Referências Bibliográficas

[1] BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis: Forcasting and Control**. San Francisco: Holden-Day, San Francisco, revised edition, 1976.

[2] BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R. ; STONE, C. J. **Classification and Regression Trees**. Wadsworth and Brooks, Monterey, CA, 1984.

[3] CHAN, K. S.; TONG, H. **On estimating threshols in autoregressive models**. J. Times Series Anal., 7:179–190, 1986.

[4] CHANDHURI, P.; LO, W.; LOH, W. Y. ; YANG, C. C. **Generalized regression trees**. Statistica Sinica, 5:641–666, 1995.

[5] CIAMPI, A. **Generalized regression trees**. Computational Statistics and Data Analysis, 12:57–78, 1991.

[6] COVER, T. M.; HART, P. E. **Nearest neighbor pattern classification**. IEEE Transactions on Information Theory, 13:21–27, 1967.

[7] DA ROSA, J. C.; VEIGA, A. ; MEDEIROS, M. C. **Tree-structured smooth transition regression models**. Computational Statistics and Data Analysis, 52:2469–2488, 2008.

[8] FAN, J.; YAO, Q. **Nonlinear Time Series: Nonparametric and Parametric Methods**. New York, 2003.

[9] FIX, E.; HODGES, J. L. **Nonparametric discrimination: Consistency properties**. Randolph Field, Texas, USA, 1951.

[10] GRANGER, C.; TERÄSVIRTA, T. **Modelling Nonlinear Economic Relationships**. Oxford, UK, 1993.

[11] GUISAN, A.; EDWARDS, T. ; HASTIE, T. **Generalized linear and generalized additive models in studies of species distributions: setting the scene**. Ecological Modeling, 157:89–100, 2002.

[12] HASTIE, T.; TIBSHIRANI, R. **Generalized additive models**. Statistical Science, 1(3):297–318, 1986.

[13] HASTIE, T.; TIBSHIRANI, R. **Generalized additive models**. London, UK, 1990.

[14] HASTIE, T.; TIBSHIRANI, R. ; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. New York, USA, 2001.

[15] HOSMER, D.; LEMESHOW, S. **Applied Logistic Regression**. New York, USA, 2nd edition, 2000.

[16] JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. London, UK, 5th edition, 2002.

[17] LEWIS, P. A. W.; STEVENS, J. G. **Nonlinear modelong of time series using multivariate adaptative regression splines**. Journal of American Statistical Association, p. 864–877, 1986.

[18] LUUKKONEN, R.; SAIKKONEN, P. ; TERÄSVIRTA, T. **Testing linearity againt smooth transition autoregressive models**. Biometrika, 75:491–499, 1988.

[19] MCCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. London, UK, 2nd edition, 1989.

[20] MEDEIROS, M.; VEIGA, A. **Diagnostic checking in a flexible nonlinear time series model**. Journal of Time Series Analysis, 24:461–482, 2003.

[21] MEDEIROS, M. C.; VEIGA, A. **A hybrid linear-neural model for time series forecasting**. IEEE Transactions on Neural Networks, 11(6):1402–1412, 2000.

[22] MEDEIROS, M. C.; VEIGA, A. **A flexible coefficient smooth transition time series models**. IEEE Transactions on Neural Networks, 16(1):97–113, 2005.

[23] MINGOTI, S. A. **Análise de Dados Através de Métodos de Estatística Multivariada**. Belo Horizonte, BR, 2005.

[24] NELDER, J. A.; WEDDERBURN, R. W. M. **Generalized linear models**. Journal of the Royal Statistical Society. Series A (General), 135(3):370–384, 1972.

[25] ORTEGA, G. V. C. **Redes neurais na identificação de perdas comerciais do setor elétrico**. Dissertação de Mestrado, Departamento de Engenharia Elétrica - PUC-Rio, Rio de Janeiro, Brasil, 2008.

[26] RECH, G.; TERÄSVIRTA, T. ; TSCHERNING, R. **A simple variable selection technique for nonlinear models**. Commun. Statist., Theory and Methods, 30:1227–1241, 2001.

[27] STONE, C. **Additive regression and other nonparametric models**. Ann. Statist, 13:689–705, 1985.

[28] TONG, H. **On a threshold model**. 1978. Leyden, Netherlands: Sijthoff and Noordhoff, 1978.

[29] TONG, H. **Non-linear Time Series: A Dynamical System Approach**, volumen 6. Oxford Statistical Science Series, Oxford, 1990.

[30] TONG, H.; LIM, K. **Threshold autoregression, limit cycles and cyclical data**. Royal Statistical Society, Series B, Methodological, (42):245–292, 1980.

[31] VAN DIJK, D.; FRANSES, P. **Modelling multiple regimes in the business cycle**. Macroeconomic Dynamics, 3:311–340, 1999.

[32] VAN DIJK, D.; TERÄSVIRTA, T. ; FRANSES, P. H. **Smooth transition autoregressive models - a survey of recent developments**. Econometric Rev., 21:1–47, 2002.

[33] WHITE, H. **A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity**. Econometrica, 48(4):817–838, 1980.

[34] YULE, G. U. **On a method of investigating periodicities in disturbed series with special reference to wolfer´s sunspot numbers**. Philosophical Transactions of th Royal Society, 226:267–298, 1927.

A
Alguns Modelos Não-lineares

Serão brevemente explicados alguns dos modelos não-lineares existentes na literatura, muitos deles utilizado na análise de séries temporais não enquadrados no contexto de classificação, porém foram escolhidos aqueles que, de alguma maneira, estão relacionados com este trabalho em sua forma estruturas, estimação e/ou previsão.

A.1
Threshold Auto Regressive (TAR)

Proposto em (28) e mais tarde desenvolvido em (30) e extensamente discutido em (29).

A idéia principal do modelo TAR é a de mudar os parâmetros de um modelo linear Auto Regressivo (AR) (ver (34) e (1)), de acordo com o valor de uma variável observável chamada variável de transição ou limiar (do inglês *threshold variable*).

Basicamente fazem uma divisão do espaço Euclidiano unidimensional de modo a obter *L* regimes, os quais são liderados por um modelo Auto Regressivo de odem $k_i, \quad i = 1, ..., L$. Tal divisão é feita de forma abrupta sendo regida por uma função indicadora, $I(z_t)$.

Formulação Matemática

$$y_t = \begin{cases} \beta_0^{(1)} + \sum_{i=1}^{k_1} \beta_i^{(1)} y_{t-i} + \epsilon_t^{(1)} & , \ se \ z_t \in \mathbb{R}_1 \\ \beta_0^{(2)} + \sum_{i=1}^{k_2} \beta_i^{(2)} y_{t-i} + \epsilon_t^{(2)} & , \ se \ z_t \in \mathbb{R}_2 \\ \vdots & \\ \beta_0^{(L)} + \sum_{i=1}^{k_L} \beta_i^{(L)} y_{t-i} + \epsilon_t^{(L)} & , \ se \ z_t \in \mathbb{R}_L \end{cases}$$

$$y_t = \sum_{j=1}^{L} \left[\beta_0^{(j)} + \sum_{i=1}^{p} \beta_i^{(j)} y_{t-i} + \epsilon_t^{(j)} \right] I^{(j)}(z_t) \tag{A-1}$$

onde z_t é a variável de transição, $\epsilon_t^{(j)} \rightarrow (0, \sigma^2)$ o erro aleatório (ruído branco) e o vetor de parâmetros lineares $\boldsymbol{\beta} = (\beta_0^{(j)}, \beta_1^{(j)}, ..., \beta_p^{(j)})'$. A função indicadora é tal que

$$I^{(j)}(z_t) = \begin{cases} 1 & , se\ z_t \in \mathbb{R}_j \\ 0 & , se\ z_t \notin \mathbb{R}_j. \end{cases}$$

A variável de transição pode ser governada pelo tempo ($z_t = t$), por uma variável exógena ($z_t = x_{t-d}$) ou ainda por um valor defasado da variável dependente, ou seja, um auto-regressor de y_t ($z_t = y_{t-d}$). A letra d representa o parâmetro de defasagem.

A.2 Self-Exiting Threshold Auto Regressive (SETAR)

A escolha da variável de transição como um auto-regressor de y_t caracteriza um modelo SETAR (ver (29)), o qual, da mesma forma que o TAR, divide o espaço das variáveis de forma abrupta, em subespaços ortogonais a somente um auto-regressor, y_{t-d}.

Formulação Matemática

$$y_t = \begin{cases} \beta_0^{(1)} + \sum_{i=1}^{k_1} \beta_i^{(1)} y_{t-i} + \epsilon_t^{(1)} & ,\ se\ y_{t-d} \in \mathbb{R}_1 \\ \beta_0^{(2)} + \sum_{i=1}^{k_2} \beta_i^{(2)} y_{t-i} + \epsilon_t^{(2)} & ,\ se\ y_{t-d} \in \mathbb{R}_2 \\ \vdots & \\ \beta_0^{(L)} + \sum_{i=1}^{k_L} \beta_i^{(L)} y_{t-i} + \epsilon_t^{(L)} & ,\ se\ y_{t-d} \in \mathbb{R}_L \end{cases}$$

$$y_t = \sum_{j=1}^{L} \left[\beta_0^{(j)} + \sum_{i=1}^{p} \beta_i^{(j)} y_{t-i} + \epsilon_t^{(j)} \right] I^{(j)}(y_{t-d}) \qquad \text{(A-2)}$$

onde y_{t-d} é a variável de transição e

$$I^{(j)}(y_{t-d}) = \begin{cases} 1 & , se\ y_{t-d} \in \mathbb{R}_j \\ 0 & , se\ y_{t-d} \notin \mathbb{R}_j. \end{cases}$$

A.3 Smooth Transition Autoregression (STAR)

Uma alteração no modelo SETAR proposta por (3), onde passamos de uma transição abrupta para uma transição suave, substituindo a função indicadora por uma função não-linear, contínua e limitada entre 0 e 1 denominada por $G(z_t; \gamma, c)$. Tais modelos limitam-se a dois regimes apenas.

Formulação Matemática

$$y_t = \beta_0^{(1)} + \beta_1^{(1)} y_{t-1} + ... + \beta_{k1}^{(1)} y_{t-k_1} + \left(\beta_0^{(2)} + \beta_1^{(2)} y_{t-1} + ... + \beta_{k1}^{(2)} y_{t-k_2}\right) G(z_t;\gamma,c) + \epsilon_t$$

$$y_t = \beta_0^{(1)} + \sum_{i=1}^{k_1} \beta_i^{(1)} y_{t-i} + \left(\beta_0^{(2)} + \sum_{j=1}^{p} \beta_j^{(2)} y_{t-j}\right) G(z_t;\gamma,c) + \epsilon_t$$

A formulação para dois regimes pode ser expressa, quando $z_t = y_{t-1}$, por

$$y_t = (\alpha_0 + \beta_0 y_{t-1}) G(y_{t-1};\gamma,c) + (\alpha_1 + \beta_1 y_{t-1})[1 - G(y_{t-1};\gamma,c)] + \epsilon_t$$

onde $G(y_{t-1};\gamma,c)$ é a função de transição, γ é chamado de parâmetro de suavização e c de parâmetro de localização ou limiar.

Especificação da Função de Transição

Nos modelos de transição suave pode-se especificar a função de transição de modo a modelar os dados sem assumir inicialmente que haverá uma mudança abrupta entre os regimes. As funções podem ser escolhidas como, por exemplo:

- Função Logística: $G(z_t;\gamma,c) = \frac{e^{-\gamma(z_t-c)}}{1+e^{-\gamma(z_t-c)}}, \quad \gamma > 0;$
- Função Exponencial: $G(z_t;\gamma,c) = 1 - e^{-\gamma(z_t-c)}, \quad \gamma > 0.$

Dependendo da função de transição e dos valores do parâmetro de suavização da mesma, o STAR é definido de formas diferentes.

A.4
Logistic Smooth Transition Autoregression (LSTAR)

Quando a função de transição utilizada para suavizar a mudança entre os regimes for a função logística, trata-se do modelo LSTAR (ver (18)).

A variação no valor do parâmetro do grau de suavidade da função, γ, remete a casos particulares do modelo LSTAR.

Se $\gamma \to \infty$ a função de transição toma a forma de uma função degrau, ou seja

$$G(z_t;\gamma,c) = \begin{cases} 1 & , se\ z_t \leq c \\ 0 & , se\ z_t > c \end{cases}$$

Esta situação caracteriza um modelo TAR em que o limiar é regido e determinado por c. Caso $z_t = c$ a observação fará parte de ambos os regimes com o mesmo grau de pertinência.

No caso em que $\gamma \to 0$ a função logística é igual a $0,5$ e teremos um processo Auto Regressivo (AR).

A.5 Exponencial Smooth Transition Autoregression (ESTAR)

Ao utilizarmos uma função de transição exponencial teremos um modelo denominado ESTAR (ver (10)).

Neste, tanto para valores de $\gamma \to \infty$ quanto para $\gamma \to 0$, teremos um modelo AR.

A.6 Multiple Regime Smooth Transition Autoregression (MRSTAR)

Como o modelo STAR abrange somente até dois regimes, vemos em (31) o tratamento da multiplicidade de regimes através do modelo MRSTAR.

Formulação Matemática

Para um modelo com 4 regimes, por exemplo, teremos a representação de um MRSTAR onde $z_t = y_{t-1}$ dada por

$$
\begin{aligned}
y_t = \quad & \{(\alpha_0 + \beta_0 y_{t-1})G_0(y_{t-1};\gamma,c) + (\alpha_1 + \beta_1 y_{t-1})[1 - G_0(y_{t-1};\gamma,c)]\} \\
& +\{(\alpha_2 + \beta_2 y_{t-1})G_1(y_{t-1};\gamma,c) + (\alpha_3 + \beta_3 y_{t-1})[1 - G_1(y_{t-1};\gamma,c)]\} + \epsilon_t
\end{aligned}
$$

A.7 Neural Coefficient Smooth Transition Autoregressive (NCSTAR)

Ainda no contexto dos modelos de múltiplos regimes cabe destacar aquele proposto em (21). Trata-se de um modelo híbrido, pois mescla os parâmetros autoregressivos de um STAR, os quais variam ao longo do tempo conforme a saída de uma Rede Neural Artificial (RNA).

O modelo STR-Tree, principal enfoque deste trabalho, tem seus processos de especificação e estimação muito semelhantes ao que é feito no caso do NCSTAR.

As técnicas de diagnóstico do ajuste de um NCSTAR encontram-se em (20) e o tema volta a ser abordado em (22).

Formulação Matemática

$$
y_t \quad = G(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\psi}) + \epsilon_t
$$

$$= \alpha_0 + \sum_{j=1}^{p} \alpha_j y_{t-j} + \sum_{i=1}^{h} \lambda_{0i} F(\boldsymbol{\omega}_i' \mathbf{x}_t - c_i)$$
$$+ \sum_{j=1}^{p} \left\{ \sum_{i=1}^{h} \lambda_{ji} F(\boldsymbol{\omega}_i' \mathbf{x}_t - c_i) \right\} y_{t-j} + \epsilon_t$$

que tem a forma vetorial dada

$$y_t = \boldsymbol{\alpha}' \mathbf{z}_t + \sum_{i=1}^{h} \boldsymbol{\lambda}' \mathbf{z}_t F(\boldsymbol{\omega}_i' \mathbf{x}_t - c_i),$$

onde $\boldsymbol{\psi} = (\boldsymbol{\alpha}', \boldsymbol{\lambda}_1', ..., \boldsymbol{\lambda}_h', \boldsymbol{\omega}_1', ..., \boldsymbol{\omega}_h', c_i, ..., c_h)' \in \mathbb{R}^r$ é o vetor de parâmetros, $r = (q+1)h + (p+1)(h+1)$, $\boldsymbol{\alpha} = (\alpha_0, ..., \alpha_p)' = (-\lambda_{00}, ..., \lambda_{p0})'$ e $\boldsymbol{\lambda}_i = (\lambda_{0i}, ..., \lambda_{pi})'$. $F(\boldsymbol{\omega}_j' \mathbf{x}_t - c_i)$ é a função de transição logística onde $\mathbf{x}_t \in \mathbb{R}^q$ é o vetor de variáveis de transição, $\boldsymbol{\omega}_i = (\omega 1i, ..., \omega_{qi})' \in \mathbb{R}^q$ e $c_i \in \mathbb{R}$ são os parâmetros não-lineares.

A.8
Smooth Transition Regression (STR)

Semelhante ao STAR onde a transição não é mais governada por autorregressores de y_t e sim por outras variáveis explicativas, $\mathbf{x}_t$, as quais podem ser dependentes do tempo, no caso de estarmos tratando de séries temporais, ou simplesmente covariáveis independentes do tempo e entre si, que serão as regressões não-lineares.

Esses modelos também acompanham as variações de acordo com o tipo de funções de ligação onde teremos os modelos LSTR e ESTR. No caso de múltiplos regimes o MSTR (ver (31)).

O modelo STR e suas variações tem sua especificação, estimação e avaliação extensamente comentadas em (10).

B
Comando do programa R 2.6.2

Serão mostrados os comandos utilizados para o caso de Fraude/Irregularidade no Consumo de Energia Elétrica.

B.1
Comandos para GAM

```
library(VGAM)

gam1 = vgam(y ∼ s(x15,df=4)+s(x16,df=4)+s(x17,df=4)+s(x18,df=4)+s(x19,df=4)+
+s(x20,df=4)+s(x21,df=4)+s(x22,df=4)+s(x23,df=4)+s(x25,df=4)+s(x26,df=4)
+s(x27,df=4),binomialff, lt)

phat=fitted.values(gam1)
Dev=deviance(gam1)
betas=as.matrix(coefficients(gam1)
cutoff=0.5 yhat=lt$y dim(yhat)=c(length(lt$y),1)
for (i in 1:length(lt$y))if (phat[[i]]>=cutoff) yhat[[i]]=1 else yhat[[i]]=0
phat_out=predict.vglm(gam1, newdata=nd, type="response")
dim(phat_out)=c(length(nd$y),1) yhat_out=nd$y dim(yhat_out)=c(length(nd$y),1)
for (i in 1:length(nd$y))if (phat_out[[i]]>=cutoff) yhat_out[[i]]=1 else
yhat_out[[i]]=0
```

B.2
Comandos para CART

```
library(tree)

cart1 = tree(y  x15+x16+x17+x18+x19+x20+x21+x22+x23+x25+x26+x27+x29+x30+x31,
data=lt)

plot(cart1) text(cart1) cv.cart1=cv.tree(cart1,rand=1:10,K=10,FUN=prune.tree)
prune.cart1=prune.tree(cart1,best=NULL)
phat=predict(cart1) yhat=predict(cart1, type=c("class")) dim(yhat)=c(length(lt$y),1)
```

B.3
Comandos para k-NN

library(kknn)

lt=as.data.frame(light[1:2430,]) nd=as.data.frame(light[-c(1:2430),])

knn1.learn ¡- lt knn1.valid ¡- nd

modknn1 = kknn(y x15+x16+x17+x18+x19+x20+x21+x22+x23+x25+x26+x27+ +x29+x30+x31,knn1.learn,knn1.valid)

phat=fitted(modknn1) table(knn1.valid$y, phat)

cutoff=0.5 yhat=lt$y dim(yhat)=c(length(lt$y),1)

for (i in 1:length(lt$y))if (phat[[i]]>=cutoff) yhat[[i]]=1 else yhat[[i]]=0

phat_out = predict(modknn1, newdata = nd, type = "response")

dim(phat_out)=c(length(nd$y),1) yhat_out=nd$y dim(yhat_out)=c(length(nd$y),1)

for (i in 1:length(nd$y))if (phat_out[[i]]>=cutoff) yhat_out[[i]]=1 else yhat_out[[i]]=0

B.4
Comandos para Regressão Logística

lt = as.data.frame(light[1:2430,]) nd = as.data.frame(light[-c(1:2430),])

glm1=glm(y x15+x16+x17+x18+x19+x20+x21+x22+x23+x25+x26+x27+x29+x30+x31, family=binomial(link=logit), data=lt)

X=model.matrix(glm1) betas=as.matrix(coefficients(glm1)) Dev=deviance(glm1)

phat=fitted(glm1) dim(phat)=c(length(lt$y),1)

cutoff=0.5 yhat=lt$y dim(yhat)=c(length(lt$y),1)

for (i in 1:length(lt$y))if (phat[[i]]>=cutoff) yhat[[i]]=1 else yhat[[i]]=0

phat_out=predict(glm1, newdata=nd, type="response")

dim(phat_out)=c(length(nd$y),1) yhat_out=nd$y dim(yhat_out)=c(length(nd$y),1)

for (i in 1:length(nd$y))if (phat_out[[i]]>=cutoff) yhat_out[[i]]=1 else yhat_out[[i]]=0

C
Estatísticas Descritivas

C.1
E-mail/Spam

Tabela C.1: Estatísticas Descritivas - Spam

	Máximo	Mínimo	Média	Variância	Desvio Padrão
our	10.00	0.00	0.32	0.49	0.70
over	5.88	0.00	0.09	0.08	0.28
remove	5.40	0.00	0.12	0.16	0.40
internet	4.68	0.00	0.10	0.13	0.35
free	20.00	0.00	0.27	0.91	0.95
business	5.12	0.00	0.15	0.21	0.45
hp	18.18	0.00	0.55	2.90	1.70
hpl	10.86	0.00	0.27	0.80	0.89
george	33.33	0.00	0.75	10.92	3.30
1999	4.54	0.00	0.13	0.14	0.37
remove	20.00	0.00	0.28	0.97	0.99
edu	22.05	0.00	0.21	1.04	1.02
char_!	19.13	0.00	0.28	0.57	0.75
char_$	5.30	0.00	0.08	0.06	0.24
CAPMAX	2204.00	1.00	50.31	18348.72	135.46
CAPTOT	9163.00	1.00	267.41	265164.09	514.94

C.2
Doenças Cardíacas na África do Sul

Tabela C.2: Estatísticas Descritivas - DCAS

	Máximo	Mínimo	Média	Variância	Desvio Padrão
sbp	218.00	101.00	138.33	420.10	20.50
tobacco	31.20	0.00	3.64	21.10	4.59
ldl	15.33	0.98	4.74	4.29	2.07
obesity	46.58	14.70	26.04	17.76	4.21
alcohol	147.19	0.00	17.04	599.32	24.48
age	64.00	15.00	42.82	213.42	14.61

C.3
Fraude/Irregularidade no Consumo de Energia Elétrica

Tabela C.3: Estatísticas Descritivas - Fraude no Consumo de Energia

	Máximo	Mínimo	Média	Variância	Desvio Padrão
consumo	0.92	0.00	0.32	0.07	0.27
comsumo_ano_ant	0.94	0.00	0.36	0.07	0.26
consumo_ano_base	51.83	0.00	0.63	3.41	1.85
media_3	0.91	0.00	0.33	0.06	0.24
media_6	0.91	0.00	0.34	0.04	0.21
media_12	0.91	0.00	0.34	0.03	0.18
media_12_24	0.91	0.00	0.37	0.03	0.18
indic_trimestral_1	1.00	0.00	0.09	0.04	0.19
indic_trimestral_2	1.00	0.00	0.09	0.03	0.19
indic_trimestral_3	1.00	0.00	0.09	0.04	0.19
indic_anual	1.00	0.00	0.16	0.06	0.25
indic_ajuste	1.00	0.00	0.20	0.09	0.30
indic_tendencia	1.00	0.00	0.24	0.11	0.33
temperatura_min	0.81	0.33	0.51	0.02	0.13
temperatura_max	0.82	0.29	0.44	0.01	0.11
carga	0.70	0.19	0.36	0.02	0.13

D
Estimativas dos Coeficientes

D.1
E-mail/Spam

Tabela D.1: Coeficientes - Spam

	GAM	Regressão Logística	Análise Discrimina	STLR-Tree (REGIME 1)	STLR-Tree (REGIME 2)
Intercepto	-1.3979	-1.577	-	-0.391	-0.014
our	0.0110	0.010	0.250	-3.153	11.598
over	0.0184	0.018	0.243	-10.772	-7.960
remove	0.0350	0.042	0.414	9.870	12.162
internet	0.0157	0.020	0.243	-5.669	-1.185
free	0.0141	0.015	0.257	3.111	2.307
business	0.0130	0.012	0.142	0.037	-0.962
hp	-0.0313	-0.025	-0.205	2.401	-5.604
hpl	0.0020	-0.019	-0.122	1.802	0.145
george	-0.0428	-0.060	-0.210	-2.879	-328.586
1999	-	-	-0.148	-	-
remove	-0.0164	-	-0.149	-7.036	13.064
edu	-0.0198	-0.014	-0.168	-14.609	-7.917
char_!	0.0291	-0.034	0.266	3.451	-18.344
char_$	0.0819	0.027	0.328	19.250	14.834
CAPMAX	-0.0051	0.096	0.125	1.528	0.014
CAPTOT	0.0002	-0.003	0.282	0.017	0.017

D.2
Doenças Cardíacas na África do Sul

Tabela D.2: Coeficientes - DCAS

	GAM	Regressão Logística	Análise Discrimina	STLR-Tree (REGIME 1)	STLR-Tree (REGIME 2)
Intercepto	-4.1713	-4.3762	-	130.491	-3.629
sbp	0.0067	0.0051	0.104	-3.583	0.002
tobacco	0.0027	0.0033	0.389	2.664	0.081
ldl	0.0029	0.0026	0.393	6.386	-0.004
alcohol	-0.0003	-0.0001	0.016	-8.911	0.002
k-NN	0.0440	0.0514	0.563	-3.926	0.051

D.3
Fraude/Irregularidade no Consumo de Energia Elétrica

Tabela D.3: Coeficientes - Fraude no Consumo de Energia

	GAM	Regressão Logística	Análise Discrimina	STLR-Tree (REGIME 1)	STLR-Tree (REGIME 2)	STLR-Tree (REGIME 3)	STLR-Tree (REGIME 4)	STLR-Tree (REGIME 5)
Intercepto	0.52	-1.23	-	-2.38	-0.54	-3.86	33.85	-16.84
consumo	1.47	-	0.14	6.40	33.76	7.82	-35.36	-0.69
consumo_ano_ant	0.27	-	0.19	-1.48	-23.65	-1.65	-32.45	-0.23
consumo_ano_base	0.00	0.00	0.05	35.20	-6.64	1.03	15.22	-0.01
media_3	-0.83	-	-0.44	0.28	0.46	-0.11	0.98	0.98
media_6	-2.76	-1.55	-0.55	0.14	-7.70	-0.03	-19.19	-2.61
media_12	0.98	1.13	0.48	-0.78	-0.12	-3.32	6.09	1.51
media_12_24	-0.45	-0.77	-0.36	0.03	-1.54	6.50	40.83	-0.06
indic_trimestral_1	1.24	0.60	0.35	-	-	-	-	-
indic_trimestral_2	1.04	0.61	0.27	-	-	-	-	-
indic_anual	-0.34	-	0.10	-1.51	0.15	-5.74	-75.25	-0.72
indic_ajuste	-0.19	-	-0.34	-0.55	3.56	1.96	35.18	-0.24
indic_tendencia	0.14	-	-0.08	2.79	-4.65	-0.92	-25.79	-0.15
temperatura_min	-	5.27	1.34	-1.62	3.62	0.60	-18.00	1.90
temperatura_max	-	-	0.01	0.95	-0.59	-0.36	-29.10	0.34
carga	-	-2.76	-0.70	0.59	0.74	-4.70	56.38	-1.23

Tabela D.4: Pesos Redes Neurais - Fraude no Consumo de Energia

	Neurônio 1	Neurônio 2	Neurônio 3	Neurônio 4	Neurônio 5	Neurônio 6	Neurônio 7	Neurônio 8
consumo	-1.13	-31.57	-1.96	5.49	-3.29	2.38	2.91	-3.51
comsumo_ano_ant	0.37	-6.63	-9.18	-5.30	3.31	-0.48	1.11	-3.30
consumo_ano_base	-0.04	3.68	-2.92	-1.66	1.20	-0.19	-0.19	0.33
media_3	10.60	20.32	5.88	-3.43	1.59	-4.19	-8.30	10.77
media_6	-23.91	40.65	-1.44	-5.18	3.62	-0.02	12.74	-17.20
media_12	4.05	70.20	24.88	3.61	-2.35	-1.29	-4.75	8.71
media_12_24	0.38	-4.99	-12.80	-1.84	1.61	1.26	-5.31	5.01
indic_trimestral_1	8.61	-27.56	-0.36	2.37	-2.55	2.60	-1.82	1.02
indic_trimestral_2	2.65	18.67	5.08	5.79	-3.97	-0.21	-3.29	4.99
indic_trimestral_3	10.73	-26.45	0.05	4.47	-2.38	0.69	1.37	0.41
indic_anual	5.15	-20.05	5.52	2.48	-2.14	-2.34	8.68	-9.34
indic_ajuste	9.35	21.16	0.59	-1.99	1.48	-0.41	0.51	-0.77
indic_tendencia	-3.57	-0.91	2.39	0.83	-0.65	0.54	-0.14	0.61
temperatura_min	44.50	5.67	3.68	-24.13	18.24	-79.18	4.18	-2.26
temperatura_max	22.60	-1.91	42.95	21.04	-17.69	74.52	-17.42	21.25
carga	-48.36	-7.46	-31.66	-5.02	4.59	-8.68	13.64	-19.26

D.4
Coeficientes dos parâmetros não-lineares

Tabela D.5: Coeficientes Não-lineares

		c	γ
E-mail/Spam	1ª divisão	0.00	50.00
DCAS	1ª divisão	4.08	50.00
Consumo Energia	1ª divisão	3.54	10.22
	2ª divisão	2.82	50.03
	3ª divisão	3.74	50.00
	4ª divisão	3.57	99.99