

## 7

### Resultados

Neste capítulo, utilizaremos os mesmos dados sintéticos gerados para realizar os experimentos no capítulo 4. Testaremos a aproximação implícita da variedade com duas subdivisões espaciais diferentes e, com a estrutura da árvore definida, testaremos a capacidade preditiva do modelo, comparando-o com alguns métodos conhecidos para regressão. O modelo RCVI será ainda utilizado para realizar regressão com três conjuntos distintos de dados reais e uma aplicação em geologia.

#### 7.1

##### Dados sintéticos

Para cada experimento, indicamos a subdivisão espacial: malha  $2^n$ -ádica ou Árvore BSP, a função peso utilizada: polinomial ou gaussiana, o grau do polinômio: 1, 2 ou 3, fornecemos uma cota superior para o erro de aproximação local para a função  $F_i, i = 1, \dots, n - m$ , onde  $n$  indica a dimensão do espaço onde a variedade está e  $m$  é a dimensão da variedade estudada, determinamos o nível máximo da árvore, isto é, a profundidade máxima admitida pela árvore e definimos uma cota superior para o desvio padrão dos valores de  $y$  em cada folha.

Após a árvore construída, verificamos o número de folhas obtido e os critérios de parada ocorridos, os quais são classificados do seguinte modo: critério de parada 0 significa que o nível máximo da árvore foi atingido; critério de parada 1, obtivemos o erro local  $<$  cota estabelecida; critério de parada 2, número de dados insuficiente para aproximar uma nova função aos dados; critério de parada 3, o desvio padrão dos valores de  $y$  é menor do que a cota estabelecida.

Calculamos ainda a aproximação obtida para cada função  $F_i, i = 1, \dots, n - m$  para todos os dados de entrada do seguinte modo:

$$\sum_{j=1}^N F_i(x_j) = \sum_{j=1}^N \sum_{k=1}^L \varphi_k(x_j) F_{ik}(x_j), \quad i = 1, \dots, n - m.$$

Assim, o resultado exibido nas tabelas com relação ao erro de aproximação da  $F_i$ , é dado por:

$$\frac{\sum_{j=1}^N F_i(x_j)^2}{N}, \quad i = 1, \dots, n - m.$$

Com relação a subdivisão espacial dada pela Árvore BSP temos ainda que determinar o valor do parâmetro  $\lambda$ . Essa escolha foi realizada a partir de um conjunto de valores para  $\lambda$  variando de 0.5 a 15. A técnica empregada foi a 5-validação cruzada (*5-fold cross-validation*), onde 5 define a quantidade de execuções a ser realizada e também o número de divisões das amostras em treinamento e teste, que serão usadas durante a validação. Assim, todas as amostras são usadas para treinamento e teste. Para cada divisão dos dados em dados treino/dados teste, uma árvore é construída com os dados treino e uma regressão utilizando a estimativa global é realizada, testada no conjunto teste e o erro quadrado médio é calculado. O parâmetro  $\lambda$  que apresentar o menor erro quadrado médio é escolhido.

Por fim, uma vez que temos a árvore construída, realizamos uma regressão. Para cada dado de entrada associamos a saída  $y = f(\mathbf{x}) = x_1x_2 + x_3$ . Realizamos uma regressão com os dados dentro da amostra (regressão *in-sample*) e, em seguida, realizamos uma regressão com os dados fora da amostra, isto é, *out-of-sample*. Neste segundo caso, os dados foram divididos, de forma aleatória, em 10 partes iguais e, 9 dessas partes foram usadas para treino e 1 parte para teste. O teste foi repetido 10 vezes e em cada rodada avaliamos os erros obtidos pela estimativa local, chamada  $y_{\text{local}}$  e estimativa global, denotada de  $y_{\text{global}}$ , conforme definidos no capítulo 6, utilizando as estatísticas do erro quadrado médio (EQM) e do erro percentual absoluto médio (MAPE):

$$EQM = \frac{\sum_{j=1}^N (y_i - \hat{y}_j)^2}{N} \quad MAPE = \frac{1}{N} \sum_{j=1}^N \left| \frac{y_j - \hat{y}_j}{y_j} \right|$$

onde  $y_i$  representa o valor de  $y$  real e  $\hat{y}_j$  o valor previsto de  $y$ .

Calculamos ainda a estimativa MAD (*Median Absolute Deviation*) para cada uma das estatísticas acima. O MAD é definido como a mediana dos desvios absolutos da mediana dos dados:

$$MAD = \text{Mediana}_i(|X_i - \text{mediana}_j(X_j)|),$$

para um conjunto de dados univariados  $X_1, X_2, \dots, X_n$ .

Para avaliar o desempenho do RCVI, utilizamos o CART, tradicional método de regressão em árvores e os métodos de aprendizagem supervisionada: SVR e *Bayesian Ridge Regression* (BayesianRR). Para implementar o método CART utilizamos o pacote *tree* (55) do software livre R (59). Este pacote foi escrito por Ripley, cuja obra (56) apresenta um capítulo dedicado ao tema de previsão com árvores, além de ser uma boa referência em mineração de dados. Para implementar o SVR e o BayesianRR utilizamos o módulo Python *scikit-learn*, o qual, no caso do SVR utiliza a biblioteca LIBSVM (12). Para os parâmetros do SVR empregamos os valores padrões: a função núcleo é a *radial basis function* ( $e(-\gamma \|x - x'\|^2)$ ,  $\gamma > 0$ ),  $\gamma = 0.5$ ,  $C = 1.0$ ,  $\text{cache\_size} = 200$ ,  $\text{coef0} = 0.0$ ,  $\text{degree} = 3$ ,  $\text{epsilon} = 0.1$ ,  $\text{probability} = \text{False}$ ,  $\text{scale\_C} = \text{False}$ ,  $\text{shrinking} = \text{True}$ ,  $\text{tol} = 0.001$ . Os valores padrões para o método BayesianRR são:  $\alpha_1 = \alpha_2 = \lambda_1 = \lambda_2 = 10^{-6}$ ,  $\text{tol} = 0.001$ ,  $n\_iter = 300$ ,  $\text{normalize} = \text{False}$ ,  $\text{fit\_intercept} = \text{True}$ ,  $\text{verbose} = \text{False}$ . Tanto o método SVR quanto o BayesianRR foram brevemente explicados no capítulo 3.

### 7.1.1

#### Curva no $\mathbb{R}^3$

Geramos  $N = 3000$  dados, tomamos 15 vizinhos e  $\sigma = 3$ . A tabela 7.1 exhibe os resultados da estimativa da dimensionalidade correta (DE) e o erro de orientação obtidos.

Curva no $\mathbb{R}^3$	
DE	100%
mediana dos ângulos (graus)	0.45
média dos ângulos (graus)	0.50

Tabela 7.1: Curva no  $\mathbb{R}^3$ . Taxa da estimativa correta da dimensionalidade (DE). As estatísticas dos ângulos referem-se aos erros de orientação obtidos.

#### Resultados na aproximação da variedade implícita

As tabelas a seguir mostram os resultados obtidos na aproximação da função  $F : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ , onde  $F_i : \mathbb{R}^3 \rightarrow \mathbb{R}$ ,  $i = 1, 2$  e  $F_{ik}$  representa a aproximação obtida para  $F_i$  na folha  $k$ .

#### Subdivisão espacial: Malha $2^n$ -ádica

Cota do erro de aproximação de $F_{ik}$ em cada folha $k$ , $i = 1, 2$	$10^{-8}$
Nível Máximo da árvore	3
Cota do desvio padrão dos $y$	0.1

Tabela 7.2: Subdivisão espacial da curva no  $\mathbb{R}^3$  por Malha  $2^n$ -ádica.

<b>Aproximação da curva no <math>\mathbb{R}^3</math> por Malha <math>2^n</math>-ádica e peso polinomial</b>			
Grau do polinômio	1	2	3
Nº folhas	279	241	212
Crítérios de parada 0:(%dados)	54%	2%	1%
Crítérios de parada 1:(%dados)	0%	92%	96%
Crítérios de parada 2:(%dados)	0%	1%	0%
Crítérios de parada 3:(%dados)	46%	5%	3%
Erro de aproximação da $F_1$	$6.41 \times 10^{-8}$	$3.20 \times 10^{-11}$	$1.42 \times 10^{-11}$
Erro de aproximação da $F_2$	$4.75 \times 10^{-5}$	$8.19 \times 10^{-11}$	$1.02 \times 10^{-10}$
Máximo dos EQM nas folhas para $F_1$	$1.31 \times 10^{-2}$	$1.72 \times 10^{-8}$	$1.42 \times 10^{-8}$
Mínimo dos EQM nas folhas para $F_1$	$5.60 \times 10^{-11}$	$4.34 \times 10^{-11}$	$8.11 \times 10^{-12}$
Média dos EQM nas folhas para $F_1$	$4.89 \times 10^{-5}$	$2.69 \times 10^{-9}$	$2.54 \times 10^{-9}$
Máximo dos EQM nas folhas para $F_2$	$3.36 \times 10^{-2}$	$1.98 \times 10^{-6}$	$1.61 \times 10^{-8}$
Mínimo dos EQM nas folhas para $F_2$	$7.92 \times 10^{-8}$	$3.15 \times 10^{-11}$	$2.74 \times 10^{-12}$
Média dos EQM nas folhas para $F_2$	$4.78 \times 10^{-4}$	$1.20 \times 10^{-8}$	$2.54 \times 10^{-9}$
Regressão dentro da amostra			
EQM ylocal	$2.7 \times 10^{-3}$	$3.5 \times 10^{-3}$	$4.3 \times 10^{-3}$
EQM yglobal	$5.09 \times 10^{-5}$	$7.44 \times 10^{-5}$	$1.0 \times 10^{-4}$
MAPE ylocal	0.1458	0.1505	0.1631
MAPE yglobal	0.026	0.0284	0.0326
<b>Aproximação da curva no <math>\mathbb{R}^3</math> por Malha <math>2^n</math>-ádica e peso gaussiano</b>			
Grau do polinômio	1	2	3
Nº folhas	279	241	212
Crítérios de parada 0:(%dados)	54%	2%	1%
Crítérios de parada 1:(%dados)	0%	92%	96%
Crítérios de parada 2:(%dados)	0%	1%	0%
Crítérios de parada 3:(%dados)	46%	5%	3%
Erro de aproximação da $F_1$	$1.63 \times 10^{-7}$	$2.42 \times 10^{-11}$	$1.17 \times 10^{-11}$
Erro de aproximação da $F_2$	$2.61 \times 10^{-5}$	$1.11 \times 10^{-10}$	$7.42 \times 10^{-11}$
Máximo dos EQM nas folhas para $F_1$	$1.31 \times 10^{-2}$	$1.72 \times 10^{-8}$	$1.42 \times 10^{-8}$
Mínimo dos EQM nas folhas para $F_1$	$5.60 \times 10^{-11}$	$4.34 \times 10^{-11}$	$8.11 \times 10^{-12}$
Média dos EQM nas folhas para $F_1$	$4.89 \times 10^{-5}$	$2.69 \times 10^{-9}$	$2.32 \times 10^{-9}$
Máximo dos EQM nas folhas para $F_2$	$3.36 \times 10^{-2}$	$1.98 \times 10^{-6}$	$1.61 \times 10^{-8}$
Mínimo dos EQM nas folhas para $F_2$	$7.92 \times 10^{-8}$	$3.16 \times 10^{-11}$	$2.74 \times 10^{-12}$
Média dos EQM nas folhas para $F_2$	$4.78 \times 10^{-4}$	$1.21 \times 10^{-8}$	$2.64 \times 10^{-9}$
Regressão dentro da amostra			
EQM ylocal	$4.40 \times 10^{-3}$	$5.60 \times 10^{-3}$	$6.60 \times 10^{-3}$
EQM yglobal	$2.41 \times 10^{-4}$	$3.50 \times 10^{-4}$	$4.45 \times 10^{-4}$
MAPE ylocal	0.174	0.1945	0.2233
MAPE yglobal	0.057	0.0647	0.0695

Tabela 7.3: Aproximação da curva no  $\mathbb{R}^3$  por Malha  $2^n$ -ádica.

**Subdivisão espacial: Árvore BSP**

Cota do erro de aproximação de $F_{ik}$ em cada folha k	$10^{-10}$
Cota do desvio padrão dos y	0.1

Tabela 7.4: Dados da subdivisão espacial da curva no  $\mathbb{R}^3$  pela Árvore BSP.

Com relação a aproximação da curva no  $\mathbb{R}^3$ , usando a Malha  $2^n$ -ádica e a Árvore BSP, com polinômios de grau 1, independente dos pesos adotados, os resultados foram semelhantes. Nestes casos, o critério de parada se divide quase que de forma igual entre a árvore ter atingido sua profundidade máxima e termos obtido o desvio padrão nas saídas menor do que a cota adotada, assim, o polinômio de grau 1 oferece uma melhor aproximação das saídas, fato confirmado ao se realizar a regressão com os dados dentro da amostra, onde também observamos que a estimativa global apresenta erros menores do que a estimativa local.

Já ao se utilizar polinômios de grau 2 ou 3, os resultados são diferentes. Na Malha  $2^n$ -ádica, o critério de parada dominante (com mais de 90% dos casos) se deve ao erro de aproximação local da função polinomial ter sido menor do que a cota estabelecida. Isso indica que os polinômios de grau maior, no caso, 2 e 3, aproximam melhor a estrutura, fato que pode ser confirmado pelos melhores resultados obtidos para esses graus de polinômio nos erros de aproximação para  $F_1$  e  $F_2$ , quando comparados aos erros de aproximação dados pelos polinômios de grau 1. Na árvore BSP, o critério de parada dominante se deve a termos alcançado a profundidade máxima estabelecida, embora que, com polinômios de grau 2, em cerca de 22% dos casos, o critério de parada se deu pelo fato do desvio padrão das saídas ter sido menor do que a cota dada. Por essa razão, na regressão dentro da amostra também se verifica um melhor resultado do que com polinômios de grau 3.

**Resultados da regressão fora da amostra**

Para esta etapa, realizamos a regressão utilizando o peso polinomial da Malha  $2^n$ -ádica com polinômios de grau 1 e testamos a Árvore BSP com peso gaussiano e polinômios de grau 1 e, repetimos este último com polinômios de grau 3. Os resultados obtidos usando a Árvore BSP e estimativas globais foram melhores do que com a Malha  $2^n$ -ádica, conforme pode ser visto na tabela 7.6.

Vamos utilizar apenas a Malha  $2^n$ -ádica (estimativas local e global) e a Árvore BSP (estimativa global) de grau 1 para comparar com outros métodos.

Em todos os casos apresentados na tabela 7.7 o método RCVI supera os demais métodos.

Vamos realizar uma nova regressão fora da amostra, mas agora a cada dado de entrada  $\mathbf{x}$  associaremos uma nova saída  $y$  dada por:  $f(\mathbf{x}) = \sin(x_1) + \cos(x_3) + e^{x_2}$ . Esta regressão será identificada como regressão fora da amostra 2 e os resultados são exibidos na tabela 7.8.

Os resultados foram mantidos, isto é, os métodos RCVI superaram os outros métodos utilizados para comparação e as estimativas globais apresentam melhores resultados do que as estimativas locais.

<b>Aproximação da curva no <math>\mathbb{R}^3</math> pela Árvore BSP e peso polinomial</b>			
Grau do polinômio	1	2	3
Nível Máximo da árvore	7	6	5
Nº folhas	67	51	31
Critérios de parada 0:(%dados)	48%	78%	97%
Critérios de parada 1:(%dados)	0%	0%	0%
Critérios de parada 2:(%dados)	0%	0%	0%
Critérios de parada 3:(%dados)	52%	22%	3%
Erro de aproximação da $F_1$	$9.2 \times 10^{-1}$	0.358	$3.2 \times 10^{-1}$
Erro de aproximação da $F_2$	1.3483	$2.63 \times 10^{-1}$	$2.5 \times 10^{-1}$
Máximo dos EQM nas folhas para $F_1$	$8.42 \times 10^{-6}$	$1.64 \times 10^{-8}$	$8.32 \times 10^{-9}$
Mínimo dos EQM nas folhas para $F_1$	$3.85 \times 10^{-10}$	$3.81 \times 10^{-10}$	$2.12 \times 10^{-10}$
Média dos EQM nas folhas para $F_1$	$3.24 \times 10^{-7}$	$4.15 \times 10^{-9}$	$2.84 \times 10^{-9}$
Máximo dos EQM nas folhas para $F_2$	$1.3 \times 10^{-2}$	$1.97 \times 10^{-8}$	$8.08 \times 10^{-9}$
Mínimo dos EQM nas folhas para $F_2$	$7.9 \times 10^{-8}$	$1.71 \times 10^{-10}$	$9.75 \times 10^{-11}$
Média dos EQM nas folhas para $F_2$	$8.2 \times 10^{-4}$	$6.50 \times 10^{-9}$	$3.12 \times 10^{-9}$
Regressão dentro da amostra			
Lambda	0.5	0.5	0.5
EQM ylocal	0.1892	0.1931	0.2122
EQM yglobal	$2.12 \times 10^{-9}$	$3.66 \times 10^{-7}$	$1.52 \times 10^{-3}$
MAPE ylocal	2.1181	2.121	2.1237
MAPE yglobal	$1.45 \times 10^{-4}$	$3.76 \times 10^{-3}$	$6.53 \times 10^{-2}$
<b>Aproximação da curva no <math>\mathbb{R}^3</math> pela Árvore BSP e peso gaussiano</b>			
Grau do polinômio	1	2	3
Nível Máximo da árvore	7	6	5
Nº folhas	67	51	31
Critérios de parada 0:(%dados)	48%	78%	97%
Critérios de parada 1:(%dados)	0%	0%	0%
Critérios de parada 2:(%dados)	0%	0%	0%
Critérios de parada 3:(%dados)	52%	22%	3%
Erro de aproximação da $F_1$	$9.5 \times 10^{-2}$	$1.33 \times 10^{-1}$	$2.6 \times 10^{-2}$
Erro de aproximação da $F_2$	$3.4 \times 10^{-1}$	$8.82 \times 10^{-2}$	$4.5 \times 10^{-2}$
Máximo dos EQM nas folhas para $F_1$	$8.42 \times 10^{-6}$	$1.64 \times 10^{-8}$	$8.32 \times 10^{-9}$
Mínimo dos EQM nas folhas para $F_1$	$3.85 \times 10^{-10}$	$3.61 \times 10^{-10}$	$2.12 \times 10^{-10}$
Média dos EQM nas folhas para $F_1$	$3.24 \times 10^{-7}$	$4.15 \times 10^{-9}$	$3.29 \times 10^{-9}$
Máximo dos EQM nas folhas para $F_2$	$1.3 \times 10^{-2}$	$1.97 \times 10^{-8}$	$8.08 \times 10^{-9}$
Mínimo dos EQM nas folhas para $F_2$	$7.9 \times 10^{-8}$	$1.71 \times 10^{-10}$	$3.69 \times 10^{-12}$
Média dos EQM nas folhas para $F_2$	$8.2 \times 10^{-4}$	$6.63 \times 10^{-9}$	$3.29 \times 10^{-9}$
Regressão dentro da amostra			
Lambda	0.5	0.5	0.5
EQM ylocal	0.7769	0.7865	0.8079
EQM yglobal	$2.02 \times 10^{-14}$	$1.20 \times 10^{-13}$	$1.02 \times 10^{-8}$
MAPE ylocal	1.2601	1.2482	1.2208
MAPE yglobal	$8.24 \times 10^{-7}$	$2.21 \times 10^{-6}$	$4.70 \times 10^{-4}$

Tabela 7.5: Aproximação da curva no  $\mathbb{R}^3$  pela Árvore BSP.

EQM	Malha $2^n$ -ádica e grau1		Árvore BSP e grau 1		Árvore BSP e grau 3	
	ylocal	yglobal	ylocal	yglobal	ylocal	yglobal
Mínimo	$1.87 \times 10^{-3}$	$3.62 \times 10^{-5}$	$6.0 \times 10^{-1}$	$3.41 \times 10^{-14}$	$5.98 \times 10^{-1}$	$1.26 \times 10^{-8}$
Máximo	$2.13 \times 10^{-3}$	$5.51 \times 10^{-5}$	$6.91 \times 10^{-1}$	$4.10 \times 10^{-14}$	$7.45 \times 10^{-1}$	$2.33 \times 10^{-8}$
Mediana	$1.98 \times 10^{-3}$	$4.71 \times 10^{-5}$	$6.65 \times 10^{-1}$	$3.64 \times 10^{-14}$	$6.81 \times 10^{-1}$	$1.72 \times 10^{-8}$
MAD	$7.46 \times 10^{-5}$	$5.21 \times 10^{-6}$	$1.55 \times 10^{-2}$	$7.14 \times 10^{-16}$	$2.36 \times 10^{-2}$	$1.42 \times 10^{-9}$
MAPE	ylocal	yglobal	ylocal	yglobal	ylocal	yglobal
Mínimo	$8.00 \times 10^{-2}$	$7.31 \times 10^{-3}$	2.4214	$1.46 \times 10^{-7}$	1.9710	$1.46 \times 10^{-4}$
Máximo	$2.83 \times 10^{-1}$	$7.34 \times 10^{-2}$	4.7493	$4.47 \times 10^{-7}$	4.2499	$2.86 \times 10^{-4}$
Mediana	$1.17 \times 10^{-1}$	$1.07 \times 10^{-2}$	2.8204	$1.86 \times 10^{-7}$	2.8157	$1.98 \times 10^{-4}$
MAD	$3.34 \times 10^{-2}$	$1.26 \times 10^{-3}$	0.1550	$2.50 \times 10^{-8}$	0.5867	$3.80 \times 10^{-5}$

Tabela 7.6: Curva no  $\mathbb{R}^3$ . Regressão fora da amostra com Malha  $2^n$ -ádica e Árvore BSP.

EQM	Malha $2^n$ -ádica e grau1		BSP e grau1	SVR	BayesianRR	CART
	ylocal	yglobal	yglobal			
Mínimo	$1.87 \times 10^{-3}$	$3.62 \times 10^{-5}$	$3.41 \times 10^{-14}$	$3.32 \times 10^{-3}$	$1.21 \times 10^{-1}$	$6.67 \times 10^{-2}$
Máximo	$2.13 \times 10^{-3}$	$5.51 \times 10^{-5}$	$4.10 \times 10^{-14}$	$3.84 \times 10^{-3}$	$1.33 \times 10^{-1}$	$7.39 \times 10^{-2}$
Mediana	$1.98 \times 10^{-3}$	$4.71 \times 10^{-5}$	$3.64 \times 10^{-14}$	$3.63 \times 10^{-3}$	$1.26 \times 10^{-1}$	$7.18 \times 10^{-2}$
MAD	$7.46 \times 10^{-5}$	$5.21 \times 10^{-6}$	$7.14 \times 10^{-16}$	$5.65 \times 10^{-5}$	$2.78 \times 10^{-3}$	$2.32 \times 10^{-3}$
MAPE	ylocal	yglobal	yglobal	SVR	BayesianRR	CART
Mínimo	$8.00 \times 10^{-2}$	$7.31 \times 10^{-3}$	$1.46 \times 10^{-7}$	$1.25 \times 10^{-1}$	$5.59 \times 10^{-1}$	$3.90 \times 10^{-1}$
Máximo	$2.83 \times 10^{-1}$	$7.34 \times 10^{-2}$	$4.47 \times 10^{-7}$	$7.94 \times 10^{-1}$	3.4300	1.0195
Mediana	$1.17 \times 10^{-1}$	$1.07 \times 10^{-2}$	$1.86 \times 10^{-7}$	$1.71 \times 10^{-1}$	$8.23 \times 10^{-1}$	$7.01 \times 10^{-1}$
MAD	$3.34 \times 10^{-2}$	$1.26 \times 10^{-3}$	$2.50 \times 10^{-8}$	$3.85 \times 10^{-2}$	$1.98 \times 10^{-1}$	$1.65 \times 10^{-1}$

Tabela 7.7: Curva no  $\mathbb{R}^3$ . Regressão fora da amostra.

EQM	Malha $2^n$ -ádica e grau1		BSP e grau1	SVR	BayesianRR	CART
	ylocal	yglobal	yglobal			
Mínimo	$1.85 \times 10^{-3}$	$9.49 \times 10^{-5}$	$9.54 \times 10^{-15}$	$8.34 \times 10^{-3}$	$4.40 \times 10^{-2}$	$9.24 \times 10^{-2}$
Máximo	$2.46 \times 10^{-3}$	$1.20 \times 10^{-4}$	$2.73 \times 10^{-14}$	$9.70 \times 10^{-3}$	$5.13 \times 10^{-2}$	$1.49 \times 10^{-1}$
Mediana	$2.22 \times 10^{-3}$	$1.05 \times 10^{-4}$	$1.33 \times 10^{-14}$	$8.91 \times 10^{-3}$	$4.58 \times 10^{-2}$	$1.28 \times 10^{-1}$
MAD	$7.82 \times 10^{-5}$	$6.28 \times 10^{-6}$	$3.49 \times 10^{-15}$	$4.07 \times 10^{-4}$	$1.61 \times 10^{-3}$	$1.80 \times 10^{-2}$
MAPE	ylocal	yglobal	yglobal	SVR	BayesianRR	CART
Mínimo	$6.77 \times 10^{-2}$	$1.19 \times 10^{-2}$	$1.82 \times 10^{-7}$	$1.29 \times 10^{-1}$	$2.97 \times 10^{-1}$	$4.99 \times 10^{-1}$
Máximo	$2.10 \times 10^{-1}$	$3.97 \times 10^{-2}$	$6.24 \times 10^{-7}$	$2.44 \times 10^{-1}$	2.1600	1.9148
Mediana	$9.91 \times 10^{-2}$	$1.75 \times 10^{-2}$	$3.19 \times 10^{-7}$	$1.53 \times 10^{-1}$	$5.09 \times 10^{-1}$	$8.56 \times 10^{-1}$
MAD	$2.45 \times 10^{-2}$	$1.53 \times 10^{-3}$	$9.13 \times 10^{-8}$	$1.52 \times 10^{-2}$	$1.37 \times 10^{-1}$	$1.27 \times 10^{-1}$

Tabela 7.8: Curva no  $\mathbb{R}^3$ . Regressão fora da amostra 2.



### 7.1.2

#### Superfície no $\mathbb{R}^3$ : Swiss Roll

Geramos  $N = 3000$  dados, tomamos 10 vizinhos e  $\sigma = 20$ . A tabela 7.9 exibe os resultados da estimativa da dimensionalidade correta (DE) e o erro de orientação obtidos.

Curva no $\mathbb{R}^3$	
DE	82%
Mediana dos ângulos (graus)	2.42
Média dos ângulos (graus)	3.72

Tabela 7.9: *Swiss Roll*. Taxa da estimativa correta da dimensionalidade (DE). As estatísticas dos ângulos referem-se aos erros de orientação obtidos.

#### Resultados na aproximação da variedade implícita

As tabelas a seguir mostram os resultados obtidos na aproximação da função  $F : \mathbb{R}^3 \rightarrow \mathbb{R}$ .

#### Subdivisão espacial: Malha $2^n$ -ádica

Cota do erro de aproximação de $F$ em cada folha $k$	$10^{-8}$
Nível Máximo da árvore	3
Cota do desvio padrão dos $y$	0.1

Tabela 7.10: Subdivisão espacial da superfície no  $\mathbb{R}^3$  por Malha  $2^n$ -ádica.

#### Subdivisão espacial: Árvore BSP

Na aproximação da superfície no  $\mathbb{R}^3$  utilizando a Malha  $2^n$ -ádica, independente do peso, notamos que quando empregamos o polinômio de grau 3, o critério de parada predominante, com 61% dos casos, se deve ao erro de aproximação da função ter sido menor do que a cota adotada, indicando mais uma vez que o polinômio de grau maior determina uma melhor aproximação da variedade estudada. Os outros critérios de parada foram por termos alcançado a profundidade máxima da árvore. Ao se utilizar polinômios de grau 1 e 2 vimos que em aproximadamente 90% dos casos, o critério de parada foi devido a termos alcançado o nível máximo da árvore. Com respeito a regressão dentro da amostra, os resultados foram semelhantes.

Ao empregarmos a Árvore BSP, os critérios de parada predominantes para todos os graus de polinômio e peso adotados foram devidos ao nível máximo da árvore ter sido atingido. Na regressão dentro da amostra, o uso de

<b>Aproximação da superfície no <math>\mathbb{R}^3</math> por Malha <math>2^n</math>-ádica e peso polinomial</b>			
Grau do polinômio	1	2	3
Nº folhas	462	462	416
Critérios de parada 0:(%dados)	95%	86%	35%
Critérios de parada 1:(%dados)	0%	8%	61%
Critérios de parada 2:(%dados)	3%	5%	4%
Critérios de parada 3:(%dados)	2%	1%	0%
Erro de aproximação da $F$	$4.3 \times 10^{-4}$	$1.09 \times 10^{-6}$	$1.09 \times 10^{-6}$
Máximo dos EQM nas folhas para $F$	$1.13 \times 10^{-1}$	$2.32 \times 10^{-3}$	$9.05 \times 10^{-5}$
Mínimo dos EQM nas folhas para $F$	$1.57 \times 10^{-6}$	$5.33 \times 10^{-10}$	$8.19 \times 10^{-12}$
Média dos EQM nas folhas para $F$	$4.6 \times 10^{-3}$	$1.82 \times 10^{-5}$	$8.45 \times 10^{-7}$
Regressão dentro da amostra			
EQM ylocal	$7.7 \times 10^{-3}$	$7.7 \times 10^{-3}$	$8.3 \times 10^{-3}$
EQM yglobal	$4.6 \times 10^{-4}$	$4.6 \times 10^{-4}$	$5.0 \times 10^{-4}$
MAPE ylocal	$2.67 \times 10^{-1}$	$2.67 \times 10^{-1}$	$3.17 \times 10^{-1}$
MAPE yglobal	$9.6 \times 10^{-2}$	$9.6 \times 10^{-2}$	$9.3 \times 10^{-2}$
<b>Aproximação da superfície no <math>\mathbb{R}^3</math> por Malha <math>2^n</math>-ádica e peso gaussiano</b>			
Grau do polinômio	1	2	3
Nº folhas	462	462	416
Critérios de parada 0:(%dados)	95%	86%	35%
Critérios de parada 1:(%dados)	0%	8%	61%
Critérios de parada 2:(%dados)	3%	5%	4%
Critérios de parada 3:(%dados)	2%	1%	0%
Erro de aproximação da $F$	$6.1 \times 10^{-4}$	$1.15 \times 10^{-6}$	$3.77 \times 10^{-8}$
Máximo dos EQM nas folhas para $F$	$1.13 \times 10^{-1}$	$2.32 \times 10^{-3}$	$9.05 \times 10^{-5}$
Mínimo dos EQM nas folhas para $F$	$1.57 \times 10^{-6}$	$5.33 \times 10^{-10}$	$8.18 \times 10^{-12}$
Média dos EQM nas folhas para $F$	$4.6 \times 10^{-3}$	$1.82 \times 10^{-5}$	$8.45 \times 10^{-7}$
Regressão dentro da amostra			
EQM ylocal	$1.3 \times 10^{-2}$	$1.2 \times 10^{-2}$	$1.35 \times 10^{-2}$
EQM yglobal	$2.1 \times 10^{-3}$	$2.1 \times 10^{-3}$	$2.23 \times 10^{-3}$
MAPE ylocal	$3.41 \times 10^{-1}$	$3.41 \times 10^{-1}$	$3.33 \times 10^{-1}$
MAPE yglobal	$2.05 \times 10^{-1}$	$2.05 \times 10^{-1}$	$1.93 \times 10^{-1}$

Tabela 7.11: Aproximação da superfície no  $\mathbb{R}^3$  por Malha  $2^n$ -ádica.

Cota do erro de aproximação de $F$ em cada folha $k$	$10^{-10}$
Cota do desvio padrão dos $y$	0.1

Tabela 7.12: Dados da subdivisão espacial da superfície no  $\mathbb{R}^3$  pela Árvore BSP.

polinômios de grau 1, independente do peso, apresentou melhores resultados comparado aos polinômios de grau 2 e 3 e, destaca-se ainda, em todos os casos, o melhor desempenho na predição dado pela estimativa global em relação a estimativa local.

<b>Aproximação da superfície no <math>\mathbb{R}^3</math> pela Árvore BSP e peso polinomial</b>			
Grau do polinômio	1	2	3
Nível Máximo da árvore	8	7	6
Nº folhas	255	128	64
Critérios de parada 0:(%dados)	99.5%	100%	100%
Critérios de parada 1:(%dados)	0%	0%	0%
Critérios de parada 2:(%dados)	0%	0%	0%
Critérios de parada 3:(%dados)	0.5%	0%	0%
Erro de aproximação da $F$	1.4278	0.6782	1.7427
Máximo dos EQM nas folhas para $F$	$8.24 \times 10^{-2}$	$1.37 \times 10^{-4}$	$6.78 \times 10^{-6}$
Mínimo dos EQM nas folhas para $F$	$4.02 \times 10^{-7}$	$2.18 \times 10^{-10}$	$2.40 \times 10^{-10}$
Média dos EQM nas folhas para $F$	$3.52 \times 10^{-3}$	$5.06 \times 10^{-6}$	$2.48 \times 10^{-7}$
Regressão dentro da amostra			
Lambda	0.5	0.5	0.5
EQM ylocal	$1.65 \times 10^{-1}$	$1.91 \times 10^{-1}$	$2.37 \times 10^{-1}$
EQM yglobal	$4.74 \times 10^{-7}$	$5.11 \times 10^{-5}$	$9.80 \times 10^{-4}$
MAPE ylocal	$9.52 \times 10^{-1}$	$9.42 \times 10^{-1}$	1.0818
MAPE yglobal	$2.34 \times 10^{-3}$	$2.21 \times 10^{-2}$	$1.01 \times 10^{-1}$
<b>Aproximação da superfície no <math>\mathbb{R}^3</math> pela Árvore BSP e peso gaussiano</b>			
Grau do polinômio	1	2	3
Nível Máximo da árvore	7	6	5
Nº folhas	128	64	32
Critérios de parada 0:(%dados)	100%	100%	100%
Critérios de parada 1:(%dados)	0%	0%	0%
Critérios de parada 2:(%dados)	0%	0%	0%
Critérios de parada 3:(%dados)	0%	0%	0%
Erro de aproximação da $F$	$3.95 \times 10^{-1}$	$2.62 \times 10^{-2}$	$6.51 \times 10^{-2}$
Máximo dos EQM nas folhas para $F$	$7.954 \times 10^{-2}$	$1.90 \times 10^{-3}$	$4.82 \times 10^{-6}$
Mínimo dos EQM nas folhas para $F$	$4.07 \times 10^{-6}$	$1.84 \times 10^{-9}$	$1.52 \times 10^{-10}$
Média dos EQM nas folhas para $F$	$8.10 \times 10^{-3}$	$5.30 \times 10^{-5}$	$6.62 \times 10^{-7}$
Regressão dentro da amostra			
Lambda	0.5	0.5	0.5
EQM ylocal	$8.23 \times 10^{-1}$	$8.41 \times 10^{-1}$	$8.60 \times 10^{-1}$
EQM yglobal	$1.44 \times 10^{-11}$	$3.65 \times 10^{-8}$	$2.58 \times 10^{-5}$
MAPE ylocal	$9.86 \times 10^{-1}$	$9.91 \times 10^{-1}$	$9.92 \times 10^{-1}$
MAPE yglobal	$1.48 \times 10^{-5}$	$7.69 \times 10^{-4}$	$2.10 \times 10^{-2}$

Tabela 7.13: Aproximação da superfície no  $\mathbb{R}^3$  pela Árvore BSP.

### Resultados da regressão fora da amostra

Realizamos a regressão utilizando o peso polinomial da Malha  $2^n$ -ádica (estimativas local e global) e Árvore BSP com peso gaussiano (estimativa global apenas) e, em ambos os casos, trabalhamos com polinômios de grau 1.

Com exceção da estimativa local para a Malha  $2^n$ -ádica quando comparada com o SVR, todos os outros métodos RCVI apresentados na tabela 7.14

EQM	Malha $2^n$ -ádica e grau1		BSP e grau1	SVR	BayesianRR	CART
	ylocal	yglobal	yglobal			
Mínimo	$1.28 \times 10^{-2}$	$8.03 \times 10^{-4}$	$2.19 \times 10^{-11}$	$6.16 \times 10^{-3}$	0.4331	0.4707
Máximo	$1.65 \times 10^{-2}$	$1.13 \times 10^{-3}$	$3.46 \times 10^{-11}$	$7.11 \times 10^{-3}$	0.5302	0.5678
Mediana	$1.43 \times 10^{-2}$	$9.35 \times 10^{-4}$	$2.76 \times 10^{-11}$	$6.63 \times 10^{-3}$	0.4832	0.4939
MAD	$9.71 \times 10^{-4}$	$8.34 \times 10^{-5}$	$1.25 \times 10^{-12}$	$1.25 \times 10^{-4}$	$2.0 \times 10^{-2}$	$2.90 \times 10^{-2}$
MAPE	ylocal	yglobal	yglobal	SVR	BayesianRR	CART
Mínimo	$1.55 \times 10^{-1}$	$3.89 \times 10^{-2}$	$7.22 \times 10^{-6}$	$1.53 \times 10^{-1}$	1.3837	1.5145
Máximo	1.6943	$6.1 \times 10^{-1}$	$8.98 \times 10^{-5}$	$3.57 \times 10^{-1}$	4.5568	2.9879
Mediana	$2.25 \times 10^{-1}$	$7.15 \times 10^{-2}$	$1.12 \times 10^{-5}$	$2.47 \times 10^{-1}$	2.2162	2.2751
MAD	$5.74 \times 10^{-2}$	$1.64 \times 10^{-2}$	$1.88 \times 10^{-6}$	$8.56 \times 10^{-2}$	0.551	0.3915

Tabela 7.14: Superfície no  $\mathbb{R}^3$ . Regressão fora da amostra.

apresentam melhor desempenho do que os métodos usados na comparação.

Um nova regressão fora da amostra também será efetuada, no entanto, a cada dado de entrada  $\mathbf{x}$  associaremos uma nova saída  $y$  dada por:  $f(\mathbf{x}) = \sin(x_1) + \cos(x_3) + e^{x_2}$ . Esta regressão será identificada como regressão fora da amostra 2 e os resultados são exibidos na tabela 7.15.

EQM	Malha $2^n$ -ádica e grau1		BSP e grau1	SVR	BayesianRR	CART
	ylocal	yglobal	yglobal			
Mínimo	$1.36 \times 10^{-2}$	$6.19 \times 10^{-4}$	$5.79 \times 10^{-10}$	$6.37 \times 10^{-3}$	$3.53 \times 10^{-1}$	$2.47 \times 10^{-1}$
Máximo	$1.68 \times 10^{-2}$	$8.17 \times 10^{-4}$	$8.21 \times 10^{-10}$	$7.35 \times 10^{-3}$	$4.46 \times 10^{-1}$	$2.89 \times 10^{-1}$
Mediana	$1.50 \times 10^{-2}$	$6.90 \times 10^{-4}$	$6.88 \times 10^{-10}$	$6.88 \times 10^{-3}$	$4.09 \times 10^{-1}$	$2.74 \times 10^{-1}$
MAD	$1.06 \times 10^{-3}$	$3.51 \times 10^{-5}$	$8.63 \times 10^{-11}$	$1.70 \times 10^{-4}$	$8.78 \times 10^{-3}$	$1.84 \times 10^{-2}$
MAPE	ylocal	yglobal	yglobal	SVR	BayesianRR	CART
Mínimo	$2.44 \times 10^{-1}$	$3.40 \times 10^{-2}$	$4.16 \times 10^{-5}$	$8.89 \times 10^{-2}$	$9.11 \times 10^{-1}$	$5.22 \times 10^{-1}$
Máximo	2.8749	$8.32 \times 10^{-1}$	$5.98 \times 10^{-4}$	$1.97 \times 10^{-1}$	3.1895	3.1335
Mediana	$2.44 \times 10^{-1}$	$5.55 \times 10^{-2}$	$5.26 \times 10^{-5}$	$1.26 \times 10^{-1}$	1.1676	1.0057
MAD	$6.67 \times 10^{-2}$	$1.19 \times 10^{-2}$	$1.06 \times 10^{-5}$	$1.76 \times 10^{-2}$	$2.13 \times 10^{-1}$	$4.97 \times 10^{-1}$

Tabela 7.15: Superfície no  $\mathbb{R}^3$ . Regressão fora da amostra 2.

Os resultados foram mantidos, isto é, os métodos RCVI superaram os outros métodos utilizados para comparação, com exceção da estimativa local da aproximação com Malha  $2^n$ -ádica e polinômio de grau 1, quando comparado com o método SVR. Verifica-se que as estimativas globais apresentam melhores resultados do que as estimativas locais.

### 7.1.3

#### Curva no $\mathbb{R}^4$

Para realizar a aproximação de uma curva implícita no  $\mathbb{R}^4$  geramos  $N = 3000$  dados, tomamos 20 vizinhos e  $\sigma = 3$ . A tabela 7.16 exhibe

os resultados da estimativa da dimensionalidade correta (DE) e o erro de orientação obtidos.

Curva no $\mathbb{R}^4$	
DE	100%
mediana dos ângulos (graus)	0.36
média dos ângulos (graus)	0.41

Tabela 7.16: Curva no  $\mathbb{R}^4$ . Taxa da estimativa correta da dimensionalidade (DE). As estatísticas dos ângulos referem-se aos erros de orientação obtidos.

### Resultados na aproximação da variedade implícita

As tabelas a seguir mostram os resultados obtidos na aproximação da função  $F : \mathbb{R}^4 \rightarrow \mathbb{R}^3$ , onde  $F_i : \mathbb{R}^4 \rightarrow \mathbb{R}, i = 1, 2, 3$  e  $F_{ik}$  representa a aproximação obtida para  $F_i$  na folha  $k$ .

#### Subdivisão espacial: Malha $2^n$ -ádica

Cota do erro de aproximação de $F_{ik}$ em cada folha $k, i = 1, 2$	$10^{-8}$
Nível Máximo da árvore	2
Cota do desvio padrão dos $y$	0.1

Tabela 7.17: Subdivisão espacial da curva no  $\mathbb{R}^4$  por Malha  $2^n$ -ádica.

<b>Aproximação da curva no <math>\mathbb{R}^4</math> por Malha <math>2^n</math>-ádica e peso polinomial</b>			
Grau do polinômio	1	2	3
Nº folhas	191	191	156
CrITÉrios de parada 0:(%dados)	89%	22%	1%
CrITÉrios de parada 1:(%dados)	0%	78%	97%
CrITÉrios de parada 2:(%dados)	0%	0%	1%
CrITÉrios de parada 3:(%dados)	11%	0%	1%
Erro de aproximação da $F_1$	$3.77 \times 10^{-5}$	$1.90 \times 10^{-12}$	$7.17 \times 10^{-12}$
Erro de aproximação da $F_2$	$5.08 \times 10^{-4}$	$1.97 \times 10^{-8}$	$1.09 \times 10^{-11}$
Erro de aproximação da $F_3$	$4.89 \times 10^{-4}$	$1.70 \times 10^{-7}$	$3.44 \times 10^{-9}$
Máximo dos EQM nas folhas para $F_1$	$1.31 \times 10^{-2}$	$1.51 \times 10^{-8}$	$1.39 \times 10^{-8}$
Mínimo dos EQM nas folhas para $F_1$	$1.39 \times 10^{-11}$	$9.73 \times 10^{-12}$	$1.24 \times 10^{-12}$
Média dos EQM nas folhas para $F_1$	$3.53 \times 10^{-4}$	$1.09 \times 10^{-9}$	$9.64 \times 10^{-10}$
Máximo dos EQM nas folhas para $F_2$	$3.86 \times 10^{-2}$	$6.72 \times 10^{-6}$	$1.04 \times 10^{-7}$
Mínimo dos EQM nas folhas para $F_2$	$1.11 \times 10^{-9}$	$1.74 \times 10^{-11}$	$8.87 \times 10^{-12}$
Média dos EQM nas folhas para $F_2$	$4.69 \times 10^{-3}$	$2.11 \times 10^{-7}$	$1.84 \times 10^{-9}$
Máximo dos EQM nas folhas para $F_3$	$4.99 \times 10^{-2}$	$1.04 \times 10^{-4}$	$3.94 \times 10^{-4}$
Mínimo dos EQM nas folhas para $F_3$	$7.96 \times 10^{-7}$	$4.17 \times 10^{-11}$	$2.79 \times 10^{-12}$
Média dos EQM nas folhas para $F_3$	$8.02 \times 10^{-3}$	$1.97 \times 10^{-6}$	$2.53 \times 10^{-6}$
Regressão dentro da amostra			
EQM ylocal	$1.28 \times 10^{-1}$	$1.28 \times 10^{-1}$	$1.34 \times 10^{-1}$
EQM yglobal	$6.09 \times 10^{-2}$	$6.09 \times 10^{-2}$	3.5918
MAPE ylocal	$7.04 \times 10^{-1}$	$7.04 \times 10^{-1}$	$7.03 \times 10^{-1}$
MAPE yglobal	$7.24 \times 10^{-2}$	$7.24 \times 10^{-2}$	1.4482
<b>Aproximação da curva no <math>\mathbb{R}^4</math> por Malha <math>2^n</math>-ádica e peso gaussiano</b>			
Grau do polinômio	1	2	3
Nº folhas	191	191	156
CrITÉrios de parada 0:(%dados)	89%	22%	1%
CrITÉrios de parada 1:(%dados)	0%	77%	97%
CrITÉrios de parada 2:(%dados)	0%	0%	1%
CrITÉrios de parada 3:(%dados)	11%	1%	1%
Erro de aproximação da $F_1$	$3.75 \times 10^{-5}$	$1.85 \times 10^{-12}$	$5.77 \times 10^{-12}$
Erro de aproximação da $F_2$	$3.10 \times 10^{-4}$	$1.64 \times 10^{-8}$	$1.32 \times 10^{-11}$
Erro de aproximação da $F_3$	$3.86 \times 10^{-4}$	$1.84 \times 10^{-7}$	$1.41 \times 10^{-8}$
Máximo dos EQM nas folhas para $F_1$	$1.31 \times 10^{-2}$	$1.51 \times 10^{-8}$	$139 \times 10^{-8}$
Mínimo dos EQM nas folhas para $F_1$	$1.39 \times 10^{-11}$	$7.25 \times 10^{-12}$	$1.24 \times 10^{-12}$
Média dos EQM nas folhas para $F_1$	$3.53 \times 10^{-4}$	$1.09 \times 10^{-9}$	$9.64 \times 10^{-10}$
Máximo dos EQM nas folhas para $F_2$	$3.86 \times 10^{-2}$	$6.72 \times 10^{-6}$	$1.04 \times 10^{-7}$
Mínimo dos EQM nas folhas para $F_2$	$1.11 \times 10^{-9}$	$1.74 \times 10^{-11}$	$8.59 \times 10^{-12}$
Média dos EQM nas folhas para $F_2$	$4.69 \times 10^{-3}$	$2.11 \times 10^{-7}$	$1.80 \times 10^{-9}$
Máximo dos EQM nas folhas para $F_3$	$4.99 \times 10^{-2}$	$1.04 \times 10^{-4}$	$3.94 \times 10^{-4}$
Mínimo dos EQM nas folhas para $F_3$	$7.96 \times 10^{-7}$	$4.17 \times 10^{-11}$	$1.94 \times 10^{-13}$
Média dos EQM nas folhas para $F_3$	$8.02 \times 10^{-3}$	$1.97 \times 10^{-6}$	$2.53 \times 10^{-6}$
Regressão dentro da amostra			
EQM ylocal	$1.48 \times 10^{-1}$	$1.48 \times 10^{-1}$	$1.45 \times 10^{-1}$
EQM yglobal	$6.12 \times 10^{-2}$	$6.12 \times 10^{-2}$	$6.16 \times 10^{-2}$
MAPE ylocal	$7.99 \times 10^{-1}$	$7.99 \times 10^{-1}$	$6.52 \times 10^{-1}$
MAPE yglobal	$1.18 \times 10^{-1}$	$1.18 \times 10^{-1}$	$1.63 \times 10^{-1}$

Tabela 7.18: Aproximação da curva no  $\mathbb{R}^4$  por Malha  $2^n$ -ádica.

**Subdivisão espacial: Árvore BSP**

Cota do erro de aproximação de $F_{ik}$ em cada folha k	$10^{-10}$
Cota do desvio padrão dos y	0.1

Tabela 7.19: Dados da subdivisão espacial da curva no  $\mathbb{R}^4$  pela Árvore BSP.

Com relação a aproximação da curva no  $\mathbb{R}^4$ , analisando os resultados usando a Malha  $2^n$ -ádica, verificamos que quando se trabalha com o polinômio de grau 1 o critério de parada em 89% dos casos é devido a árvore ter alcançado a profundidade máxima estabelecida. Já ao utilizarmos polinômios de grau 2 ou 3, vemos que em 78% e 97% dos casos, respectivamente, o critério de parada foi por termos alcançado o erro na aproximação da variedade inferior a cota superior estabelecida, indicando um melhor ajuste a variedade, fato este que pode ser confirmado ao se observar o erro obtido de aproximação da  $F_1$ ,  $F_2$  e  $F_3$  ao utilizarmos polinômios de grau 2 ou 3. Com relação a regressão dentro da amostra, os polinômios de grau 1 e 2 apresentam os mesmos resultados, porque provavelmente obtivemos a mesma divisão dos dados e, esses polinômios exibem um melhor desempenho do que o polinômio de grau 3.

Já ao se empregar a Árvore BSP e usar o polinômio de grau 1, o critério de parada na subdivisão da árvore se divide quase que igualmente nas seguintes situações: a profundidade máxima admitida foi atingida e obteve-se um desvio padrão nos dados de saída inferior a cota estabelecida. Quando se adota o polinômio de grau 2, em 86% dos casos, a parada na subdivisão da árvore se deve a termos alcançado o nível máximo da árvore e 14% por termos conseguido desvio padrão nos dados de saída menor do que a cota dada. Com o polinômio de grau 3, a parada na subdivisão espacial se deve 100% a termos chegado a profundidade máxima da árvore. Os resultados na regressão dentro da amostra com os polinômios de grau 1 e 2 são semelhantes e melhores do que os obtidos com o polinômio de grau 3. Esse resultado na regressão dentro da amostra mostra-se coerente com os critérios de parada encontrados.

<b>Aproximação da curva no <math>\mathbb{R}^4</math> pela Árvore BSP e peso polinomial</b>			
Grau do polinômio	1	2	3
Nº folhas	72	56	32
CrITÉrios de parada 0:(%dados)	44%	86%	100%
CrITÉrios de parada 1:(%dados)	0%	0%	0%
CrITÉrios de parada 2:(%dados)	0%	0%	0%
CrITÉrios de parada 3:(%dados)	56%	14%	0%
Erro de aproximação da $F_1$	$3.38 \times 10^{-1}$	$5.23 \times 10^{-1}$	1.7781
Erro de aproximação da $F_2$	$2.35 \times 10^{-1}$	1.91851221	36.14847983
Erro de aproximação da $F_3$	1.0695	$5.87 \times 10^{-1}$	1.55061501
Máximo dos EQM nas folhas para $F_1$	$1.28 \times 10^{-6}$	$1.08 \times 10^{-8}$	$6.39 \times 10^{-9}$
Mínimo dos EQM nas folhas para $F_1$	$2.64 \times 10^{-10}$	$1.78 \times 10^{-10}$	$6.77 \times 10^{-12}$
Média dos EQM nas folhas para $F_1$	$2.77 \times 10^{-8}$	$3.55 \times 10^{-9}$	$2.14 \times 10^{-9}$
Máximo dos EQM nas folhas para $F_2$	$1.30 \times 10^{-2}$	$1.64 \times 10^{-8}$	$7.77 \times 10^{-9}$
Mínimo dos EQM nas folhas para $F_2$	$2.32 \times 10^{-10}$	$1.65 \times 10^{-10}$	$3.17 \times 10^{-11}$
Média dos EQM nas folhas para $F_2$	$1.81 \times 10^{-4}$	$5.12 \times 10^{-9}$	$2.56 \times 10^{-9}$
Máximo dos EQM nas folhas para $F_3$	$2.02 \times 10^{-2}$	$1.73 \times 10^{-8}$	$8.36 \times 10^{-9}$
Mínimo dos EQM nas folhas para $F_3$	$1.29 \times 10^{-7}$	$1.05 \times 10^{-10}$	$4.32 \times 10^{-11}$
Média dos EQM nas folhas para $F_3$	$7.30 \times 10^{-4}$	$5.23 \times 10^{-9}$	$263 \times 10^{-9}$
Regressão dentro da amostra			
Lambda	0.5	0.5	0.5
EQM ylocal	$2.26 \times 10^{-1}$	$2.30 \times 10^{-1}$	$2.54 \times 10^{-1}$
EQM yglobal	$4.99 \times 10^{-10}$	$1.39 \times 10^{-9}$	$4.21 \times 10^{-5}$
MAPE ylocal	1.6599	1.6384	1.6516
MAPE yglobal	$5.61 \times 10^{-5}$	$1.43 \times 10^{-4}$	$2.37 \times 10^{-2}$
<b>Aproximação da curva no <math>\mathbb{R}^4</math> pela Árvore BSP e peso gaussiano</b>			
Grau do polinômio	1	2	3
Nº folhas	72	56	32
CrITÉrios de parada 0:(%dados)	44.5%	86%	100%
CrITÉrios de parada 1:(%dados)	0%	0%	0%
CrITÉrios de parada 2:(%dados)	0%	0%	0%
CrITÉrios de parada 3:(%dados)	55.5%	14%	0%
Erro de aproximação da $F_1$	$2.59 \times 10^{-1}$	$3.28 \times 10^{-1}$	1.5093
Erro de aproximação da $F_2$	$1.29 \times 10^{-1}$	$3.72 \times 10^{-2}$	$4.86 \times 10^{-2}$
Erro de aproximação da $F_3$	$3.07 \times 10^{-1}$	$3.30 \times 10^{-1}$	$3.45 \times 10^{-1}$
Máximo dos EQM nas folhas para $F_1$	$1.28 \times 10^{-6}$	$1.08 \times 10^{-8}$	$6.39 \times 10^{-9}$
Mínimo dos EQM nas folhas para $F_1$	$2.64 \times 10^{-10}$	$1.78 \times 10^{-10}$	$6.77 \times 10^{-12}$
Média dos EQM nas folhas para $F_1$	$2.77 \times 10^{-8}$	$3.55 \times 10^{-9}$	$2.12 \times 10^{-9}$
Máximo dos EQM nas folhas para $F_2$	$1.30 \times 10^{-2}$	$1.64 \times 10^{-8}$	$7.79 \times 10^{-9}$
Mínimo dos EQM nas folhas para $F_2$	$2.32 \times 10^{-10}$	$1.65 \times 10^{-10}$	$1.07 \times 10^{-10}$
Média dos EQM nas folhas para $F_2$	$1.81 \times 10^{-4}$	$5.12 \times 10^{-9}$	$3.15 \times 10^{-9}$
Máximo dos EQM nas folhas para $F_3$	$2.02 \times 10^{-2}$	$1.73 \times 10^{-8}$	$8.98 \times 10^{-9}$
Mínimo dos EQM nas folhas para $F_3$	$1.29 \times 10^{-7}$	$1.05 \times 10^{-10}$	$1.29 \times 10^{-10}$
Média dos EQM nas folhas para $F_3$	$7.30 \times 10^{-4}$	$5.11 \times 10^{-9}$	$2.95 \times 10^{-9}$
Regressão dentro da amostra			
Lambda	0.5	0.5	0.5
EQM ylocal	$7.81 \times 10^{-1}$	$7.89 \times 10^{-1}$	$8.08 \times 10^{-1}$
EQM yglobal	$6.49 \times 10^{-14}$	$7.06 \times 10^{-13}$	$9.84 \times 10^{-8}$
MAPE ylocal	1.1230	1.1172	1.1002
MAPE yglobal	$1.25 \times 10^{-6}$	$4.45 \times 10^{-6}$	$1.87 \times 10^{-3}$

Tabela 7.20: Aproximação da curva no  $\mathbb{R}^4$  pela Árvore BSP.



### Resultados da regressão fora da amostra

Para esta etapa, realizamos a regressão utilizando o peso polinomial da Malha  $2^n$ -ádica com polinômio de grau 1 (estimativas local e global) e testamos a árvore BSP (estimativa global) com peso gaussiano e polinômio de grau 1.

EQM	Malha $2^n$ -ádica e grau1		BSP e grau1	SVR	BayesianRR	CART
	ylocal	yglobal	yglobal			
Mínimo	$7.77 \times 10^{-2}$	$2.35 \times 10^{-2}$	$1.66 \times 10^{-13}$	$3.72 \times 10^{-3}$	$1.13 \times 10^{-1}$	$6.17 \times 10^{-2}$
Máximo	$1.46 \times 10^{-1}$	$9.28 \times 10^{-2}$	$2.29 \times 10^{-12}$	$4.16 \times 10^{-3}$	$1.29 \times 10^{-1}$	$6.99 \times 10^{-2}$
Mediana	$1.01 \times 10^{-1}$	$4.88 \times 10^{-2}$	$4.06 \times 10^{-13}$	$3.88 \times 10^{-3}$	$1.15 \times 10^{-1}$	$6.41 \times 10^{-2}$
MAD	$8.20 \times 10^{-3}$	$9.60 \times 10^{-3}$	$1.88 \times 10^{-13}$	$6.42 \times 10^{-5}$	$2.47 \times 10^{-3}$	$2.92 \times 10^{-3}$
MAPE	ylocal	yglobal	yglobal	SVR	BayesianRR	CART
Mínimo	$3.69 \times 10^{-1}$	$7.83 \times 10^{-2}$	$4.87 \times 10^{-7}$	$8.92 \times 10^{-2}$	$4.35 \times 10^{-1}$	$3.58 \times 10^{-1}$
Máximo	339.23	4.50	$4.66 \times 10^{-6}$	107.5249	441.8825	321.6187
Mediana	$8.88 \times 10^{-1}$	$2.27 \times 10^{-1}$	$1.82 \times 10^{-6}$	$2.27 \times 10^{-1}$	1.0804	$7.19 \times 10^{-1}$
MAD	$3.12 \times 10^{-1}$	$1.01 \times 10^{-1}$	$8.56 \times 10^{-7}$	$8.93 \times 10^{-2}$	$3.13 \times 10^{-1}$	$4.10 \times 10^{-1}$

Tabela 7.21: Curva no  $\mathbb{R}^4$ . Regressão fora da amostra.

Avaliando a tabela 7.21 vemos que o método RCVI utilizando a Malha  $2^n$ -ádica não tem um bom desempenho, principalmente com a estimativa local, no entanto, com a estimativa global o método apresenta bons resultados embora não supere o método SVR. Ao trabalhar com a Árvore BSP os resultados são melhores do que todos os obtidos pelos outros métodos utilizados na comparação.

Pode-se verificar alguns valores muito altos na estimativa MAPE, embora essa estatística seja vantajosa por ser fácil de interpretar e a escala ser percentual, uma desvantagem ao se utilizar essa estatística deve-se as situações onde o valor real é pequeno, e, neste caso, qualquer discrepância na previsão faz com que o MAPE “exploda”.

Vamos realizar uma nova regressão fora da amostra, mas agora a cada dado de entrada  $\mathbf{x}$  associaremos uma nova saída  $y$  dada por:  $f(\mathbf{x}) = \sin(x_1) + \cos(x_3) + e^{x_2}$ . Esta regressão será identificada como regressão fora da amostra 2 e os resultados são exibidos na tabela 7.22.

Os resultados foram mantidos, isto é, o método RCVI com Árvore BSP, peso gaussiano e grau do polinômio igual a 1 supera os outros métodos utilizados na comparação.

EQM	Malha $2^n$ -ádica e grau 1		BSP e grau 1	SVR	BayesianRR	CART
	ylocal	yglobal	yglobal			
Mínimo	$1.28 \times 10^{-1}$	$7.16 \times 10^{-2}$	$4.75 \times 10^{-14}$	$9.70 \times 10^{-3}$	$3.53 \times 10^{-2}$	$9.26 \times 10^{-2}$
Máximo	$2.72 \times 10^{-1}$	$2.14 \times 10^{-1}$	$1.48 \times 10^{-12}$	$1.08 \times 10^{-2}$	$4.12 \times 10^{-2}$	$1.10 \times 10^{-1}$
Mediana	$1.66 \times 10^{-1}$	$1.06 \times 10^{-1}$	$2.57 \times 10^{-13}$	$1.00 \times 10^{-2}$	$3.90 \times 10^{-2}$	$9.82 \times 10^{-2}$
MAD	$3.07 \times 10^{-2}$	$2.77 \times 10^{-2}$	$1.94 \times 10^{-13}$	$2.18 \times 10^{-4}$	$1.47 \times 10^{-3}$	$6.40 \times 10^{-3}$
MAPE	ylocal	yglobal	yglobal	SVR	BayesianRR	CART
Mínimo	$5.24 \times 10^{-1}$	$3.09 \times 10^{-2}$	$2.65 \times 10^{-7}$	$1.66 \times 10^{-1}$	$3.41 \times 10^{-1}$	$4.84 \times 10^{-1}$
Máximo	1.3186	$5.86 \times 10^{-2}$	$3.09 \times 10^{-6}$	$5.93 \times 10^{-1}$	1.0563	1.5515
Mediana	$6.61 \times 10^{-1}$	$3.94 \times 10^{-2}$	$8.60 \times 10^{-7}$	$3.98 \times 10^{-1}$	$6.11 \times 10^{-1}$	$6.21 \times 10^{-1}$
MAD	$1.17 \times 10^{-1}$	$5.32 \times 10^{-3}$	$3.52 \times 10^{-7}$	$7.98 \times 10^{-2}$	$2.69 \times 10^{-1}$	$9.49 \times 10^{-2}$

Tabela 7.22: Curva no  $\mathbb{R}^4$ . Regressão fora da amostra 2.

### 7.1.4

#### Superfície no $\mathbb{R}^4$ :

Geramos  $N = 3000$  dados, tomamos 20 vizinhos e  $\sigma = 15$ . A tabela 7.23 exibe os resultados da estimativa da dimensionalidade correta (DE) e o erro de orientação obtidos.

Curva no $\mathbb{R}^3$	
DE	94%
Mediana dos ângulos (graus)	1.11
Média dos ângulos (graus)	1.22

Tabela 7.23: Superfície no  $\mathbb{R}^4$ . Taxa da estimativa correta da dimensionalidade (DE). As estatísticas dos ângulos referem-se aos erros de orientação obtidos.

#### Resultados na aproximação da variedade implícita

As tabelas a seguir mostram os resultados obtidos na aproximação da função  $F : \mathbb{R}^4 \rightarrow \mathbb{R}^2$ .

#### Subdivisão espacial: Malha $2^n$ -ádica

Cota do erro de aproximação de $F_{ik}$ em cada folha $k, i = 1, 2$	$10^{-8}$
Nível Máximo da árvore	2
Cota do desvio padrão dos $y$	0.1

Tabela 7.24: Subdivisão espacial da superfície no  $\mathbb{R}^4$  por Malha  $2^n$ -ádica.

<b>Aproximação da superfície no <math>\mathbb{R}^4</math> por Malha <math>2^n</math>-ádica e peso polinomial</b>			
Grau do polinômio	1	2	3
Nº folhas	192	192	192
CrITÉrios de parada 0:(%dados)	93%	92%	42%
CrITÉrios de parada 1:(%dados)	0%	0%	45%
CrITÉrios de parada 2:(%dados)	0%	8%	13%
CrITÉrios de parada 3:(%dados)	7%	0%	0%
Erro de aproximação da $F_1$	$5.46 \times 10^{-4}$	$9.90 \times 10^{-8}$	$8.21 \times 10^{-10}$
Erro de aproximação da $F_2$	$8.56 \times 10^{-4}$	$4.06 \times 10^{-7}$	$2.69 \times 10^{-9}$
Máximo dos EQM nas folhas para $F_1$	$3.74 \times 10^{-2}$	$6.12 \times 10^{-6}$	$1.11 \times 10^{-6}$
Mínimo dos EQM nas folhas para $F_1$	$5.04 \times 10^{-7}$	$1.42 \times 10^{-9}$	$4.76 \times 10^{-12}$
Média dos EQM nas folhas para $F_1$	$5.14 \times 10^{-3}$	$1.33 \times 10^{-6}$	$2.94 \times 10^{-8}$
Máximo dos EQM nas folhas para $F_2$	$4.91 \times 10^{-2}$	$4.46 \times 10^{-5}$	$2.25 \times 10^{-6}$
Mínimo dos EQM nas folhas para $F_2$	$4.31 \times 10^{-7}$	$3.26 \times 10^{-9}$	$1.81 \times 10^{-11}$
Média dos EQM nas folhas para $F_2$	$6.42 \times 10^{-3}$	$4.93 \times 10^{-6}$	$1.43 \times 10^{-7}$
Regressão dentro da amostra			
EQM ylocal	$1.76 \times 10^{-1}$	$1.76 \times 10^{-1}$	$1.76 \times 10^{-1}$
EQM yglobal	$1.36 \times 10^{-1}$	$1.36 \times 10^{-1}$	$1.36 \times 10^{-1}$
MAPE ylocal	$4.02 \times 10^{-1}$	$4.02 \times 10^{-1}$	$4.02 \times 10^{-1}$
MAPE yglobal	$2.22 \times 10^{-1}$	$2.22 \times 10^{-1}$	$2.22 \times 10^{-1}$
<b>Aproximação da superfície no <math>\mathbb{R}^4</math> por Malha <math>2^n</math>-ádica e peso gaussiano</b>			
Grau do polinômio	1	2	3
Nº folhas	192	192	192
CrITÉrios de parada 0:(%dados)	93%	92%	43.75%
CrITÉrios de parada 1:(%dados)	0%	0%	43.75%
CrITÉrios de parada 2:(%dados)	0%	0%	12.5%
CrITÉrios de parada 3:(%dados)	7%	8%	0%
Erro de aproximação da $F_1$	$6.26 \times 10^{-4}$	$9.14 \times 10^{-8}$	$1.04 \times 10^{-9}$
Erro de aproximação da $F_2$	$1.02 \times 10^{-3}$	$4.41 \times 10^{-7}$	$3.17 \times 10^{-9}$
Máximo dos EQM nas folhas para $F_1$	$3.74 \times 10^{-2}$	$6.12 \times 10^{-6}$	$1.11 \times 10^{-6}$
Mínimo dos EQM nas folhas para $F_1$	$5.04 \times 10^{-7}$	$1.42 \times 10^{-9}$	$4.76 \times 10^{-12}$
Média dos EQM nas folhas para $F_1$	$5.14 \times 10^{-3}$	$1.33 \times 10^{-6}$	$2.94 \times 10^{-8}$
Máximo dos EQM nas folhas para $F_2$	$4.91 \times 10^{-2}$	$4.46 \times 10^{-5}$	$2.25 \times 10^{-6}$
Mínimo dos EQM nas folhas para $F_2$	$4.31 \times 10^{-7}$	$3.26 \times 10^{-9}$	$1.81 \times 10^{-11}$
Média dos EQM nas folhas para $F_2$	$6.42 \times 10^{-3}$	$4.93 \times 10^{-6}$	$1.43 \times 10^{-7}$
Regressão dentro da amostra			
EQM ylocal	$1.85 \times 10^{-1}$	$1.85 \times 10^{-1}$	$1.85 \times 10^{-1}$
EQM yglobal	$1.38 \times 10^{-1}$	$1.38 \times 10^{-1}$	$1.38 \times 10^{-1}$
MAPE ylocal	$4.89 \times 10^{-1}$	$4.89 \times 10^{-1}$	$4.89 \times 10^{-1}$
MAPE yglobal	$4.53 \times 10^{-1}$	$4.53 \times 10^{-1}$	$4.53 \times 10^{-1}$

Tabela 7.25: Aproximação da superfície no  $\mathbb{R}^4$  por Malha  $2^n$ -ádica.**Subdivisão espacial: Árvore BSP**

Pela análise das tabelas 7.25 e 7.27, vimos que quando se utiliza a Malha  $2^n$ -ádica, os polinômios de grau 1 e 2 apresentam em 93% e 92% dos

Cota do erro de aproximação de $F_{ik}$ em cada folha $k$ , $i = 1, 2$	$10^{-10}$
Cota do desvio padrão dos $y$	0.1

Tabela 7.26: Dados da subdivisão espacial da superfície no  $\mathbb{R}^4$  pela Árvore BSP.

casos, respectivamente, a parada na subdivisão espacial por ter alcançado a profundidade máxima estabelecida. Já no caso do polinômio de grau 3, o critério de parada foi 42% por conta de ter alcançado o nível máximo e 43% pelo erro obtido na aproximação da variedade ser menor do que a cota estabelecida, o que implica em menores erros de aproximação encontrados para  $F_1$  e  $F_2$ , como podemos verificar na tabela, quando comparamos com os resultados para os polinômios de grau 1 e 2. Já com relação a regressão dentro da amostra obtivemos o mesmo resultado, o que indica que, independente do grau do polinômio, a divisão do domínio foi a mesma encontrada em todos os casos.

Ao se avaliar os resultados usando a Árvore BSP vimos que o critério de parada predominante se deve a ter sido alcançado o nível máximo da árvore. Já na regressão, verifica-se o melhor desempenho conseguido pela estimativa global, com destaque para a estimativa global obtida com a subdivisão dada pela Árvore BSP e peso gaussiano com polinômio de grau 1.

### Resultados da regressão fora da amostra

Realizamos a regressão utilizando o peso polinomial da Malha  $2^n$ -ádica (estimativa global) denotado na tabela por Malha  $2^n$ -ádica1 e Árvore BSP com peso gaussiano (estimativa global) e, em ambos os casos, trabalhamos com polinômios de grau 1. Optamos por não considerar a estimativa local já que seus resultados não são comparáveis aos demais métodos.

Com relação a tabela 7.28 a estimativa global dada pela Malha  $2^n$ -ádica não apresentou bons resultados, já a estimativa global dada pelo método RCVI utilizando a Árvore BSP, peso gaussiano e grau do polinômio 1, apresentou bons resultados, com exceção apenas ao desempenho obtido pelo BayesianRR.

Um nova regressão fora da amostra também será efetuada, no entanto, a cada dado de entrada  $\mathbf{x}$  associaremos uma nova saída  $y$  dada por:  $f(\mathbf{x}) = \sin(x_1) + \cos(x_3) + e^{x_2}$ . Esta regressão será identificada como regressão fora da amostra 2 e os resultados são exibidos na tabela 7.29.

Com os resultados obtidos com a nova regressão e exibidos na tabela 7.29, o método BayesianRR não repetiu o desempenho obtido anteriormente na tabela 7.28, a estimativa global com a Malha  $2^n$ -ádica também não apresentou

<b>Aproximação da superfície no <math>\mathbb{R}^4</math> pela Árvore BSP e peso polinomial</b>			
Grau do polinômio	1	2	3
Nível Máximo da árvore	7	6	5
Nº folhas	100	56	31
Critérios de parada 0:(%dados)	88%	89%	97%
Critérios de parada 1:(%dados)	0%	0%	0%
Critérios de parada 2:(%dados)	0%	0%	0%
Critérios de parada 3:(%dados)	12%	11%	3%
Erro de aproximação da $F_1$	$1.76 \times 10^{-2}$	$1.92 \times 10^{-1}$	2.8345
Erro de aproximação da $F_2$	$3.22 \times 10^{-1}$	$2.83 \times 10^{-1}$	$7.62 \times 10^{-1}$
Máximo dos EQM nas folhas para $F_1$	$1.76 \times 10^{-3}$	$6.44 \times 10^{-7}$	$9.06 \times 10^{-9}$
Mínimo dos EQM nas folhas para $F_1$	$7.76 \times 10^{-6}$	$2.74 \times 10^{-10}$	$7.76 \times 10^{-11}$
Média dos EQM nas folhas para $F_1$	$6.64 \times 10^{-5}$	$3.00 \times 10^{-8}$	$3.47 \times 10^{-9}$
Máximo dos EQM nas folhas para $F_2$	$2.53 \times 10^{-3}$	$2.44 \times 10^{-6}$	$1.32 \times 10^{-8}$
Mínimo dos EQM nas folhas para $F_2$	$1.10 \times 10^{-5}$	$4.75 \times 10^{-10}$	$1.23 \times 10^{-10}$
Média dos EQM nas folhas para $F_2$	$1.47 \times 10^{-4}$	$8.92 \times 10^{-8}$	$4.03 \times 10^{-9}$
Regressão dentro da amostra			
Lambda	0.5	0.5	0.5
EQM ylocal	$3.92 \times 10^{-1}$	$4.46 \times 10^{-1}$	$5.35 \times 10^{-1}$
EQM yglobal	$3.58 \times 10^{-4}$	$1.89 \times 10^{-3}$	$9.46 \times 10^{-3}$
MAPE ylocal	1.5844	1.6436	1.5839
MAPE yglobal	$1.36 \times 10^{-1}$	$2.91 \times 10^{-1}$	$5.17 \times 10^{-1}$
<b>Aproximação da superfície no <math>\mathbb{R}^4</math> pela Árvore BSP e peso gaussiano</b>			
Grau do polinômio	1	2	3
Nível Máximo da árvore	7	6	5
Nº folhas	100	56	31
Critérios de parada 0:(%dados)	88%	89%	97%
Critérios de parada 1:(%dados)	0%	0%	0%
Critérios de parada 2:(%dados)	0%	0%	0%
Critérios de parada 3:(%dados)	12%	11%	3%
Erro de aproximação da $F_1$	$6.26 \times 10^{-3}$	$1.02 \times 10^{-1}$	$2.23 \times 10^{-1}$
Erro de aproximação da $F_2$	$7.15 \times 10^{-2}$	$4.95 \times 10^{-2}$	$4.76 \times 10^{-1}$
Máximo dos EQM nas folhas para $F_1$	$1.76 \times 10^{-3}$	$6.44 \times 10^{-7}$	$9.06 \times 10^{-9}$
Mínimo dos EQM nas folhas para $F_1$	$7.76 \times 10^{-6}$	$2.74 \times 10^{-10}$	$7.76 \times 10^{-11}$
Média dos EQM nas folhas para $F_1$	$6.64 \times 10^{-5}$	$3.01 \times 10^{-8}$	$3.48 \times 10^{-9}$
Máximo dos EQM nas folhas para $F_2$	$2.53 \times 10^{-3}$	$2.44 \times 10^{-6}$	$1.44 \times 10^{-8}$
Mínimo dos EQM nas folhas para $F_2$	$1.10 \times 10^{-5}$	$4.75 \times 10^{-10}$	$3.58 \times 10^{-10}$
Média dos EQM nas folhas para $F_2$	$1.47 \times 10^{-4}$	$8.92 \times 10^{-8}$	$4.87 \times 10^{-9}$
Regressão dentro da amostra			
Lambda	0.5	0.5	0.5
EQM ylocal	$8.64 \times 10^{-1}$	$8.80 \times 10^{-1}$	$9.0 \times 10^{-1}$
EQM yglobal	$1.75 \times 10^{-10}$	$9.23 \times 10^{-8}$	$2.71 \times 10^{-5}$
MAPE ylocal	$9.60 \times 10^{-1}$	$9.69 \times 10^{-1}$	$9.75 \times 10^{-1}$
MAPE yglobal	$1.36 \times 10^{-4}$	$1.66 \times 10^{-3}$	$5.60 \times 10^{-2}$

Tabela 7.27: Aproximação da superfície no  $\mathbb{R}^4$  pela Árvore BSP.

EQM	Malha $2^n$ -ádica1	BSP e grau1	SVR	BayesianRR	CART
	yglobal	yglobal			
Mínimo	$7.52 \times 10^{-3}$	$2.07 \times 10^{-11}$	$2.07 \times 10^{-4}$	$4.98 \times 10^{-21}$	$2.45 \times 10^{-2}$
Máximo	$2.48 \times 10^{-2}$	$3.32 \times 10^{-11}$	$3.29 \times 10^{-4}$	$6.09 \times 10^{-21}$	$3.15 \times 10^{-2}$
Mediana	$1.75 \times 10^{-2}$	$2.41 \times 10^{-11}$	$2.83 \times 10^{-4}$	$5.60 \times 10^{-21}$	$2.79 \times 10^{-2}$
MAD	$2.06 \times 10^{-3}$	$2.68 \times 10^{-12}$	$2.04 \times 10^{-5}$	$2.18 \times 10^{-22}$	$1.81 \times 10^{-3}$
MAPE	yglobal	yglobal	SVR	BayesianRR	CART
Mínimo	$2.18 \times 10^{-1}$	$5.19 \times 10^{-4}$	$7.90 \times 10^{-1}$	$8.75 \times 10^{-10}$	1.2866
Máximo	6.3335	$4.64 \times 10^{-1}$	1177.46	$2.67 \times 10^{-8}$	2192.419
Mediana	1.4791	$1.17 \times 10^{-3}$	2.1886	$3.03 \times 10^{-9}$	2.8205
MAD	1.1772	$4.38 \times 10^{-4}$	1	$2.10 \times 10^{-9}$	2.0065

Tabela 7.28: Superfície no  $\mathbb{R}^4$ . Regressão fora da amostra.

EQM	Malha $2^n$ -ádica 1	BSP e grau1	SVR	BayesianRR	CART
	yglobal	yglobal			
Mínimo	$5.45 \times 10^{-2}$	$2.90 \times 10^{-10}$	$3.37 \times 10^{-3}$	$8.28 \times 10^{-3}$	$9.72 \times 10^{-2}$
Máximo	$1.36 \times 10^{-1}$	$9.83 \times 10^{-5}$	$4.09 \times 10^{-3}$	$1.47 \times 10^{-2}$	$1.20 \times 10^{-1}$
Mediana	$8.57 \times 10^{-2}$	$3.57 \times 10^{-10}$	$3.71 \times 10^{-3}$	$1.22 \times 10^{-2}$	$1.10 \times 10^{-1}$
MAD	$2.05 \times 10^{-2}$	$4.55 \times 10^{-11}$	$2.35 \times 10^{-4}$	$1.93 \times 10^{-3}$	$9.17 \times 10^{-3}$
MAPE	yglobal	yglobal	SVR	BayesianRR	CART
Mínimo	$7.48 \times 10^{-2}$	$3.45 \times 10^{-5}$	$1.79 \times 10^{-1}$	$3.42 \times 10^{-1}$	$8.95 \times 10^{-1}$
Máximo	$2.77 \times 10^{-1}$	$4.70 \times 10^{-2}$	$4.98 \times 10^{-1}$	1.4454	5.5108
Mediana	$1.17 \times 10^{-1}$	$9.66 \times 10^{-5}$	$2.55 \times 10^{-1}$	$4.55 \times 10^{-1}$	1.4777
MAD	$3.73 \times 10^{-2}$	$4.38 \times 10^{-5}$	$5.77 \times 10^{-2}$	$2.63 \times 10^{-2}$	$3.18 \times 10^{-1}$

Tabela 7.29: Superfície no  $\mathbb{R}^4$ . Regressão fora da amostra 2.

resultados comparáveis com os demais métodos e a estimativa global obtida da Árvore BSP com peso gaussiano e grau do polinômio 1 superou todos os outros métodos.

### 7.1.5

#### Volume no $\mathbb{R}^4$ :

Geramos  $N = 3000$  dados, tomamos 15 vizinhos e  $\sigma = 20$ . A tabela 7.30 exibe os resultados da estimativa da dimensionalidade correta (DE) e o erro de orientação obtidos.

Curva no $\mathbb{R}^3$	
DE	61%
Mediana dos ângulos (graus)	5.21
Média dos ângulos (graus)	7.23

Tabela 7.30: Volume no  $\mathbb{R}^4$ . Taxa da estimativa correta da dimensionalidade (DE). As estatísticas dos ângulos referem-se aos erros de orientação obtidos.

### Resultados na aproximação da variedade implícita

As tabelas a seguir mostram os resultados obtidos na aproximação da função  $F : \mathbb{R}^4 \rightarrow \mathbb{R}$ .

#### Subdivisão espacial: Malha $2^n$ -ádica

Cota do erro de aproximação de $F$ em cada folha $k$	$10^{-8}$
Nível Máximo da árvore	2
Cota do desvio padrão dos $y$	0.1

Tabela 7.31: Subdivisão espacial do volume no  $\mathbb{R}^4$  por Malha  $2^n$ -ádica.

#### Subdivisão espacial: Árvore BSP

Tanto ao utilizar a partição dada pela Malha  $2^n$ -ádica quanto pela Árvore BSP, o critério de parada ao trabalhar com polinômios de grau 1 e 2 foi 100% pelo fato de atingirmos o nível máximo da árvore. No caso da Malha  $2^n$ -ádica e polinômio de grau 3 obtivemos 73% dos casos de parada por termos atingido a profundidade máxima da árvore e 26% por termos obtido um erro de aproximação da variedade menor do que a cota dada, assim, obtivemos uma melhor aproximação da variedade com o polinômio de grau 3, o que também pode ser visto pelo menor erro na aproximação da  $F$ , quando comparado com os polinômios de grau 1 e 2. Na regressão dentro da amostra temos os mesmos resultados para os diferentes graus de polinômios e mesma subdivisão espacial, o que indica que obtivemos a mesma partição. Observa-se que a estimativa global apresenta erro de previsão menor do que a estimativa local.

<b>Aproximação do volume no <math>\mathbb{R}^4</math> por Malha <math>2^n</math>-ádica e peso polinomial</b>			
Grau do polinômio	1	2	3
Nº folhas	256	256	256
Critérios de parada 0:(%dados)	100%	100%	73%
Critérios de parada 1:(%dados)	0%	0%	26%
Critérios de parada 2:(%dados)	0%	0%	1%
Critérios de parada 3:(%dados)	0%	0%	0%
Erro de aproximação da $F$	$1.90 \times 10^{-3}$	$1.18 \times 10^{-5}$	$1.63 \times 10^{-7}$
Máximo dos EQM nas folhas para $F$	$2.02 \times 10^{-2}$	$2.90 \times 10^{-4}$	$5.36 \times 10^{-5}$
Mínimo dos EQM nas folhas para $F$	$5.47 \times 10^{-4}$	$1.97 \times 10^{-7}$	$3.74 \times 10^{-10}$
Média dos EQM nas folhas para $F$	$6.14 \times 10^{-3}$	$4.98 \times 10^{-5}$	$9.03 \times 10^{-7}$
Regressão dentro da amostra			
EQM ylocal	$2.30 \times 10^{-1}$	$2.30 \times 10^{-1}$	$2.30 \times 10^{-1}$
EQM yglobal	$2.36 \times 10^{-3}$	$2.36 \times 10^{-3}$	$2.36 \times 10^{-3}$
MAPE ylocal	$6.39 \times 10^{-1}$	$6.39 \times 10^{-1}$	$6.39 \times 10^{-1}$
MAPE yglobal	$1.97 \times 10^{-1}$	$1.97 \times 10^{-1}$	$1.97 \times 10^{-1}$
<b>Aproximação do volume no <math>\mathbb{R}^4</math> por Malha <math>2^n</math>-ádica e peso gaussiano</b>			
Grau do polinômio	1	2	3
Nº folhas	256	256	256
Critérios de parada 0:(%dados)	100%	100%	73%
Critérios de parada 1:(%dados)	0%	0%	26%
Critérios de parada 2:(%dados)	0%	0%	1%
Critérios de parada 3:(%dados)	0%	0%	0%
Erro de aproximação da $F$	$1.27 \times 10^{-3}$	$6.94 \times 10^{-6}$	$1.08 \times 10^{-7}$
Máximo dos EQM nas folhas para $F$	$2.02 \times 10^{-2}$	$2.89 \times 10^{-4}$	$5.36 \times 10^{-5}$
Mínimo dos EQM nas folhas para $F$	$5.47 \times 10^{-4}$	$1.97 \times 10^{-7}$	$3.74 \times 10^{-10}$
Média dos EQM nas folhas para $F$	$6.14 \times 10^{-3}$	$4.98 \times 10^{-5}$	$9.02 \times 10^{-7}$
Regressão dentro da amostra			
EQM ylocal	$2.94 \times 10^{-1}$	$2.94 \times 10^{-1}$	$2.94 \times 10^{-1}$
EQM yglobal	$7.43 \times 10^{-2}$	$7.43 \times 10^{-3}$	$7.43 \times 10^{-3}$
MAPE ylocal	$7.17 \times 10^{-1}$	$7.17 \times 10^{-1}$	$7.17 \times 10^{-1}$
MAPE yglobal	$3.78 \times 10^{-1}$	$3.77 \times 10^{-1}$	$3.77 \times 10^{-1}$

Tabela 7.32: Aproximação do volume no  $\mathbb{R}^4$  por Malha  $2^n$ -ádica.

Cota do erro de aproximação de $F$ em cada folha $k$	$10^{-10}$
Cota do desvio padrão dos $y$	0.1

Tabela 7.33: Dados da subdivisão espacial do volume no  $\mathbb{R}^4$  pela Árvore BSP.



<b>Aproximação do volume no <math>\mathbb{R}^4</math> pela Árvore BSP e peso polinomial</b>			
Grau do polinômio	1	2	3
Nível Máximo da árvore	7	6	5
Nº folhas	128	64	32
Critérios de parada 0:(%dados)	100%	100%	97%
Critérios de parada 1:(%dados)	0%	0%	3%
Critérios de parada 2:(%dados)	0%	0%	0%
Critérios de parada 3:(%dados)	0%	0%	0%
Erro de aproximação da $F$	1.3139	$8.36 \times 10^{-1}$	2.1229
Máximo dos EQM nas folhas para $F$	$1.44 \times 10^{-2}$	$2.97 \times 10^{-4}$	$6.20 \times 10^{-6}$
Mínimo dos EQM nas folhas para $F$	$7.07 \times 10^{-8}$	$2.15 \times 10^{-9}$	$1.72 \times 10^{-11}$
Média dos EQM nas folhas para $F$	$1.40 \times 10^{-3}$	$2.45 \times 10^{-5}$	$7.60 \times 10^{-7}$
Regressão dentro da amostra			
Lambda	0.5	0.5	0.5
EQM ylocal	$5.59 \times 10^{-1}$	$6.37 \times 10^{-1}$	$7.79 \times 10^{-1}$
EQM yglobal	$3.48 \times 10^{-3}$	$1.86 \times 10^{-2}$	$1.57 \times 10^{-1}$
MAPE ylocal	$8.35 \times 10^{-1}$	$8.81 \times 10^{-1}$	1.0547
MAPE yglobal	$1.40 \times 10^{-1}$	$4.43 \times 10^{-1}$	1.0446
<b>Aproximação do volume no <math>\mathbb{R}^4</math> pela Árvore BSP e peso gaussiano</b>			
Grau do polinômio	1	2	3
Nível Máximo da árvore	7	7	7
Nº folhas	128	64	32
Critérios de parada 0:(%dados)	100%	100%	97%
Critérios de parada 1:(%dados)	0%	0%	3%
Critérios de parada 2:(%dados)	0%	0%	0%
Critérios de parada 3:(%dados)	0%	0%	0%
Erro de aproximação da $F$	$3.51 \times 10^{-1}$	$5.36 \times 10^{-2}$	$2.34 \times 10^{-1}$
Máximo dos EQM nas folhas para $F$	$1.44 \times 10^{-2}$	$2.97 \times 10^{-4}$	$6.20 \times 10^{-6}$
Mínimo dos EQM nas folhas para $F$	$7.08 \times 10^{-8}$	$2.15 \times 10^{-9}$	$1.72 \times 10^{-11}$
Média dos EQM nas folhas para $F$	$1.40 \times 10^{-3}$	$2.45 \times 10^{-5}$	$7.60 \times 10^{-7}$
Regressão dentro da amostra			
Lambda	0.5	0.5	0.5
EQM ylocal	$9.90 \times 10^{-1}$	$9.92 \times 10^{-1}$	$9.96 \times 10^{-1}$
EQM yglobal	$4.66 \times 10^{-7}$	$4.14 \times 10^{-5}$	$7.45 \times 10^{-3}$
MAPE ylocal	$9.998 \times 10^{-1}$	1.0006	1.0015
MAPE yglobal	$1.90 \times 10^{-3}$	$2.14 \times 10^{-2}$	$2.60 \times 10^{-1}$

Tabela 7.34: Aproximação do volume no  $\mathbb{R}^4$  pela Árvore BSP.

**Resultados da regressão fora da amostra**

Realizamos a regressão utilizando o peso polinomial da Malha  $2^n$ -ádica (estimativa global) e Árvore BSP com peso gaussiano (estimativa global) e, em ambos os casos, trabalhamos com polinômios de grau 1.

EQM	Malha $2^n$ -ádica1	BSP e grau1	SVR	BayesianRR	CART
	yglobal	yglobal			
Mínimo	$4.02 \times 10^{-4}$	$3.06 \times 10^{-8}$	$4.89 \times 10^{-4}$	$5.98 \times 10^{-2}$	$6.25 \times 10^{-2}$
Máximo	$6.56 \times 10^{-4}$	$5.54 \times 10^{-2}$	$6.35 \times 10^{-4}$	$6.84 \times 10^{-2}$	$7.67 \times 10^{-2}$
Mediana	$5.15 \times 10^{-4}$	$6.69 \times 10^{-8}$	$5.20 \times 10^{-4}$	$6.48 \times 10^{-2}$	$7.03 \times 10^{-2}$
MAD	$7.22 \times 10^{-5}$	$1.40 \times 10^{-8}$	$1.95 \times 10^{-5}$	$2.72 \times 10^{-3}$	$5.75 \times 10^{-3}$
MAPE	yglobal	yglobal	SVR	BayesianRR	CART
Mínimo	$1.04 \times 10^{-1}$	$1.10 \times 10^{-3}$	$1.12 \times 10^{-1}$	1.3341	1.2814
Máximo	$4.28 \times 10^{-1}$	$5.64 \times 10^{-1}$	$4.06 \times 10^{-1}$	2.7944	9.5311
Mediana	$1.70 \times 10^{-1}$	$2.08 \times 10^{-3}$	$2.13 \times 10^{-1}$	1.7937	6.8786
MAD	$6.13 \times 10^{-2}$	$6.73 \times 10^{-4}$	$7.58 \times 10^{-2}$	$2.76 \times 10^{-1}$	3.0714

Tabela 7.35: Volume no  $\mathbb{R}^4$ . Regressão fora da amostra.

A estimativa global dada pela Árvore BSP com peso gaussiano e grau do polinômio igual 1 supera todos os outros métodos. Já a estimativa global dada pela Malha  $2^n$ -ádica, embora não supere aquela dada pela Árvore BSP, apresenta desempenho melhor do que o CART e o BayesianRR e até o SVR, onde só não apresenta o menor valor do máximo do EQM e MAPE e do MAD do EQM e do MAPE, mas apresenta valores muito próximos, mostrando seu bom desempenho.

Um nova regressão fora da amostra também será efetuada, no entanto, a cada dado de entrada  $\mathbf{x}$  associaremos uma nova saída  $y$  dada por:  $f(\mathbf{x}) = \sin(x_1) + \cos(x_3) + e^{x_2}$ . Esta regressão será identificada como regressão fora da amostra 2 e os resultados são exibidos na tabela 7.36.

Na regressão fora da amostra exibida na tabela 7.36, vemos que os métodos RCVI com subdivisão espacial dada pela Malha  $2^n$ -ádica, peso polinomial, grau do polinômio 1 e pela Árvore BSP, peso gaussiano, grau do polinômio 1 superam os outros métodos utilizados para comparação e, a partição dada pela árvore BSP supera aquela dada pela Malha  $2^n$ -ádica, ressaltando-se que utilizou-se apenas as estimativas globais.

EQM	Malha $2^n$ -ádica1	BSP e grau1	SVR	BayesianRR	CART
	yglobal	yglobal			
Mínimo	$8.72 \times 10^{-4}$	$6.05 \times 10^{-8}$	$2.78 \times 10^{-3}$	$2.77 \times 10^{-2}$	$5.12 \times 10^{-2}$
Máximo	$1.15 \times 10^{-3}$	3.0982	$3.13 \times 10^{-3}$	$3.93 \times 10^{-2}$	$6.43 \times 10^{-2}$
Mediana	$9.94 \times 10^{-4}$	$1.21 \times 10^{-7}$	$2.99 \times 10^{-3}$	$3.58 \times 10^{-2}$	$5.78 \times 10^{-2}$
MAD	$6.88 \times 10^{-5}$	$4.61 \times 10^{-8}$	$1.09 \times 10^{-4}$	$5.71 \times 10^{-4}$	$3.14 \times 10^{-2}$
MAPE	yglobal	yglobal	SVR	BayesianRR	CART
Mínimo	$1.19 \times 10^{-2}$	$1.08 \times 10^{-4}$	$2.52 \times 10^{-2}$	$7.10 \times 10^{-2}$	$9.42 \times 10^{-2}$
Máximo	$1.31 \times 10^{-2}$	1.0086	$2.88 \times 10^{-2}$	$8.55 \times 10^{-2}$	$1.11 \times 10^{-1}$
Mediana	$1.25 \times 10^{-2}$	$1.46 \times 10^{-4}$	$2.76 \times 10^{-2}$	$8.10 \times 10^{-2}$	$1.04 \times 10^{-1}$
MAD	$4.09 \times 10^{-4}$	$3.02 \times 10^{-5}$	$3.38 \times 10^{-4}$	$1.32 \times 10^{-3}$	$4.73 \times 10^{-3}$

Tabela 7.36: Volume no  $\mathbb{R}^4$ . Regressão fora da amostra 2.

## 7.2

### Dados reais

Os dados utilizados foram obtidos do Arquivo de Aprendizagem de Máquina Irvine, da Universidade da Califórnia, disponível no endereço <http://www.ics.uci.edu/mlearn/MLRepository.html> e do arquivo DELVE da Universidade de Toronto, disponível em <http://www.cs.toronto.edu/delve>. Foram utilizados os seguintes dados: “Abalone”, “Boston housing” e “Computer activity”.

- a) **Dados *Abalone***: Nesse conjunto de dados, a variável a ser prevista representa a idade de um tipo de molusco e as variáveis que serão usadas nesta previsão são medições físicas deste molusco. O conjunto de dados possui 10 atributos. Como um dos atributos refere-se ao gênero, um mapeamento é realizado dos códigos de gênero M(macho), F(fêmea) e I (criança) para (1,0,0), (0,1,0) e (0,0,1), respectivamente. Os dados são normalizados coordenada a coordenada para média zero e variância unitária. O conjunto de dados normalizados é aleatoriamente dividido em 3000 dados para treino e 1177 para teste. O particionamento ocorre 10 vezes independentemente.
- b) **Dados *Boston Housing***: A variável a ser prevista representa o valor de uma moradia na área de Boston Massachusetts. O conjunto de dados possui 14 atributos, incluindo o valor a ser estimado, possui 506 dados que são divididos em 481 para treino e 25 para teste. Devido ao pequeno tamanho do conjunto teste, o experimento é repetido 20 vezes sempre utilizando como conjunto teste, observações que não haviam sido incluídos no conjunto teste anteriormente, portanto, usando virtualmente

todas as observações para teste. Esta estratégia é a mesma que foi utilizada no trabalho de (47).

- c) **Dados *Computer Activity***: Neste conjunto de dados, busca-se prever a atividade de um sistema computacional a partir de medidas de desempenho do sistema. O conjunto de dados possui 22 atributos, incluindo a variável a ser prevista, 8192 dados, dos quais, 2000 são utilizados para treino e o restante para teste. Os resultados foram obtidos depois de 10 rodadas do experimento com conjunto treino e teste selecionados de forma aleatória.

Os resultados obtidos pelo método RCVI são comparados com outros métodos e podem ser vistos na tabela 7.37, a qual foi obtida de (47). Os métodos utilizados nessa comparação já foram brevemente explicados no capítulo 3. Para o método RCVI, apenas consideramos as estimativas globais por apresentarem os melhores resultados, fato observado nos testes anteriores e consideramos apenas a partição BSP gaussiano, bem como polinômios de grau 1. A partição usando malha  $2^n$ -ádica fica inviável devido ao grande número de atributos que os dados possuem. Na tabela o nosso método está identificado por RCVI (BSP gaussiano1).

As estatísticas empregadas na comparação foram o erro absoluto médio (MAE, *Mean Absolute Error*) e a raiz do erro quadrado médio (RMS, *Root Mean Squared Error*) e cujas fórmulas são:

$$MAE = \frac{1}{N} \sum_{j=1}^N |y_j - \hat{y}_j| \quad RMS = \sqrt{\frac{\sum_{j=1}^N (y_i - \hat{y}_j)^2}{N}}$$

onde  $y$  representa o valor de  $y$  real e  $\hat{y}$  o valor previsto de  $y$ .

Os métodos BSM e GPR1 utilizaram 400 dados para treino. Os métodos Sparse GPR e GPR2 utilizaram 4000 amostras para treinar o conjunto de dados Abalone. O número de amostras de treino e teste para Online GPR, BCM2, inductive e transductive SRM não são fornecidos pelos autores. Os métodos não citados utilizaram a mesma divisão treino/teste utilizada pelo nosso modelo.

Com os dados Abalone, o método RCVI apresentou um bom desempenho, superando os métodos Sparse GPR, GPR2 e Tensor Voting e alcançando resultados comparáveis com os métodos restantes. Já com os dados Boston Housing não obtivemos um bom desempenho em relação ao MAE, embora com o RMS superamos o dos métodos RVM e SVM. Um dos motivos para o desempenho

	Abalone		Housing		Computer	
	MAE	RMS	MAE	RMS	MAE	RMS
BCM (77)	-	-	-	3.100	-	-
GPR1 (77)	-	-	-	3.013	-	-
RVM (74)	-	-	-	8.04	-	-
SVM (74)	-	-	-	7.46	-	-
Sparse GPR (66)	1.785	-	-	-	-	-
GPR2 (66)	1.782	-	-	-	-	-
Online GPR (63)	-	2.111	-	-	-	-
BCM2 (63)	-	2.111	-	-	-	-
Inductive SRM (63)	-	2.109	-	-	-	-
Transductive SRM (63)	-	2.109	-	-	-	-
SVR (13)	1.421	2.141	2.13	3.205	2.28	3.715
BSVR (13)	1.464	2.134	2.19	3.513	2.33	4.194
GPR-ARD (13)	1.493	2.134	2.01	2.884	1.686	2.362
BSVR-ARD (13)	1.454	2.119	1.86	2.645	1.687	2.408
Tensor voting (47)	1.630	2.500	1.272	1.860	1.970	2.815
RCVI (BSP gaussiano1)	1.52	2.21	2.61	3.96	3.51	5.40

Tabela 7.37: Regressão com dados reais

obtido se deve a termos 13 atributos e somente 481 dados para treino, o que dificulta uma boa aproximação da variedade e, uma boa previsão das saídas. Com relação ao conjunto de dados Computer Activity, o desempenho obtido não foi satisfatório e o motivo também pode ser o fato de trabalharmos com 21 atributos e apenas 2000 dados para treino. Além disso, o algoritmo estimou a dimensionalidade dos dados como sendo 1 (um), isto implica em termos que ajustar 20 polinômios a cada subdivisão da árvore e tendo apenas 2000 dados treino, o que dificultou o ajuste. Para tentar obter um melhor ajuste consideramos um novo experimento com 5000 dados treino e o restante para teste. O resultado obtido foi MAE=1.917 e RMS=2.701. Com essa nova divisão dos dados treino/teste, o método RCVI teve uma melhora considerável no seu desempenho. Em resumo, acreditamos que dentro das limitações encontradas devido ao número de amostras insuficientes para descrever variedades complexas, ainda assim, o método RCVI teve um desempenho satisfatório.

### 7.3

#### Aplicação em Geologia

As atividades e estudos que visam definir em termos qualitativos e quantitativos o potencial de um reservatório de petróleo, isto é, a sua capacidade produtiva e a valoração das suas reservas de óleo e gás são chamadas de “Avaliação de Formações”, conforme explicado no trabalho de Thomas (73) e no

qual nos baseamos para explicar esta aplicação.

A avaliação das formações baseia-se principalmente na perfilagem a poço aberto <sup>1</sup>, no teste de formação <sup>2</sup> a poço aberto, nos teste de pressão a poço revestido e na perfilagem de produção. Também são aproveitadas todas as informações anteriores à perfilagem do intervalo de interesse, sejam elas obtidas na etapa do estudo geológico e geofísico da área ou na etapa de perfuração do poço. A integração de todos os dados disponíveis permite a avaliação efetiva do reservatório.

Com base na análise dos perfis, decide-se quais intervalos do poço são de interesse econômico potencial para se executar os testes de formação. Se não houver intervalos de interesse o poço é abandonado. Os testes de formação têm sido amplamente utilizados na indústria petrolífera para se estimar a capacidade produtiva do poço. Entretanto, muitos dados sobre poços contém um grande número de dados faltantes que poderiam auxiliar na tomada de decisões. Portanto, uma aplicação que nos propomos a realizar é a estimativa desses dados faltantes dentro de uma amostra já obtida. Para isso, nos propomos a realizar uma regressão onde a variável de predição é composta por esses dados faltantes e as variáveis explicativas são os dados completos.

A implementação do método RCVI baseia-se na hipótese de que os dados de entrada podem fazer parte de um espaço de baixa dimensionalidade, assim, o objetivo é identificar a estrutura da qual os dados fazem parte e daí aproximar essa estrutura e, em seguida, realizar a previsão. O mesmo experimento é realizado com outros métodos e os resultados obtidos e a informação sobre cada um dos poços utilizados é dada a seguir.

Os dois tipos de perfis utilizados como variável a ser prevista são:

- RHOB: Detecta os raios gama defletidos pelos elétrons orbitais dos elementos componentes das rochas, após terem sido emitidos por uma fonte colimada situada dentro do poço. Além da densidade das camadas, permite o cálculo da porosidade e a identificação das zonas de gás. Também é utilizado como apoio à sísmica para o cálculo do sismograma sintético.

<sup>1</sup>determina o perfil de um poço obtido através do deslocamento contínuo de um sensor de perfilagem (sonda) dentro do poço com a finalidade de medir algumas propriedades das formações, fundamentais para caracterização e avaliação econômica. Esses perfis são genericamente denominados de perfis elétricos, independentemente do processo físico de medição utilizado.

<sup>2</sup>o teste de formação, isto é, a colocação do poço em fluxo, pode confirmar, com segurança, a presença de hidrocarbonetos na formação e fornecer dados a respeito das condições de fluxo nas imediações do poço.

- DT: Mede a diferença nos tempos de trânsito de uma onda mecânica através das rochas. É utilizado para estimativas de porosidade, correlação poço a poço, estimativas do grau de compactação das rochas ou estimativa das constantes elásticas, detecção de fraturas e apoio à sísmica para a elaboração do sismograma sintético.

O experimento foi realizado com dados de 4 poços diferentes localizados no Alaska, EUA e cujos dados podem ser obtidos no endereço <http://nerslweb.cr.usgs.gov>. Cada poço foi rodado separadamente e de duas formas diferentes. Em uma primeira etapa, consideramos RHOB como a variável a ser prevista e excluimos a variável DT das variáveis explicativas utilizadas. Já em um segundo momento, tomamos DT como a variável a ser prevista e excluimos RHOB dentre as variáveis explicativas utilizadas para previsão. Para cada teste tomamos 90% dos dados para treino e 10% para teste e o experimento foi rodado 10 vezes e em cada uma das vezes uma amostra diferente foi obtida aleatoriamente.

As variáveis explicativas incluem:

- DEPTH: profundidade do ponto de leitura do poço. A profundidade possui correlação com outras variáveis de perfis. De forma geral, a velocidade sônica das rochas ( $1/DT$ ) e a densidade (RHOB) crescem com a profundidade em função do soterramento e consequente compactação das rochas;
- Raios Gama (GR): é a medida da radioatividade total presente nas rochas. Este perfil é utilizado principalmente para separar tipos diferentes de rochas, já que as argilas apresentam elevados teores de elementos radioativos e os arenitos e carbonatos têm baixa radioatividade;
- Caliper (CALI): perfil que fornece o diâmetro medido ponto a ponto em toda extensão de um poço. É um indicativo importante da qualidade da leitura dos perfis. Os perfis que correm centralizados no poço são de forma geral os mais afetados (ILD, DT) pelas variações no diâmetro dos poços ao passo que os que correm com sapatas de contato nas paredes do poço são menos afetados;
- NPHI: utilizado para estimativas de porosidade, litologia e detecção de hidrocarbonetos leves ou gás. Assim, quanto maior o valor de NPHI, maior a quantidade de poros e, conseqüentemente, maior a probabilidade do fluido ali existente ser extraído;
- SP: potencial espontâneo, medido pela diferença de potencial entre dois eletrodos, um na superfície e um dentro do poço. É usado para detectar

camadas permoporosas, calcular a argilosidade das rochas e auxiliar na correlação de informações nos poços vizinhos;

- LLS e LLD: perfis de eletrorresistividade;
- ILD: fornece leitura aproximada da resistividade total por meio da medição de propagação de ondas eletromagnéticas e de leituras diagnósticas, onde existem trocas iônicas na superfície de grãos metálicos, tal como acontece em sulfitos. É empregado pela engenharia na diferenciação litológica e em estudos de salinidade de lençóis de água subterrânea;
- ILM: perfil de indução média;
- LL8: *Laterolog raso* mede a resistividade das rochas e dos fluidos nela contidos, por meio de correntes elétricas focalizadas.

Os dados de cada poço utilizado e resultados obtidos são exibidos a seguir. O método RCVI utilizado é o que utiliza subdivisão espacial dada pela Árvore BSP com peso gaussiano e grau do polinômio igual a 1. Os métodos utilizados para comparação foram os mesmos que utilizamos quando trabalhamos com os dados sintéticos. Trabalhamos com SVR por ser um método de aprendizagem estatística que tem tido sucesso em aplicações na área de petróleo, como pode ser visto nos trabalhos (36), (37), (65), (71) e (82). Também utilizamos o método Bayesian Ridge Regression e, como o nosso método é um método de regressão em árvore, utilizamos o tradicional método CART.

### 7.3.1

#### Poço 1

O poço 1, cuja localização pode ser vista na figura 7.1, possui 8 atributos e 7664 dados para previsão de RHOB e 7565 para previsão de DT. A tabela 7.38 exhibe um resumo das variáveis de predição e as estatísticas utilizadas são o EQM, RMS e o MAE.



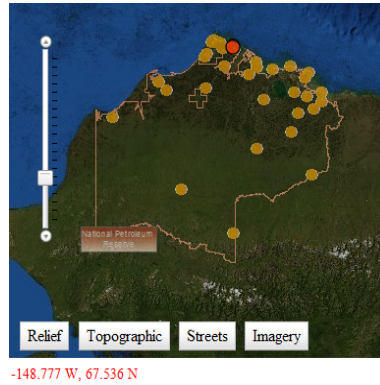


Figura 7.1: Poço 1

Poço 1: DT				
Mínimo	Máximo	Média	Mediana	Desvio padrão
53.2702	201.0129	109.4689	106.0714	20.4662
Poço 1: RHOB				
Mínimo	Máximo	Média	Mediana	Desvio padrão
1.9525	2.7327	2.4011	2.4291	0.1063

Tabela 7.38: Resumo dos dados do poço 1.

EQM	RCVI	SVR	BayesianRR	CART
Mínimo	<b>15.7940</b>	17.5113	41.3493	32.4903
Máximo	<b>21.2437</b>	22.2272	47.1066	43.5926
Mediana	<b>17.1240</b>	20.6438	45.0081	37.4314
Média	<b>18.0168</b>	20.5681	44.7352	37.5349
MAD	1.1140	<b><math>7.69 \times 10^{-1}</math></b>	1.4287	1.3238
RMS	RCVI	SVR	BayesianRR	CART
Mínimo	<b>3.9742</b>	4.1846	6.4303	5.7000
Máximo	<b>4.6091</b>	4.7146	6.8634	6.6025
Mediana	<b>4.1381</b>	4.5435	6.7088	6.1181
Média	<b>4.2390</b>	4.5330	6.6871	6.1223
MAE	RCVI	SVR	BayesianRR	CART
Mínimo	2.8102	<b>2.7993</b>	5.0440	4.2472
Máximo	3.1853	<b>3.0396</b>	5.2807	4.8175
Mediana	<b>2.8750</b>	2.9624	5.1306	4.5851
Média	<b>2.9303</b>	2.9512	5.1536	4.5644

Tabela 7.39: Dados do poço 1. Variável prevista: DT.

As tabelas 7.39 e 7.40 exibem o ótimo desempenho do RCVI em relação aos outros métodos.

EQM	RCVI	SVR	BayesianRR	CART
Mínimo	$1.19 \times 10^{-3}$	$1.34 \times 10^{-3}$	$3.54 \times 10^{-3}$	$2.26 \times 10^{-3}$
Máximo	$1.48 \times 10^{-3}$	$1.70 \times 10^{-3}$	$4.22 \times 10^{-3}$	$3.28 \times 10^{-3}$
Mediana	$1.38 \times 10^{-3}$	$1.60 \times 10^{-3}$	$3.88 \times 10^{-3}$	$2.69 \times 10^{-3}$
Média	$1.35 \times 10^{-3}$	$1.56 \times 10^{-3}$	$3.90 \times 10^{-3}$	$2.73 \times 10^{-3}$
MAD	$7.16 \times 10^{-5}$	$6.19 \times 10^{-5}$	$1.87 \times 10^{-4}$	$2.24 \times 10^{-4}$
RMS	RCVI	SVR	BayesianRR	CART
Mínimo	$3.46 \times 10^{-2}$	$3.65 \times 10^{-2}$	$5.95 \times 10^{-2}$	$4.76 \times 10^{-2}$
Máximo	$3.85 \times 10^{-2}$	$4.13 \times 10^{-2}$	$6.50 \times 10^{-2}$	$5.73 \times 10^{-2}$
Mediana	$3.71 \times 10^{-2}$	$3.40 \times 10^{-2}$	$6.22 \times 10^{-2}$	$5.19 \times 10^{-2}$
Média	$3.67 \times 10^{-2}$	$3.95 \times 10^{-2}$	$6.24 \times 10^{-2}$	$5.22 \times 10^{-2}$
MAE	RCVI	SVR	BayesianRR	CART
Mínimo	$2.57 \times 10^{-2}$	$2.70 \times 10^{-2}$	$4.51 \times 10^{-2}$	$3.63 \times 10^{-2}$
Máximo	$2.80 \times 10^{-2}$	$3.00 \times 10^{-2}$	$4.94 \times 10^{-2}$	$4.18 \times 10^{-2}$
Mediana	$2.70 \times 10^{-2}$	$2.85 \times 10^{-2}$	$4.70 \times 10^{-2}$	$3.79 \times 10^{-2}$
Média	$2.68 \times 10^{-2}$	$2.85 \times 10^{-2}$	$4.74 \times 10^{-2}$	$3.86 \times 10^{-2}$

Tabela 7.40: Dados do poço 1. Variável prevista: RHOB.

### 7.3.2 Poço 2

A localização é dada na figura 7.2, possui 9 atributos e 10901 dados para previsão de RHOB e 14937 para previsão de DT. A tabela 7.41 exibe um resumo das variáveis de predição.

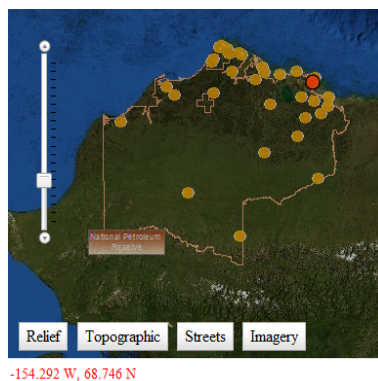


Figura 7.2: Poço 2

Com os dados do poço 2 o método RCVI apresenta resultados muito bons. As tabelas 7.42 e 7.43 mostram os resultados.

Poço 2: DT				
Mínimo	Máximo	Média	Mediana	Desvio padrão
42.6508	164.4016	93.5966	95.2924	21.2711
Poço 2: RHOB				
Mínimo	Máximo	Média	Mediana	Desvio padrão
1.6513	2.7057	2.4393	2.4766	0.1123

Tabela 7.41: Resumo dos dados do poço 2.

EQM	RCVI	SVR	BayesianRR	CART
Mínimo	<b>9.8629</b>	12.6846	30.4346	40.1390
Máximo	<b>12.9973</b>	15.8787	34.2639	48.4148
Mediana	<b>12.2797</b>	13.5495	31.9069	45.1355
Média	<b>11.7975</b>	13.7747	32.1813	44.9466
MAD	<b><math>6.42 \times 10^{-1}</math></b>	$7.44 \times 10^{-1}$	$0.8.98 \times 10^{-1}$	1.7829
EQM	RCVI	SVR	BayesianRR	CART
Mínimo	<b>3.1405</b>	3.5615	5.5168	6.3355
Máximo	<b>3.6052</b>	3.9848	5.8535	6.9581
Mediana	<b>3.5038</b>	3.6809	5.6486	6.7183
Média	<b>3.4307</b>	3.7091	5.6718	6.7021
EQM	RCVI	SVR	BayesianRR	CART
Mínimo	<b>2.0998</b>	2.2974	4.1789	4.8888
Máximo	<b>2.3903</b>	2.4721	4.3890	5.2924
Mediana	<b>2.2415</b>	2.4033	4.2624	5.1389
Média	<b>2.2475</b>	2.3993	4.2705	5.1245

Tabela 7.42: Dados do poço 2. Variável prevista: DT.

EQM	RCVI	SVR	BayesianRR	CART
Mínimo	$6.24 \times 10^{-4}$	<b><math>6.17 \times 10^{-4}</math></b>	$3.03 \times 10^{-3}$	$1.75 \times 10^{-3}$
Máximo	<b><math>7.16 \times 10^{-4}</math></b>	$8.54 \times 10^{-4}$	$4.09 \times 10^{-3}$	$2.52 \times 10^{-3}$
Mediana	$6.68 \times 10^{-4}$	<b><math>6.45 \times 10^{-4}</math></b>	$3.30 \times 10^{-3}$	$1.94 \times 10^{-3}$
Média	<b><math>6.70 \times 10^{-4}</math></b>	$7.01 \times 10^{-4}$	$3.42 \times 10^{-3}$	$2.01 \times 10^{-3}$
MAD	<b><math>2.38 \times 10^{-5}</math></b>	$2.76 \times 10^{-5}$	$8.28 \times 10^{-5}$	$1.31 \times 10^{-4}$
EQM	RCVI	SVR	BayesianRR	CART
Mínimo	$2.50 \times 10^{-2}$	<b><math>2.48 \times 10^{-2}</math></b>	$5.50 \times 10^{-2}$	$4.18 \times 10^{-2}$
Máximo	<b><math>2.68 \times 10^{-2}</math></b>	$2.92 \times 10^{-2}$	$6.40 \times 10^{-2}$	$5.01 \times 10^{-2}$
Mediana	$2.58 \times 10^{-2}$	<b><math>2.54 \times 10^{-2}</math></b>	$4.06 \times 10^{-2}$	$4.40 \times 10^{-2}$
Média	<b><math>2.59 \times 10^{-2}</math></b>	$2.64 \times 10^{-2}$	$5.85 \times 10^{-2}$	$4.47 \times 10^{-2}$
EQM	RCVI	SVR	BayesianRR	CART
Mínimo	$1.76 \times 10^{-2}$	<b><math>1.75 \times 10^{-2}</math></b>	$4.06 \times 10^{-2}$	$3.10 \times 10^{-2}$
Máximo	<b><math>1.87 \times 10^{-2}</math></b>	$1.88 \times 10^{-2}$	$4.36 \times 10^{-2}$	$3.54 \times 10^{-2}$
Mediana	$1.83 \times 10^{-2}$	<b><math>1.80 \times 10^{-2}</math></b>	$4.17 \times 10^{-2}$	$3.23 \times 10^{-2}$
Média	$1.82 \times 10^{-2}$	<b><math>1.81 \times 10^{-2}</math></b>	$4.18 \times 10^{-2}$	$3.25 \times 10^{-2}$

Tabela 7.43: Dados do poço 2. Variável prevista: RHOB.

### 7.3.3 Poço 3

A localização é dada na figura 7.3, possui 9 atributos e 15488 dados para previsão de RHOB e 15466 para previsão de DT. A tabela 7.44 exhibe um resumo das variáveis de predição.

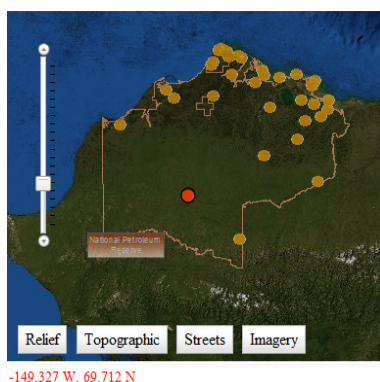


Figura 7.3: Poço 3

Poço 3: DT				
Mínimo	Máximo	Média	Mediana	Desvio padrão
39.1966	180.1528	72.6131	70.2786	14.2646
Poço 3: RHOB				
Mínimo	Máximo	Média	Mediana	Desvio padrão
1.4475	2.9668	2.5538	2.6194	0.1942

Tabela 7.44: Resumo dos dados do poço 3.

O bom desempenho do RCVI obtido nos poços anteriores também se observa no poço 3, cujos resultados estão exibidos nas tabelas 7.45,7.46.

EQM	RCVI	SVR	BayesianRR	CART
Mínimo	<b>18.4642</b>	18.8710	32.3644	29.7693
Máximo	34.3426	<b>24.7444</b>	41.3085	45.4806
Mediana	<b>21.3641</b>	23.2188	37.7640	39.4313
Média	22.7132	<b>22.5947</b>	37.5402	38.5176
MAD	1.7370	<b>1.2254</b>	3.1433	2.4834
RMS	RCVI	SVR	BayesianRR	CART
Mínimo	<b>4.2970</b>	4.3441	5.6890	5.4561
Máximo	5.8603	<b>4.9744</b>	6.4272	6.7439
Mediana	<b>4.6221</b>	4.8186	6.1450	6.2794
Média	<b>4.7459</b>	4.7497	6.1213	6.1972
MAE	RCVI	SVR	BayesianRR	CART
Mínimo	2.5583	<b>2.5031</b>	3.6177	3.9445
Máximo	3.2231	<b>2.7668</b>	3.9360	4.3978
Mediana	2.7561	<b>2.6178</b>	3.7649	4.1529
Média	2.8066	<b>2.6244</b>	3.7598	4.1622

Tabela 7.45: Dados do poço 3. Variável prevista: DT.

EQM	RCVI	SVR	BayesianRR	CART
Mínimo	<b><math>2.99 \times 10^{-3}</math></b>	$5.14 \times 10^{-3}$	$1.66 \times 10^{-2}$	$7.28 \times 10^{-3}$
Máximo	$7.01 \times 10^{-3}$	<b><math>6.68 \times 10^{-3}</math></b>	$1.97 \times 10^{-2}$	$9.85 \times 10^{-3}$
Mediana	<b><math>3.78 \times 10^{-3}</math></b>	$5.90 \times 10^{-3}$	$1.84 \times 10^{-2}$	$8.68 \times 10^{-3}$
Média	<b><math>4.22 \times 10^{-3}</math></b>	$5.86 \times 10^{-3}$	$1.84 \times 10^{-2}$	$8.52 \times 10^{-3}$
MAD	$5.66 \times 10^{-4}$	<b><math>2.86 \times 10^{-4}</math></b>	$9.35 \times 10^{-4}$	$7.66 \times 10^{-4}$
RMS	RCVI	SVR	BayesianRR	CART
Mínimo	<b><math>5.47 \times 10^{-2}</math></b>	$7.17 \times 10^{-2}$	$1.29 \times 10^{-2}$	$8.53 \times 10^{-2}$
Máximo	$8.37 \times 10^{-2}$	<b><math>8.17 \times 10^{-2}</math></b>	$1.40 \times 10^{-2}$	$9.92 \times 10^{-2}$
Mediana	<b><math>6.15 \times 10^{-2}</math></b>	$7.67 \times 10^{-2}$	$1.36 \times 10^{-2}$	$9.31 \times 10^{-2}$
Média	<b><math>6.44 \times 10^{-2}</math></b>	$7.65 \times 10^{-2}$	$1.39 \times 10^{-2}$	$9.2 \times 10^{-2}$
MAE	RCVI	SVR	BayesianRR	CART
Mínimo	<b><math>3.51 \times 10^{-2}</math></b>	$3.82 \times 10^{-2}$	$8.40 \times 10^{-2}$	$5.24 \times 10^{-2}$
Máximo	<b><math>4.19 \times 10^{-2}</math></b>	$4.33 \times 10^{-2}$	$9.05 \times 10^{-2}$	$5.96 \times 10^{-2}$
Mediana	<b><math>3.74 \times 10^{-2}</math></b>	$4.05 \times 10^{-2}$	$8.82 \times 10^{-2}$	$5.59 \times 10^{-2}$
Média	<b><math>3.75 \times 10^{-2}</math></b>	$4.03 \times 10^{-2}$	$8.74 \times 10^{-2}$	$5.54 \times 10^{-2}$

Tabela 7.46: Dados do poço 3. Variável prevista: RHOB.

### 7.3.4

#### Poço 4

A localização é dada na figura 7.4, possui 9 atributos e 19895 dados tanto para previsão de RHOB quanto para previsão de DT. A tabela 7.47 exhibe um resumo das variáveis de predição.

Com relação aos resultados obtidos no poço 4 e exibidos nas tabelas 7.48 e 7.49 podemos destacar o excelente desempenho do método RCVI e observar ainda que, conforme nos mostram as tabelas 7.38, 7.41, 7.44 e 7.47

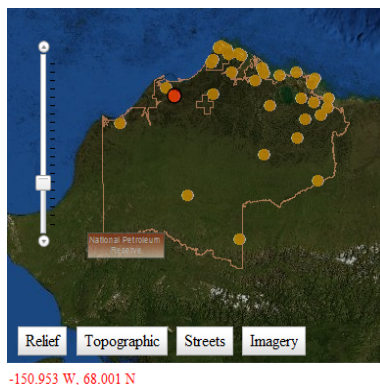


Figura 7.4: Poço 4

Poço 4: DT				
Mínimo	Máximo	Média	Mediana	Desvio padrão
50.0114	158.4382	84.6397	85.8582	12.4714
Poço 4: RHOB				
Mínimo	Máximo	Média	Mediana	Desvio padrão
1.4431	3.0157	2.4191	2.4609	0.2138

Tabela 7.47: Resumo dos dados do poço 4.

EQM	RCVI	SVR	BayesianRR	CART
Mínimo	<b>9.6324</b>	13.1235	30.6563	24.3780
Máximo	<b>17.8791</b>	18.1596	37.9107	29.9389
Mediana	<b>12.0044</b>	14.9013	32.6755	27.3235
Média	<b>12.3103</b>	14.9013	33.3315	27.1970
MAD	1.1263	1.1300	<b><math>9.68 \times 10^{-1}</math></b>	2.7637
RMS	RCVI	SVR	BayesianRR	CART
Mínimo	<b>3.1036</b>	3.6226	5.5368	4.9374
Máximo	<b>4.2284</b>	4.2614	6.1572	5.4716
Mediana	<b>3.4646</b>	3.8602	5.7162	5.2271
Média	<b>3.4956</b>	3.8515	5.7703	5.2117
MAE	RCVI	SVR	BayesianRR	CART
Mínimo	<b>2.1759</b>	2.5194	4.1990	3.6442
Máximo	<b>2.3814</b>	2.7034	4.4581	3.9042
Mediana	<b>2.2674</b>	2.5940	4.2832	3.8523
Média	<b>2.2766</b>	2.6020	4.2960	3.7983

Tabela 7.48: Dados do poço 4. Variável prevista: DT.

a variabilidade nas saídas do poço 4 são semelhantes as obtidas nos outros poços, este poço apresenta o mesmo número de atributos que os outros poços, isto é, 9 atributos, com exceção do poço 1 que apresenta 8 atributos, entretanto, este poço é o que possui o maior número de amostras e foi o poço que o método RCVI teve o seu melhor desempenho, o que mostra que com

EQM	RCVI	SVR	BayesianRR	CART
Mínimo	$3.38 \times 10^{-3}$	$6.42 \times 10^{-3}$	$1.76 \times 10^{-2}$	$1.02 \times 10^{-2}$
Máximo	$5.62 \times 10^{-3}$	$9.39 \times 10^{-3}$	$2.10 \times 10^{-2}$	$1.55 \times 10^{-2}$
Mediana	$4.23 \times 10^{-3}$	$7.72 \times 10^{-3}$	$1.84 \times 10^{-2}$	$1.16 \times 10^{-2}$
Média	$4.35 \times 10^{-3}$	$7.86 \times 10^{-3}$	$1.87 \times 10^{-2}$	$1.20 \times 10^{-2}$
MAD	$4.47 \times 10^{-4}$	$5.30 \times 10^{-4}$	$4.71 \times 10^{-4}$	$1.26 \times 10^{-3}$
RMS	RCVI	SVR	BayesianRR	CART
Mínimo	$5.81 \times 10^{-2}$	$8.01 \times 10^{-2}$	$1.33 \times 10^{-1}$	$1.01 \times 10^{-1}$
Máximo	$7.50 \times 10^{-2}$	$9.69 \times 10^{-2}$	$1.44 \times 10^{-1}$	$1.25 \times 10^{-1}$
Mediana	$6.50 \times 10^{-2}$	$8.79 \times 10^{-2}$	$1.36 \times 10^{-1}$	$1.08 \times 10^{-1}$
Média	$6.57 \times 10^{-2}$	$8.86 \times 10^{-2}$	$1.37 \times 10^{-1}$	$1.09 \times 10^{-1}$
MAE	RCVI	SVR	BayesianRR	CART
Mínimo	$4.24 \times 10^{-2}$	$5.14 \times 10^{-2}$	$9.41 \times 10^{-2}$	$7.27 \times 10^{-2}$
Máximo	$4.81 \times 10^{-2}$	$5.61 \times 10^{-2}$	$1.01 \times 10^{-1}$	$7.97 \times 10^{-2}$
Mediana	$4.37 \times 10^{-2}$	$5.33 \times 10^{-2}$	$9.77 \times 10^{-2}$	$7.49 \times 10^{-2}$
Média	$4.42 \times 10^{-2}$	$5.34 \times 10^{-2}$	$9.75 \times 10^{-2}$	$7.50 \times 10^{-2}$

Tabela 7.49: Dados do poço 4. Variável prevista: RHOB.

o número de observações suficientes para representar a variedade, pode-se alcançar resultados muito satisfatórios.