \mathbb{P}(Y = 1 | X)$  por meio da função logística:

$$p(X) = \frac{\exp\{\beta_0 + \beta_1 X\}}{1 + \exp\{\beta_0 + \beta_1 X\}} \quad (7)$$

Ademais, manipulando a equação (7), pode-se obter:

$$\text{logit}(p(X)) = \log\left(\frac{p(X)}{1 - p(X)}\right) = \beta_0 + \beta_1 X$$

Além disso, definindo  $Z_i \in \{C_1, C_2\}$  uma variável de classificação, tal que  $p_i = \mathbb{P}(Z_i = C_1 | \mathbf{Y}_i)$ , e assumindo que as observações são computadas por  $\{\mathbf{Y}_i, Z_i\}$  de forma que  $\mathbf{Y}_i$  é um vetor diádico de medições sequências. Então, pode-se propor o seguinte modelo de classificação (Wang *et al.*, 2007):

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1 - p_i}\right) = \mathbf{d}_i^T \boldsymbol{\theta} + \delta_i, \quad \delta_i \sim \mathcal{N}(0, \tau^2) \quad (8)$$

onde  $\mathbf{d}_i$  são os coeficientes de ondaletas associados à  $i$ -ésima observação,  $\boldsymbol{\theta}$  é o vetor de coeficientes da regressão e  $\delta_i$  é o componente residual aleatório. Por meio do modelo da equação (8) é possível estabelecer uma relação entre a variável classificadora  $Z_i$  e os coeficientes de ondaletas estimados.

**Exemplo 3.** Utilizando a função SpaHet e uma modificação sua (Figura 4), referidas por  $C_1$  e  $C_2$ , respectivamente, foi feita uma simulação com 10000 observações, metade pertencente a cada classe, e uma razão sinal-ruído de 3, o resultado pode ser encontrado na Tabela 1.

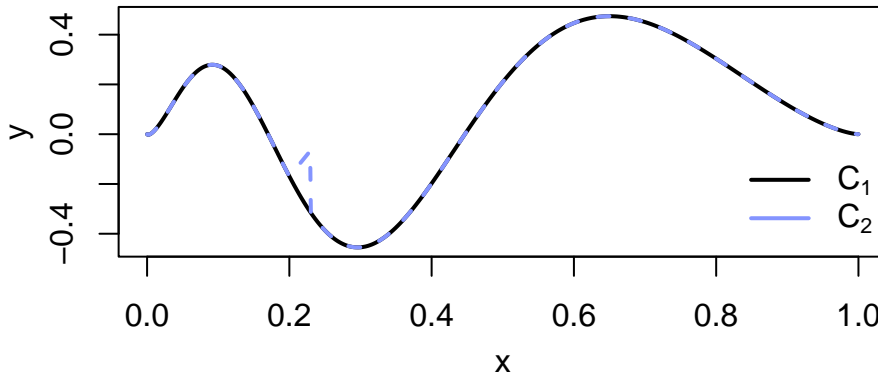


Figura 4: Classes  $C_1$  e  $C_2$

Tabela 1: Predição de 10000 observações

		Classe Verdadeira		
		$C_1$	$C_2$	Total
Classificador	$C_1$	3957	1131	5088
	$C_2$	1043	3869	4912
	Total	5000	5000	10000

## 3 Conclusão

Portanto, por meio da equação (8) e das técnicas apresentadas, é possível estimar a função por meio das ondaletas, como apresentado no Exemplo 2. Além disso, também é possível utilizar esses coeficientes para estabelecer uma relação entre a probabilidade de uma curva  $\mathbf{Y}_i$  pertencer à uma determinada classe, por meio dos seus coeficientes de ondaletas estimados, de forma a conseguir classificar curvas, como apresentado no Exemplo 3.

## Referências

- Hastie, T., Tibshirani, R. e Wainwright, M., (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press.
- Hastie, T., Tibshirani, R e Friedman, J., (2008). *The Elements of Statistical Learning*. Segunda edição. Springer.
- James, G., Witten, D., Hastie, T. e Tibshirani, R., (2014). *An Introduction to Statistical Learning with Applications in R*. Springer.
- Nason, G. P., (2008). *Wavelet Methods in Statistics with R*. Springer.
- Wang X. *et al.* (2007). *Bayesian Curve Classification Using Wavelets*, Journal of the American Statistical Association, 102:479, 962-973, DOI: 10.1198/016214507000000455
- Vidakovic, B., (1999). *Statistical Modeling by Wavelets*. Wiley, New York.