



Máquinas aleatórias

O espaço aleatório de kernels como uma combinação de
modelos de vetores de suporte

Mateus Maia - (Orientador: Anderson Araújo)

Outline

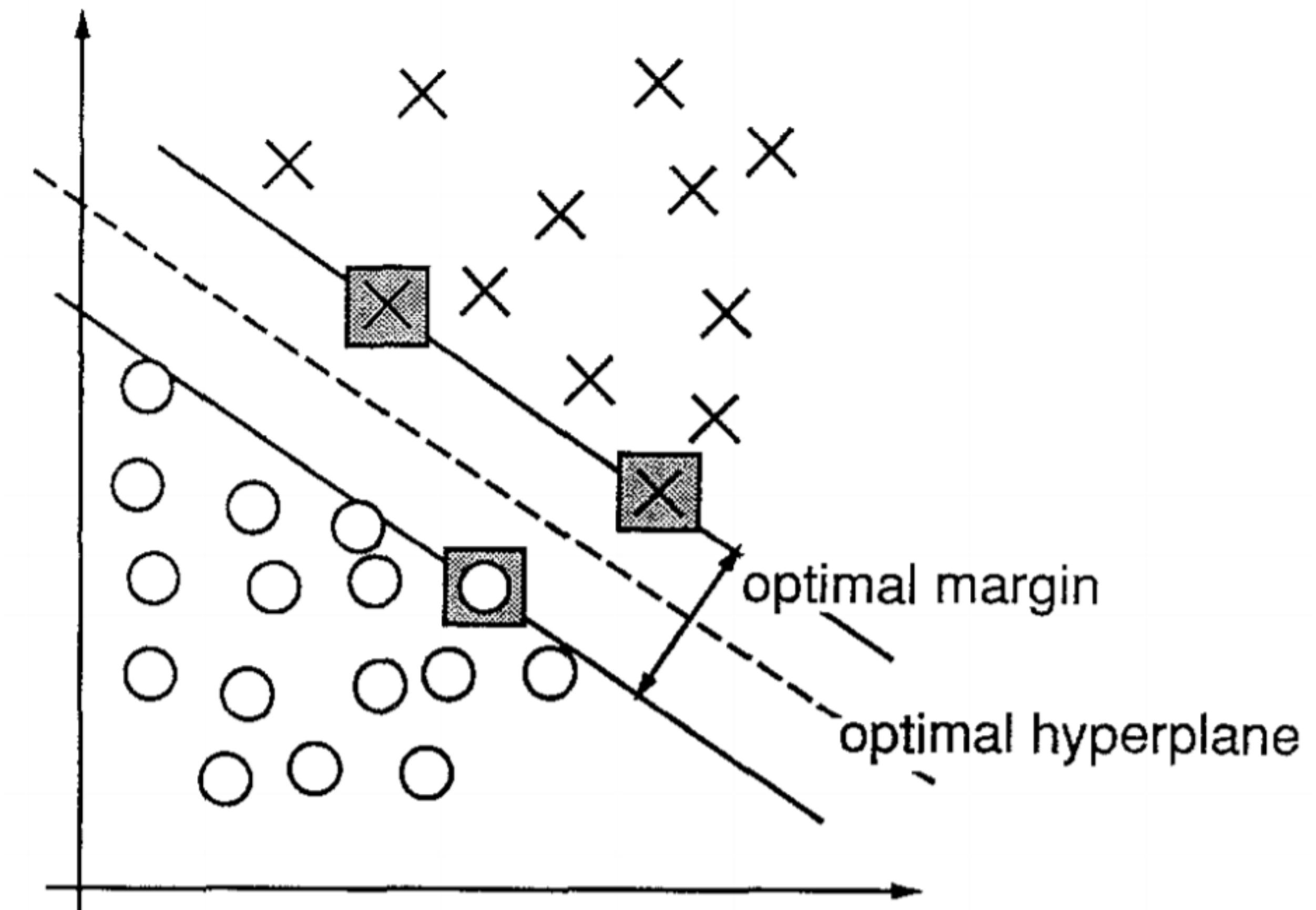
- 1. Motivação:** Por que combinar SVM e técnicas de ensemble?
- 2. Random machines: A bagged-weighted support vector model with free kernel choice** (Ara A., et al. 2021)
- 3. Regression random machines: An ensemble support vector regression model with free kernel choice** (Ara, et al., 2022)
- 4. Predictive Comparison Between Random Machines and Random Forests.** (Maia, et al. 2021)
- 5. Conclusão:** Próximos passos e direcionamentos.

Motivação

Motivação

Máquinas de Vetores de Suporte (SVM)

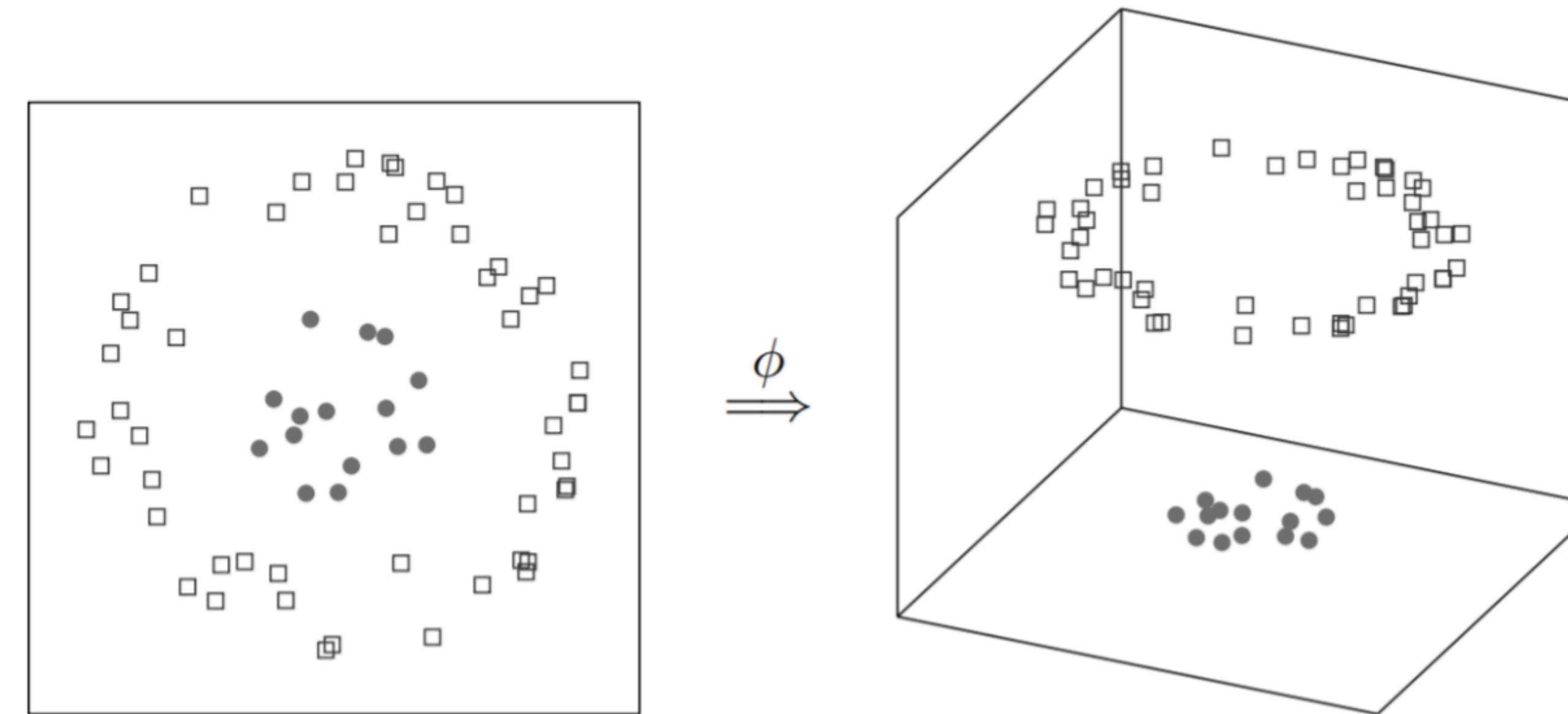
- As máquinas de vetores de suporte (Cortes e Vapnik, 1995), foram introduzidas para solucionar problemas de classificação.
- Forte capacidade preditiva, mesmo quando comparado a outros métodos que são considerados *state-of-art*.
(Maia, Mateus, et al. "Convolutional support vector models: prediction of coronavirus disease using chest X-rays." *Information* 11.12 (2020): 548.)



Retirado de Cortes e Vapnik, 1995.

Motivação

O Truque do Kernel



Retirado de Stamp, 2017

Motivação

O Truque do Kernel

- Através do **Truque do Kernel** é possível mapear implicitamente as observações em espaços dimensionais superiores, onde as classes são linearmente separáveis.

Funções de Kernel

Kernel	K(x,y)	Parâmetros
Kernel Linear	$\gamma(x \cdot y)$	γ
Kernel Polinomial	$(\gamma(x \cdot y))^d$	γ, d
Kernel Gaussiano	$e^{-\gamma x-y ^2}$	γ
Kernel Laplaciano	$e^{-\gamma x-y }$	γ

Em que $\gamma \in \mathbb{R}^+$, $d \in \mathbb{N}$.

Motivação

O Truque do Kernel

- Através do **Truque do Kernel** é possível mapear implicitamente as observações em espaços dimensionais superiores, onde as classes são linearmente separáveis.

Funções de Kernel

Kernel	$K(x,y)$	Parâmetros
Kernel Linear	$\gamma(x \cdot y)$	γ
Kernel Polinomial	$(\gamma(x \cdot y))^d$	γ, d
Kernel Gaussiano	$e^{-\gamma x-y ^2}$	γ
Kernel Laplaciano	$e^{-\gamma x-y }$	γ

Em que $\gamma \in \mathbb{R}^+$, $d \in \mathbb{N}$.



Como selecionar corretamente a Função de Kernel a ser utilizada?

Motivação

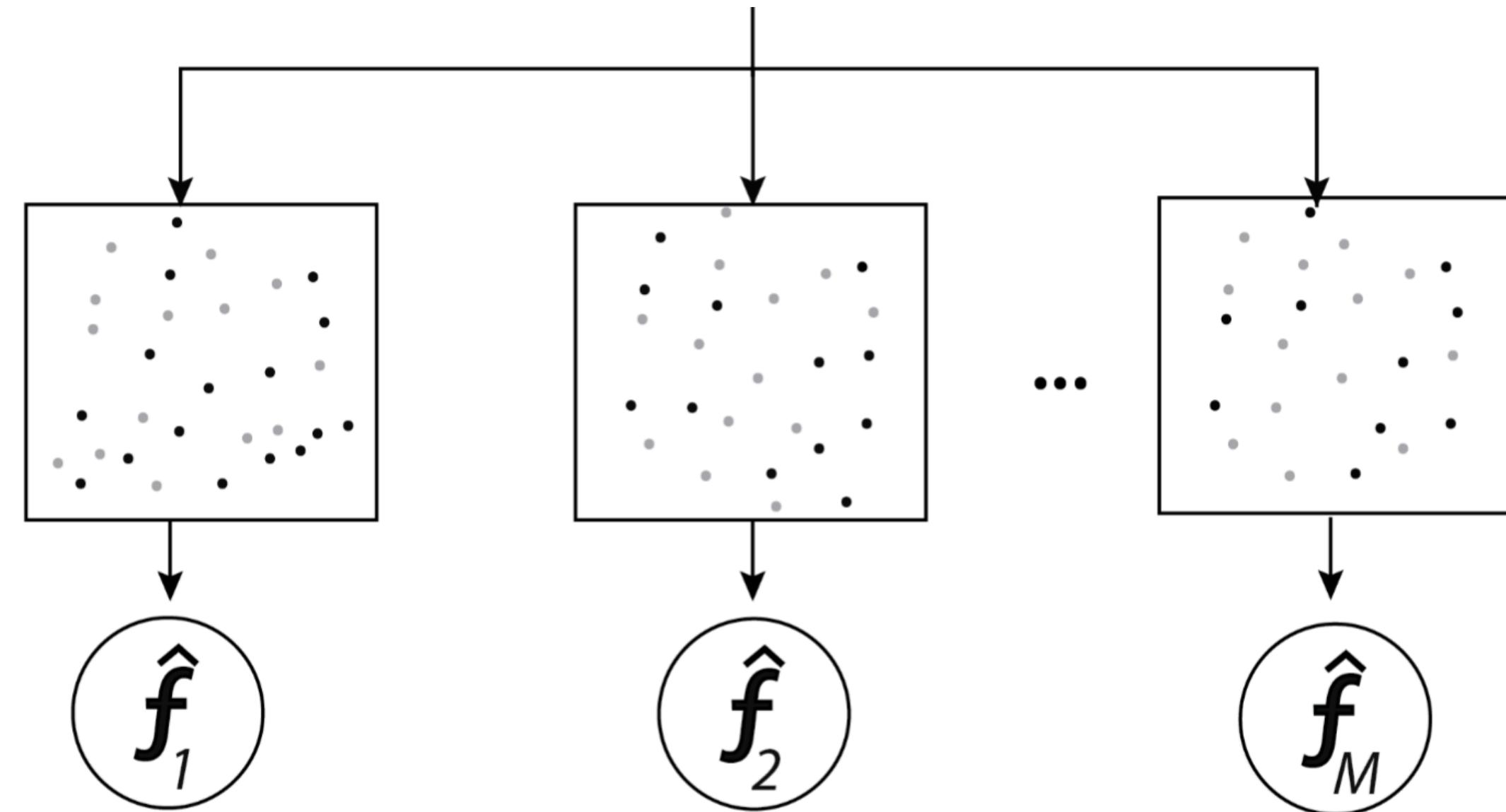
Bagging + Máquinas de Vetores de Suporte

- A utilização do **bagging** acrescentar ganhos substanciais na melhora da capacidade preditiva de um modelo (Breiman, 1996).

Pontos-chave:

Força
E

Diversidade
(Breiman, 2001)



Motivação

Bagging + Máquinas de Vetores de Suporte

- A utilização do **bagging** acrescentar ganhos substanciais na melhora da capacidade preditiva de um modelo (Breiman, 1996).

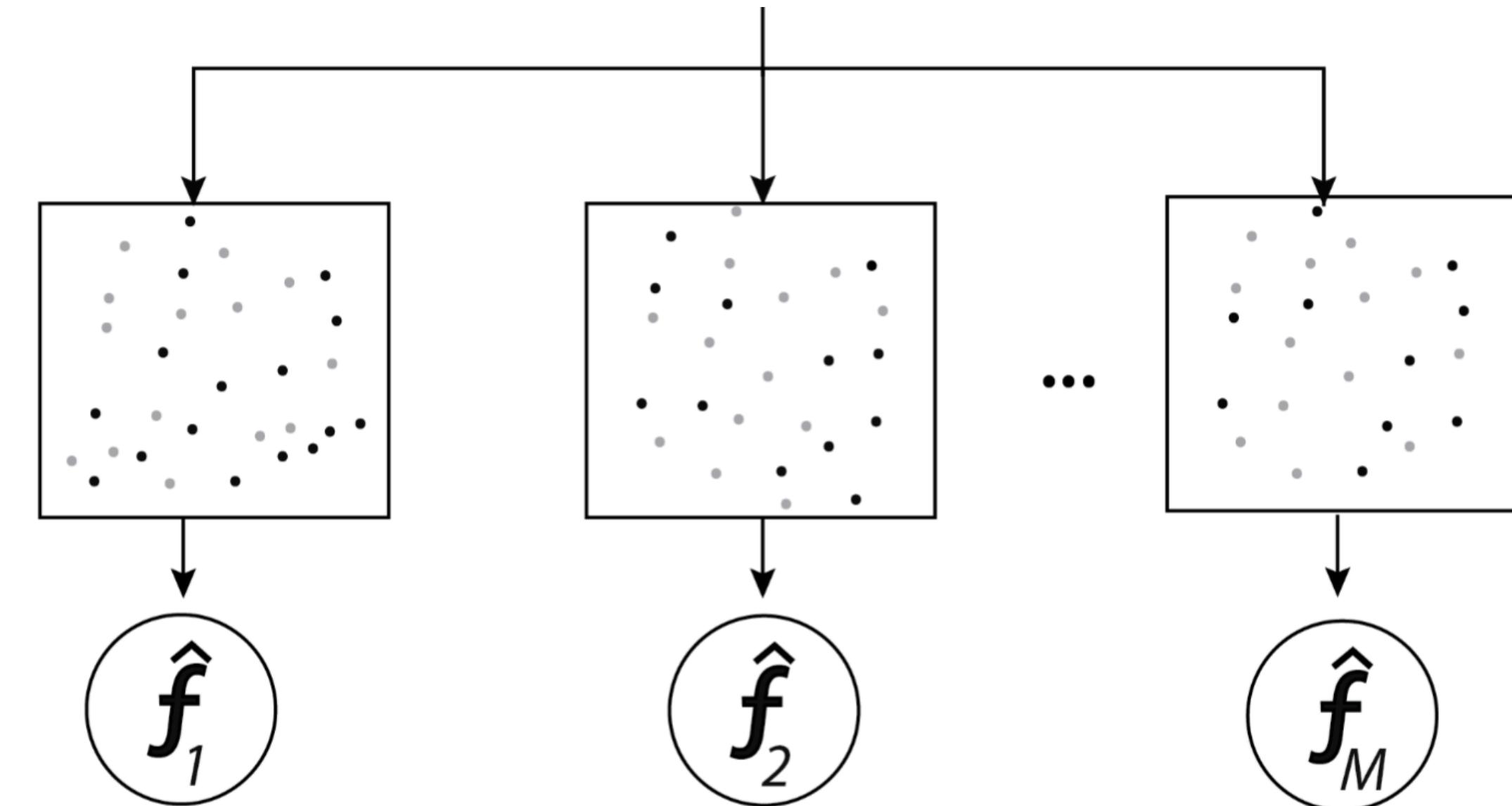
Pontos-chave:

Força

E

Diversidade

(Breiman, 2001)



“The vital element is the instability of the prediction method.

If perturbing the learning set can cause significant changes in the predictor constructed, then bagging can improve accuracy.”

(Leo Breiman, 1996)

Motivação

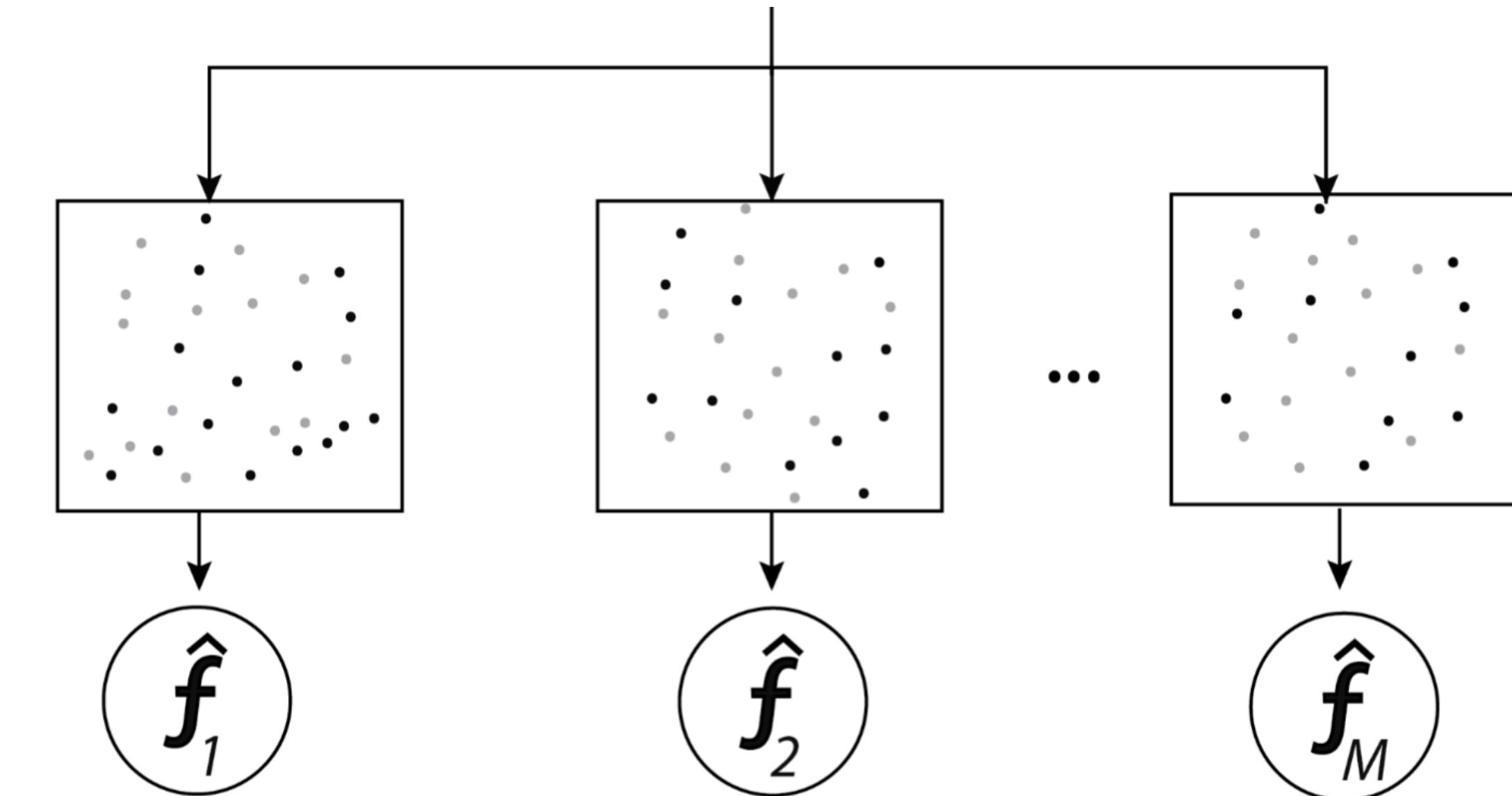
Bagging + Máquinas de Vetores de Suporte

- A utilização do **bagging** acrescentar ganhos substanciais na melhora da capacidade preditiva de um modelo (Breiman, 1996).

Pontos-chave:

Força
E

Diversidade
(Breiman, 2001)



“Demonstrating the stability of support vector machines for classification”
(Buciu, et. al, 2006)

Máquinas Aleatórias

Máquinas Aleatórias

A bagged-weighted support vector model with free kernel choice

O algoritmo propõe a realização de um bagging utilizando models de SVM como modelos base através de uma escolha “aleatória” das funções de kernel utilizadas em cada um deles

Máquinas Aleatórias

A bagged-weighted support vector model with free kernel choice

O algoritmo propõe a realização de um bagging utilizando models de SVM como modelos base através de uma escolha “aleatória” das funções de kernel utilizadas em cada um deles.



Não é necessário o *tuning* para escolha da função de kernel.



Introduz instabilidade/diversidade aos modelos SVM base, sem comprometer a força do modelo.

Máquinas Aleatórias

O algoritmo

- 1 Considere um conjunto de treinamento $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ com $\mathbf{x}_i \in \mathbb{R}^p$ e $y_i \in \{-1, 1\}$, $\forall i = 1, \dots, N$.
- 2 Considere um conjunto de validação $\{(\mathbf{x}_k, y_k)\}_{k=1}^L$ tal que $k = 1, \dots, L$.
- 3 O método é inicializando treinando modelos $h_r(\mathbf{x})$, tal que $r = 1, \dots, R$; em que R é o número total de funções de kernel que serão utilizadas no modelo de máquinas aleatórias.

Máquinas Aleatórias

O algoritmo

- 4 Cada modelo é validado sobre a amostra de validação, e então um vetor de acurácia $\mathbf{ACC} \in \mathbb{R}^R$ é obtido, em que cada coordenada corresponde à uma função de kernel.
- 5 Então, as probabilidades são calculadas para cada função do kernel λ_r é calculada a partir da equação

$$\lambda_r = \frac{\ln\left(\frac{ACC_r}{1-ACC_r}\right)}{\sum_{i=1}^R \ln\left(\frac{ACC_i}{1-ACC_i}\right)},$$

- 6 Seleciona B amostras bootstrap.
- 7 Treina B modelos $g_k, \forall k = 1, \dots, B$; em que a função de kernel que será utilizada em cada modelo é determinada aleatoriamente utilizando o vetor de probabilidades λ .

Máquinas Aleatórias

O algoritmo

- 8 Em seguida, cada modelo é avaliado utilizando as amostras Out of Bagging, calculando-se um vetor de acurárias $\Omega \in \mathbb{R}^B$.
- 9 Um peso então é associado à cada modelo, dado pela equação

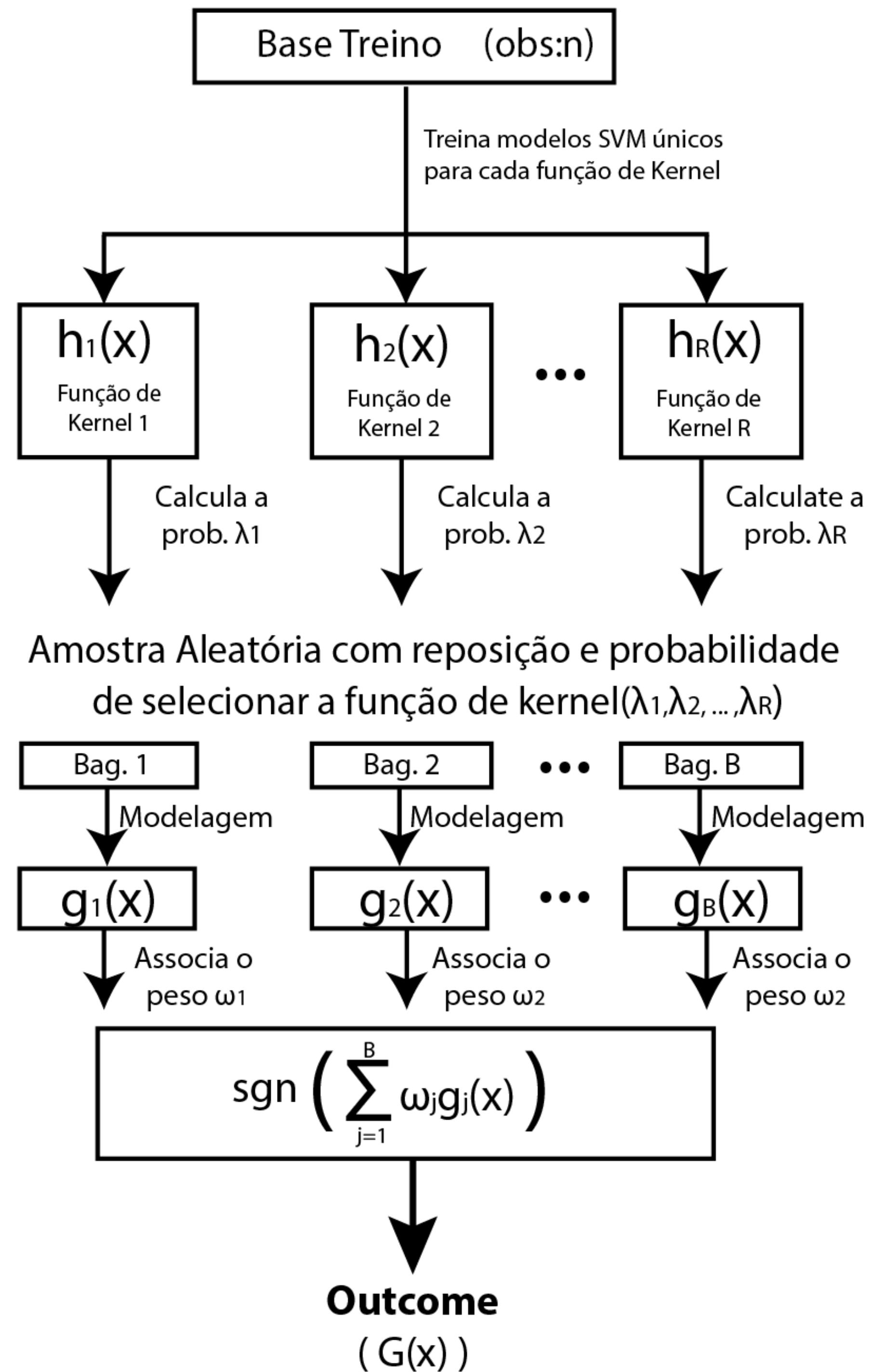
$$w_i = \frac{1}{(1 - \Omega_i)^2}, \quad i = 1, \dots, B,$$

- 10 A classificação final é dado pela equação

$$G(\mathbf{x}_i) = sgn \left(\sum_j^B w_j g_j(\mathbf{x}_i) \right), \quad i = 1, \dots, N.$$

Máquinas Aleatórias

O algoritmo

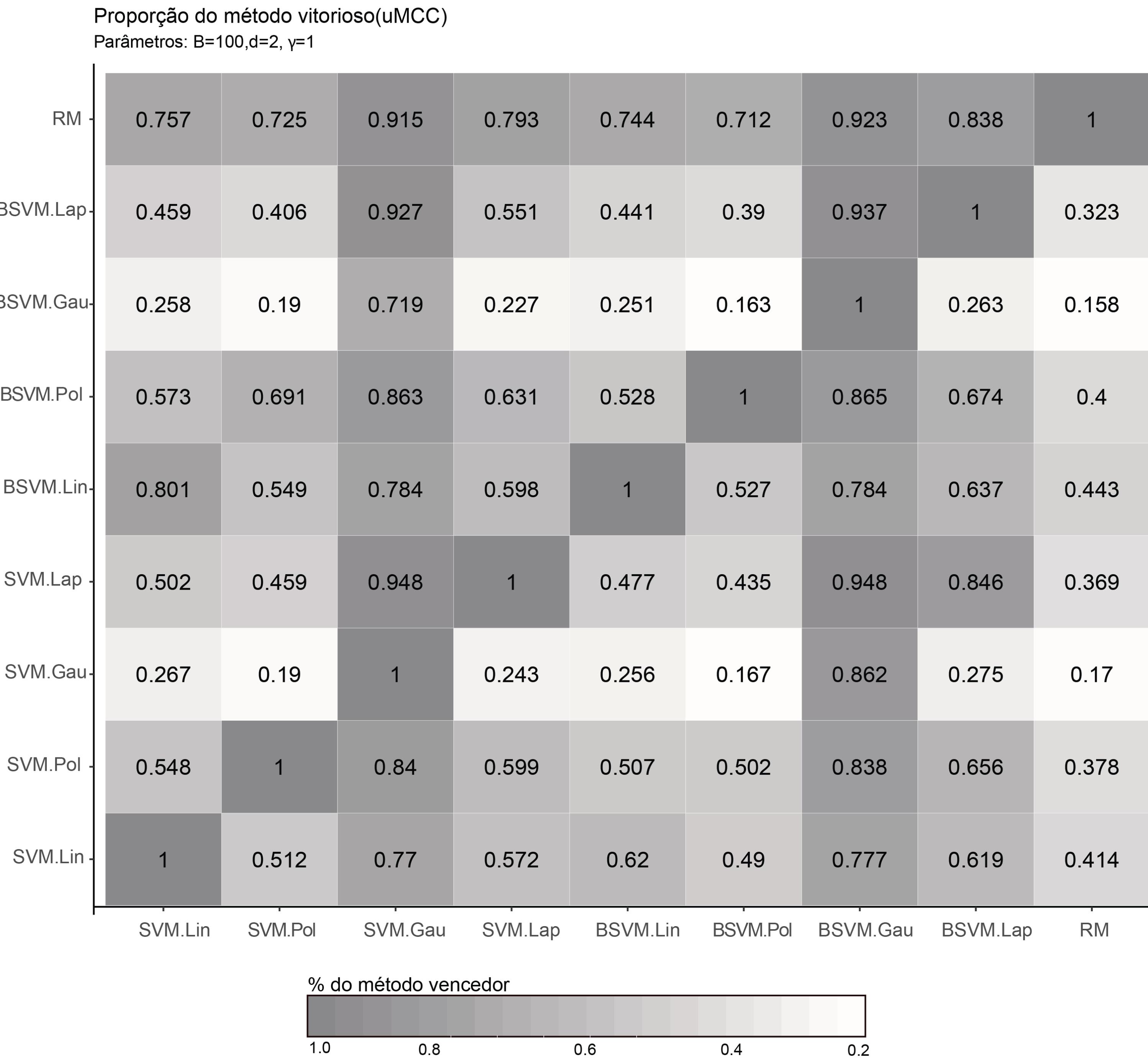


Máquinas Aleatórias: Simulações e Benchmarking

Máquinas Aleatórias

Benchmarking

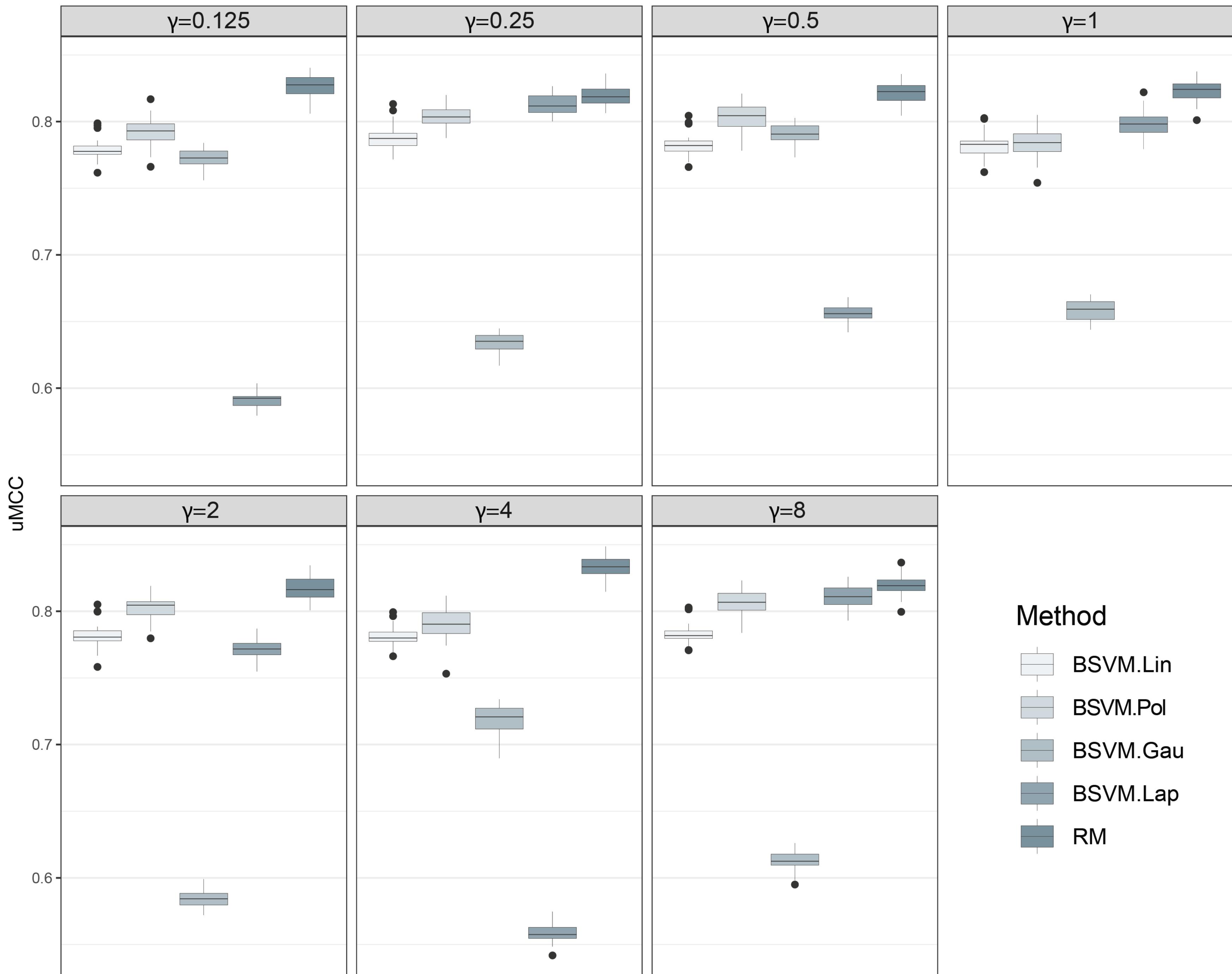
- Performance comparada em bases em 27 benchmarkings retirados, em sua maioria, do UCI Machine Learning Repository



Máquinas Aleatórias

Benchmarking

- Performance comparada em bases em 27 benchmarkings retirados, em sua maioria, do UCI Machine Learning Repository



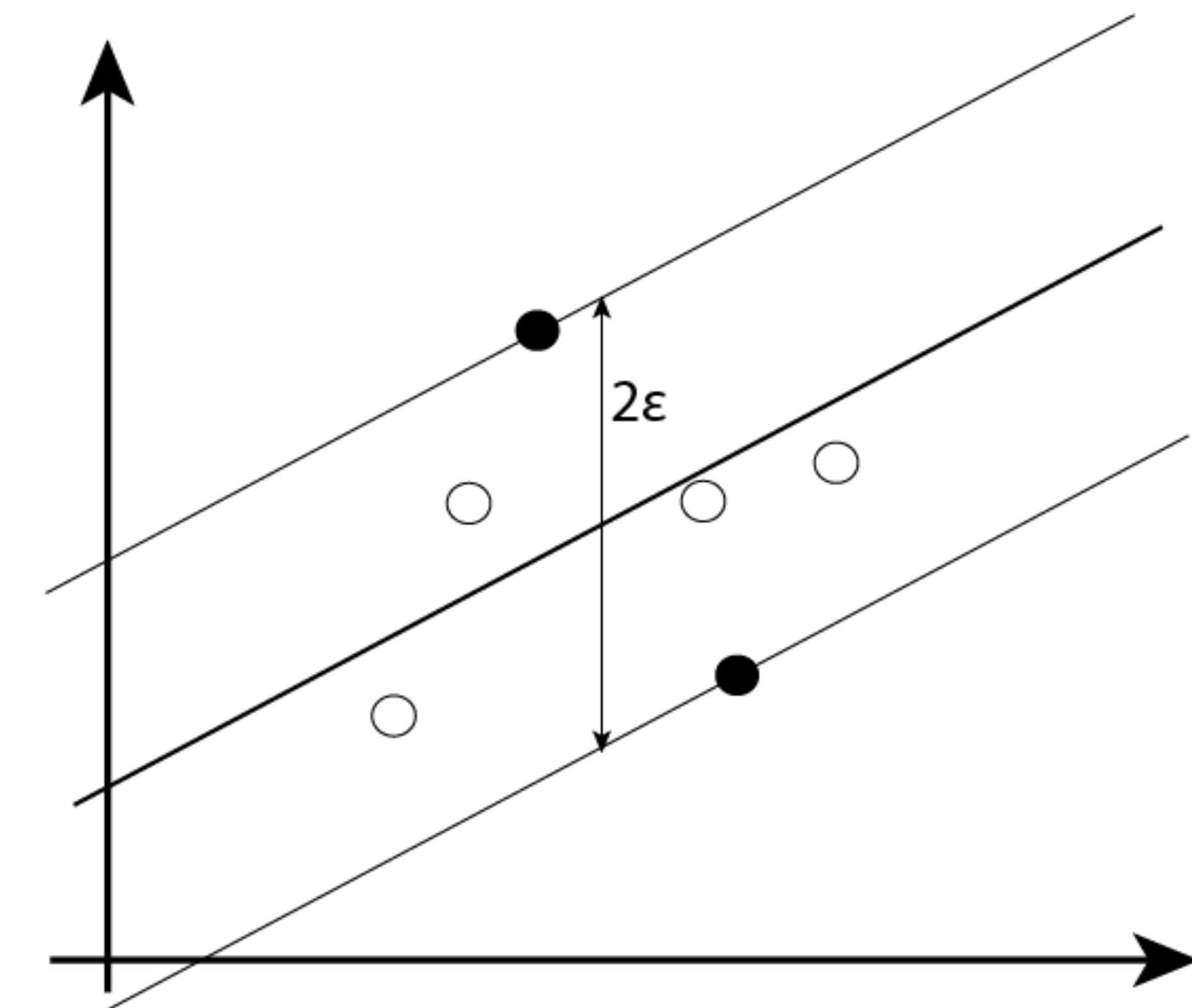
Máquinas Aleatórias:

Uma abordagem para regressão

Máquinas Aleatórias de Regressão

Vetores de Suporte de Regressão

- A utilização de modelos de vetores de suporte regressivo (Drucker et al., 1997) como modelos base, além de modificações nas ponderações e cálculo de probabilidade são principais diferenciais nesta nova abordagem.



Máquinas Aleatórias de Regressão

Principais diferenças (em relação ao RM)

Paramêtro de Probabilidade

$$\lambda_r = \frac{e^{-\beta\delta_r}}{\sum_{i=1}^R e^{-\beta\delta_i}},$$

δ_r : raiz do erro quadrático médio padronizado

β : coeficiente de correlação

Coeficiente de Peso dos Modelos Base

$$w_i = \frac{e^{-\beta\Lambda_i}}{\sum_{j=1}^B e^{-\beta\Lambda_j}} \quad i = 1, \dots, B$$

Λ : RMSE sobre as amostras Out-of-Bagging

β : coeficiente de correlação

Predição Final

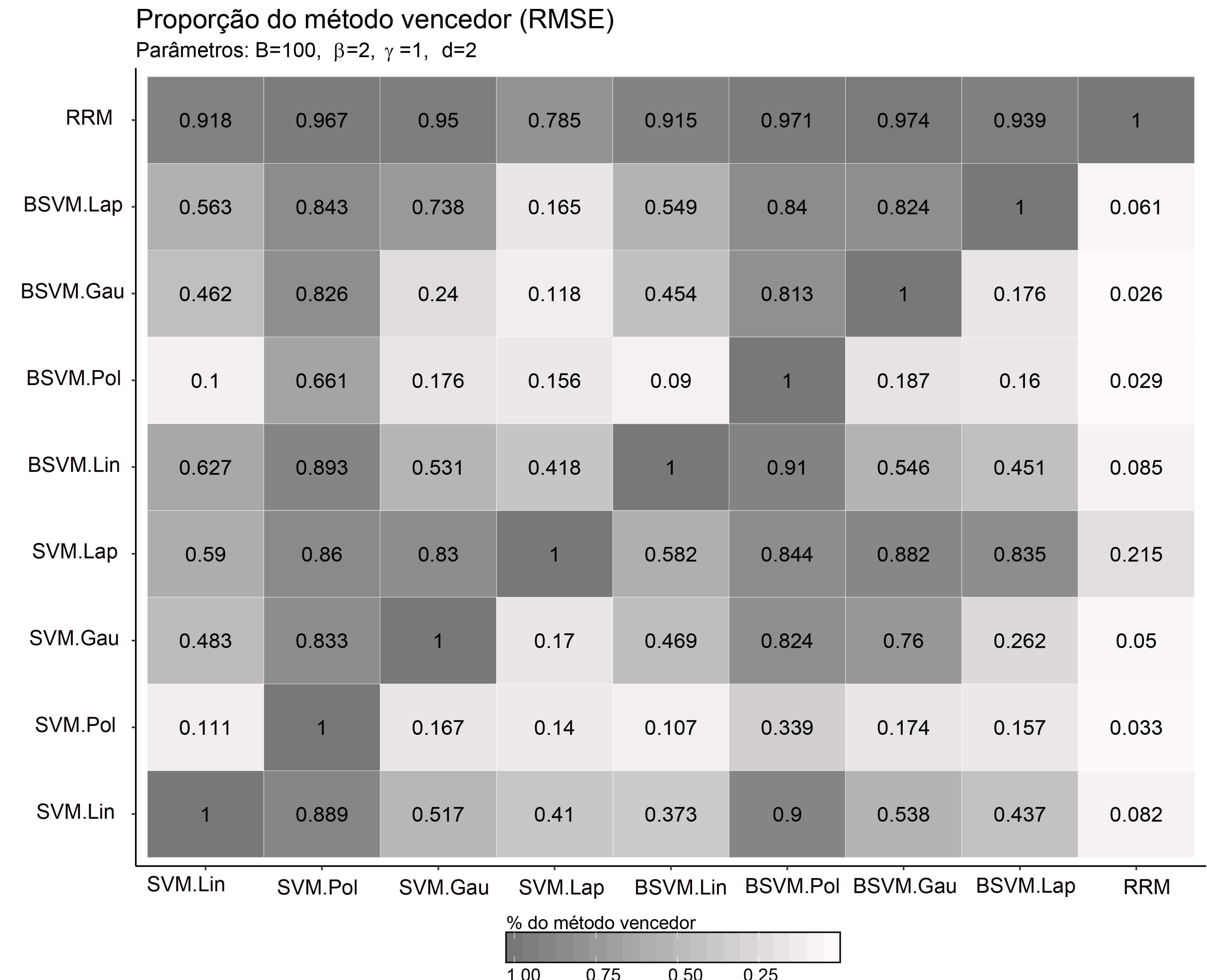
$$G(\mathbf{x}_i) = \sum_j^B w_j g_j(\mathbf{x}_i) \quad i = 1, \dots, N.$$

Máquinas Aleatórias de Regressão: Simulações e Benchmarking

Máquinas Aleatórias de Regressão

Benchmarking

- Performance comparada em bases em 26 benchmarks retirados, em sua maioria, do UCI Machine Learning Repository



Máquinas Aleatórias de Regressão Benchmarking

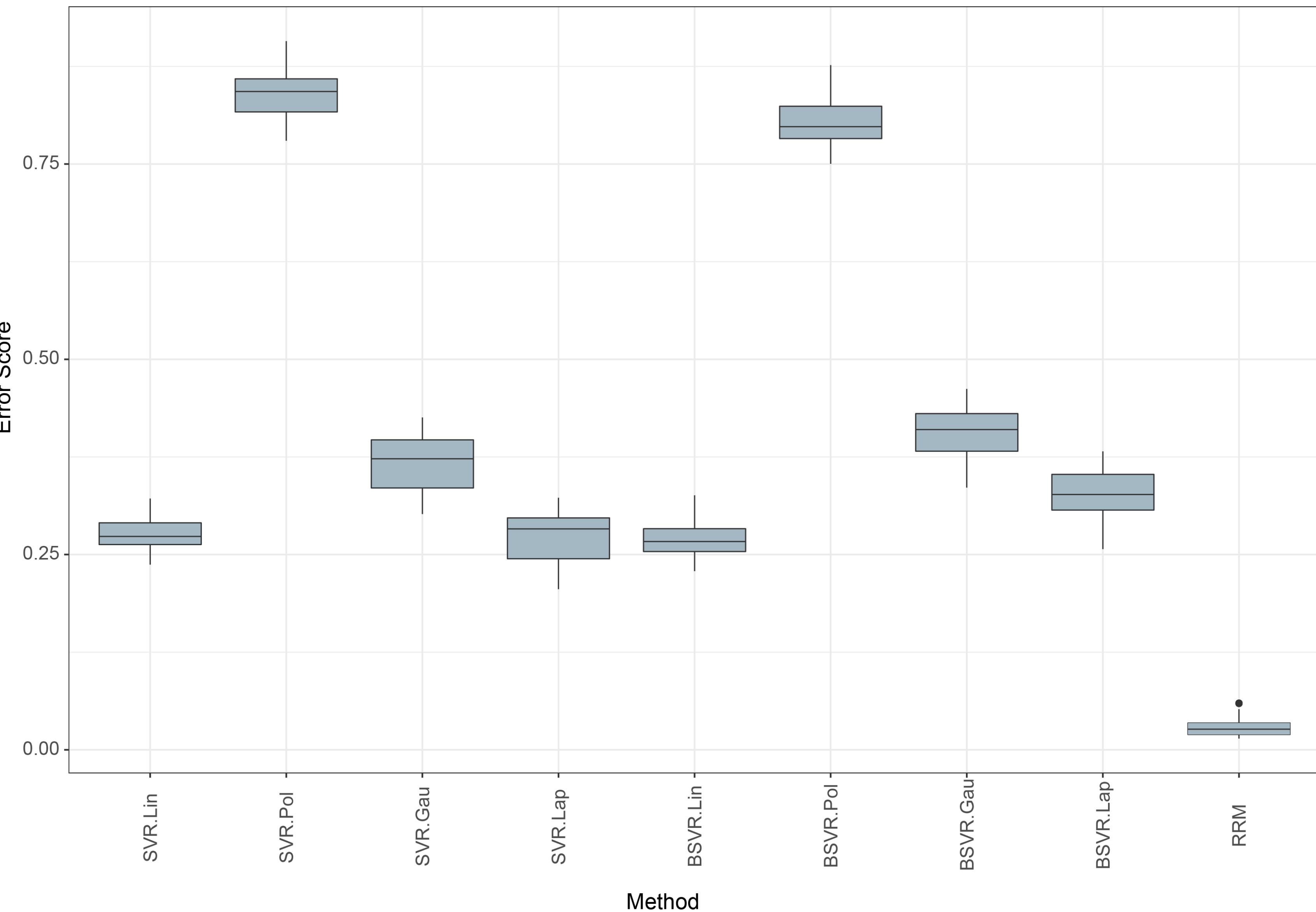
Error Score

$$ES_i = \frac{\varepsilon_i - \varepsilon_{min}}{\varepsilon_{max} - \varepsilon_{min}}$$

ε_i : RMSE referente ao método i

Error Score (ES) médio para todas bases reais

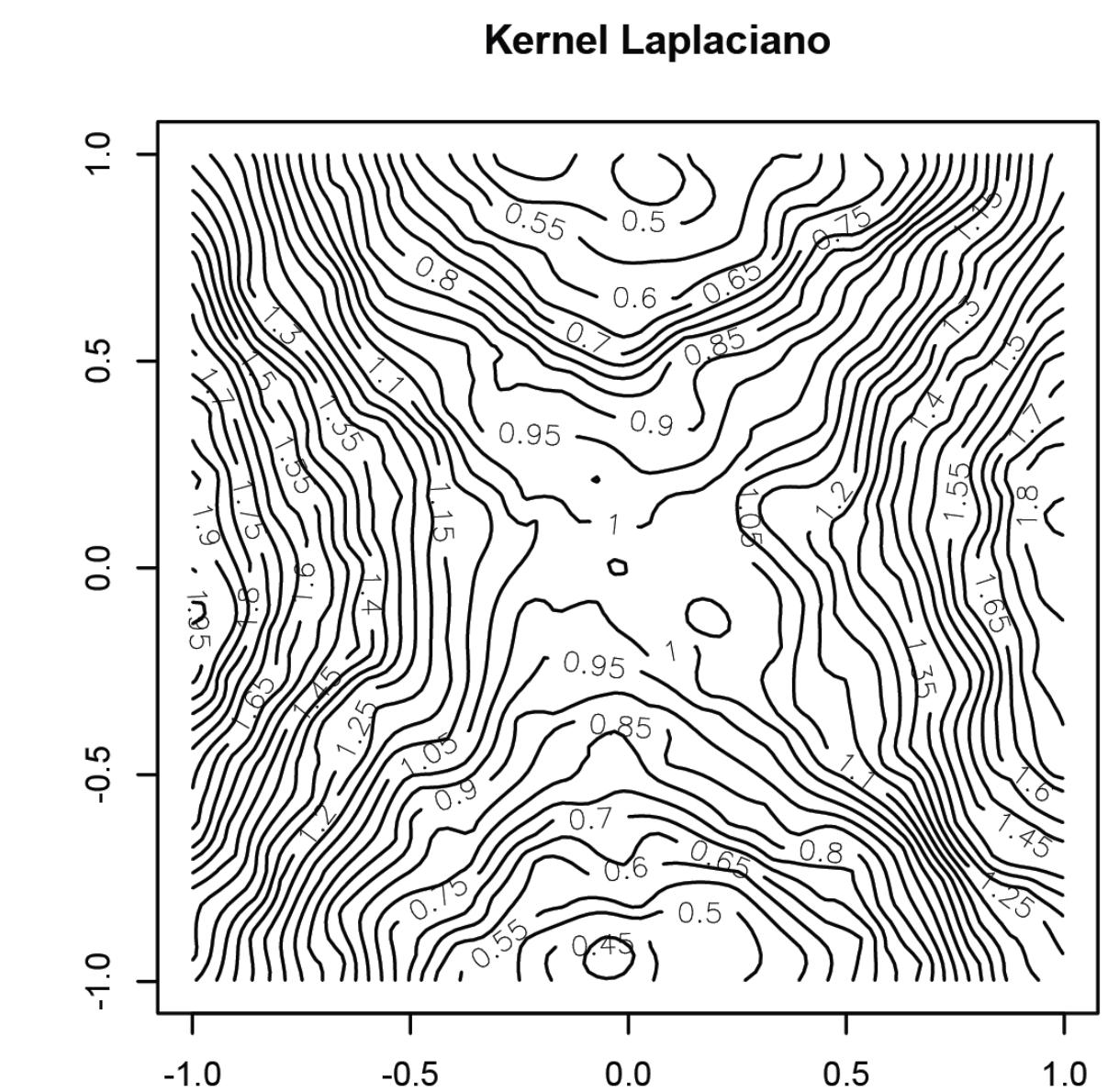
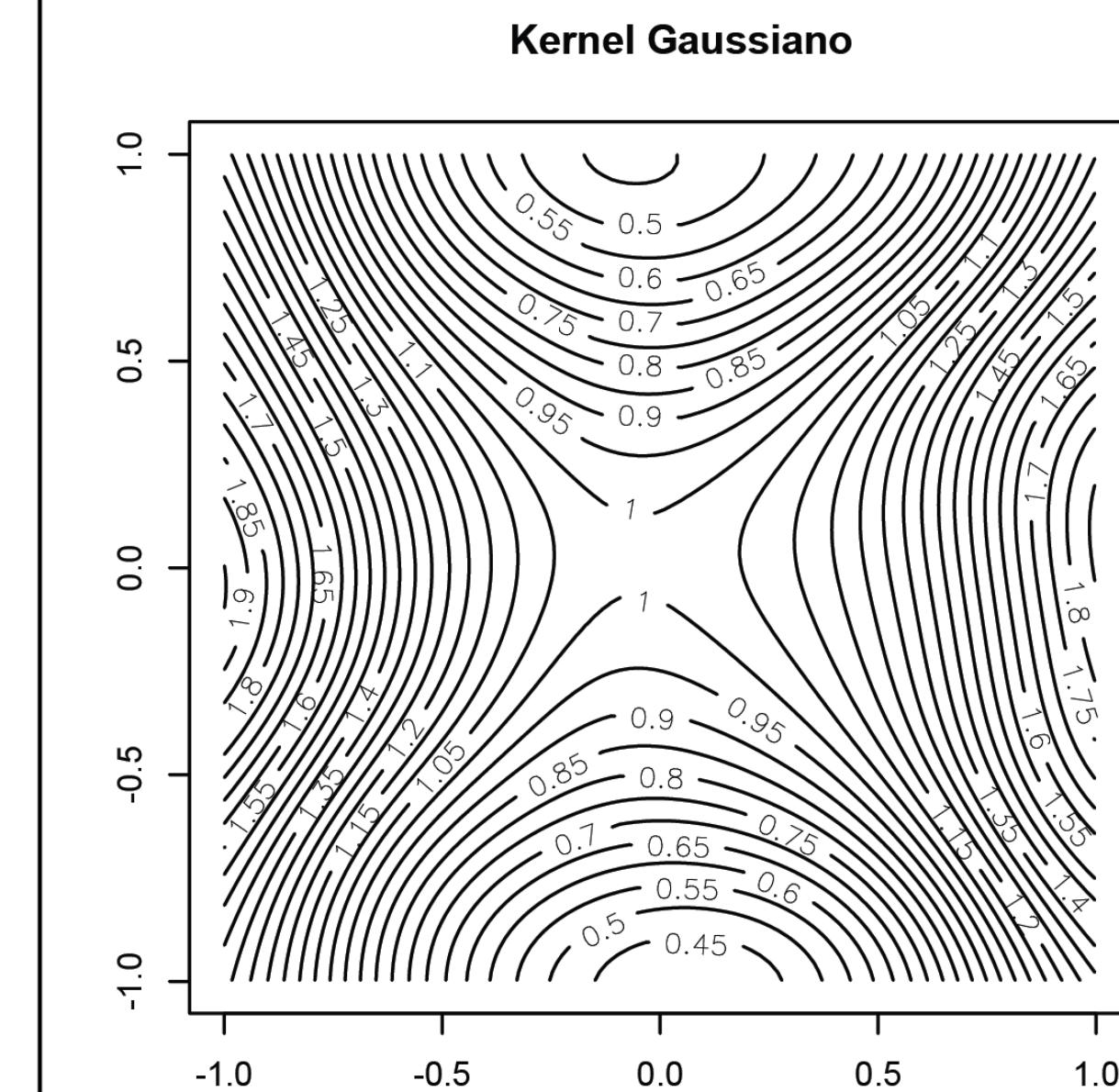
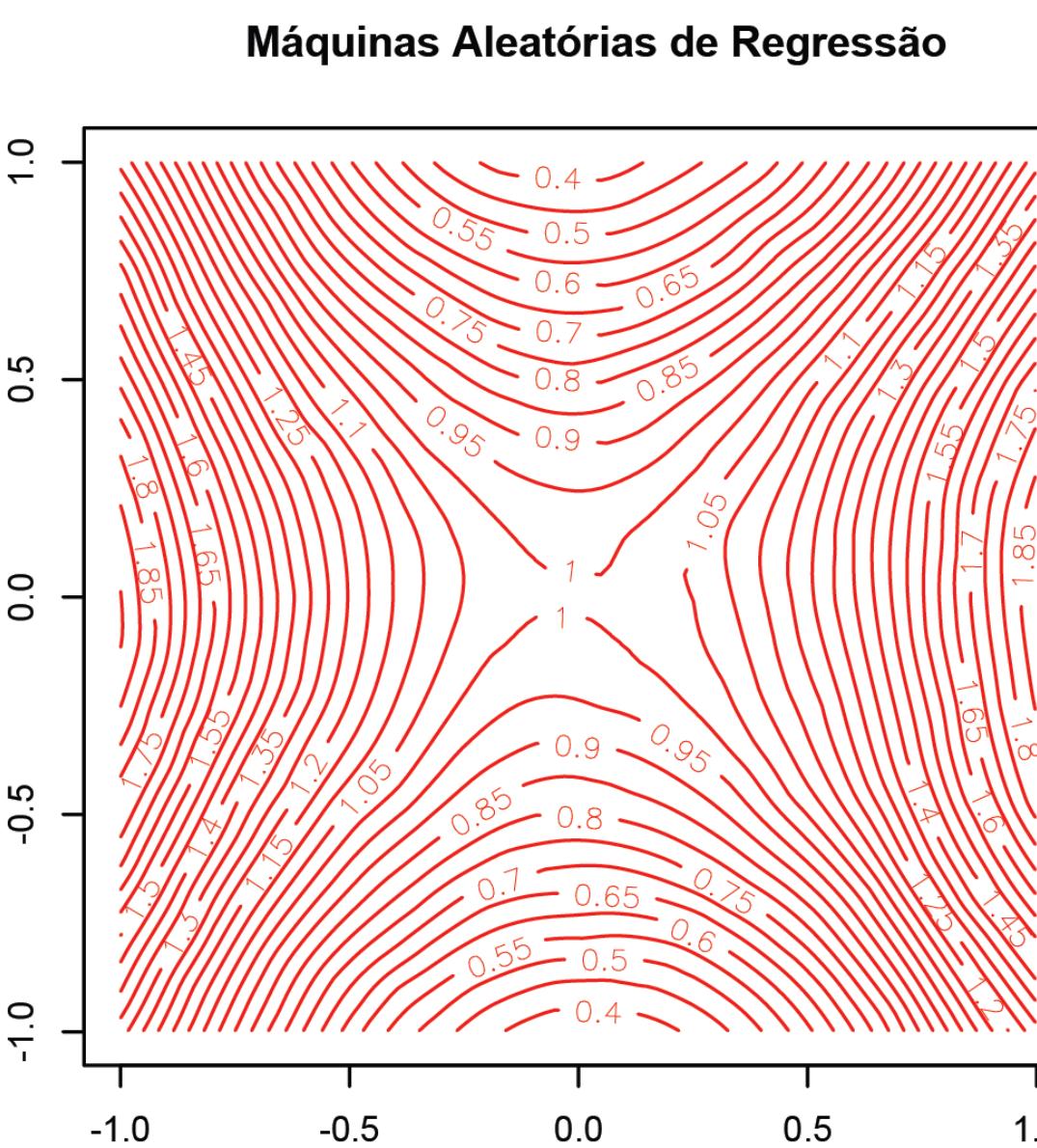
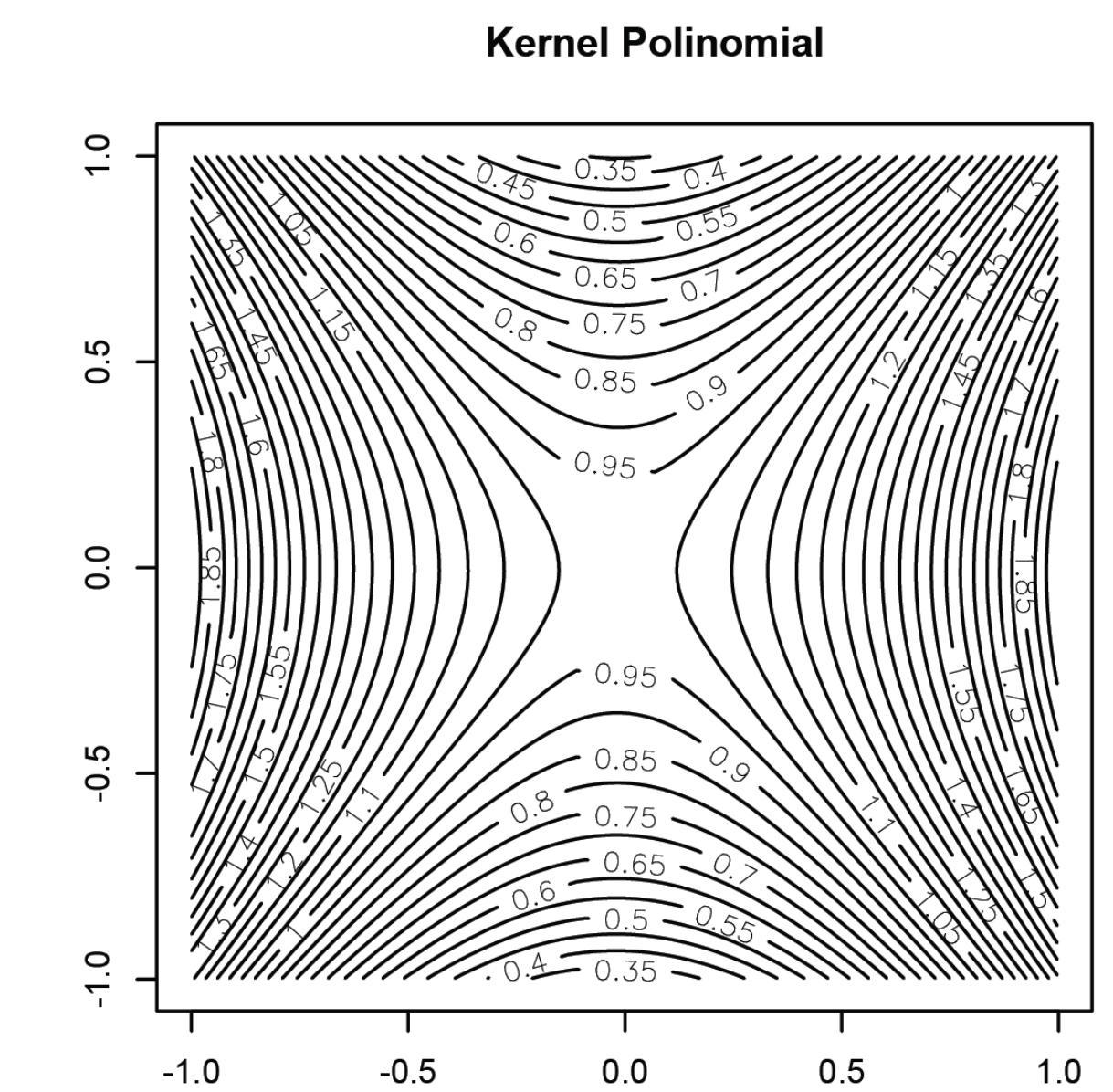
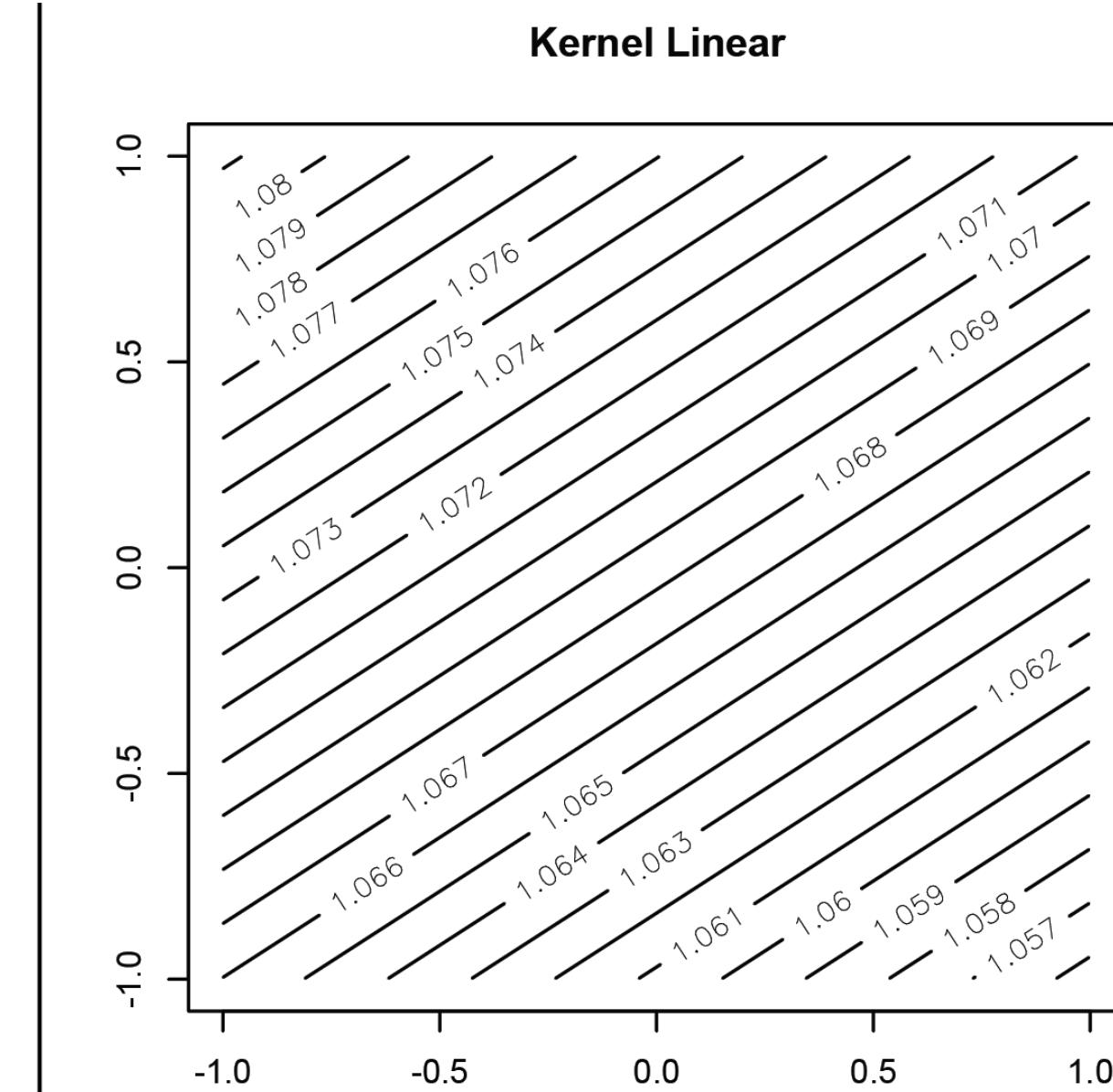
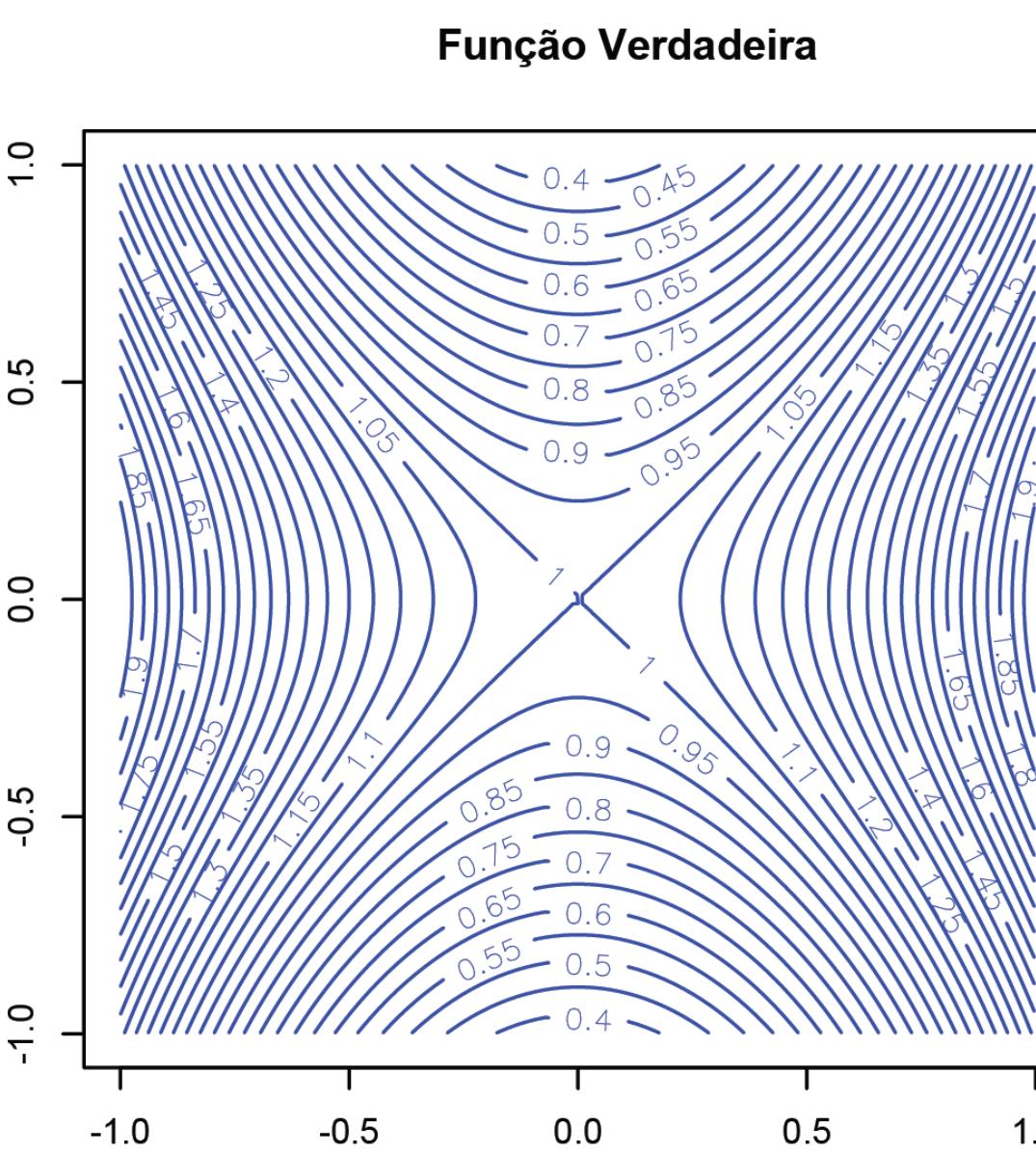
Paramêtros: B=100, $\beta = 2$, $\gamma = 1$, $d=2$



Máquinas Aleatórias de Regressão: Performance e Concordância

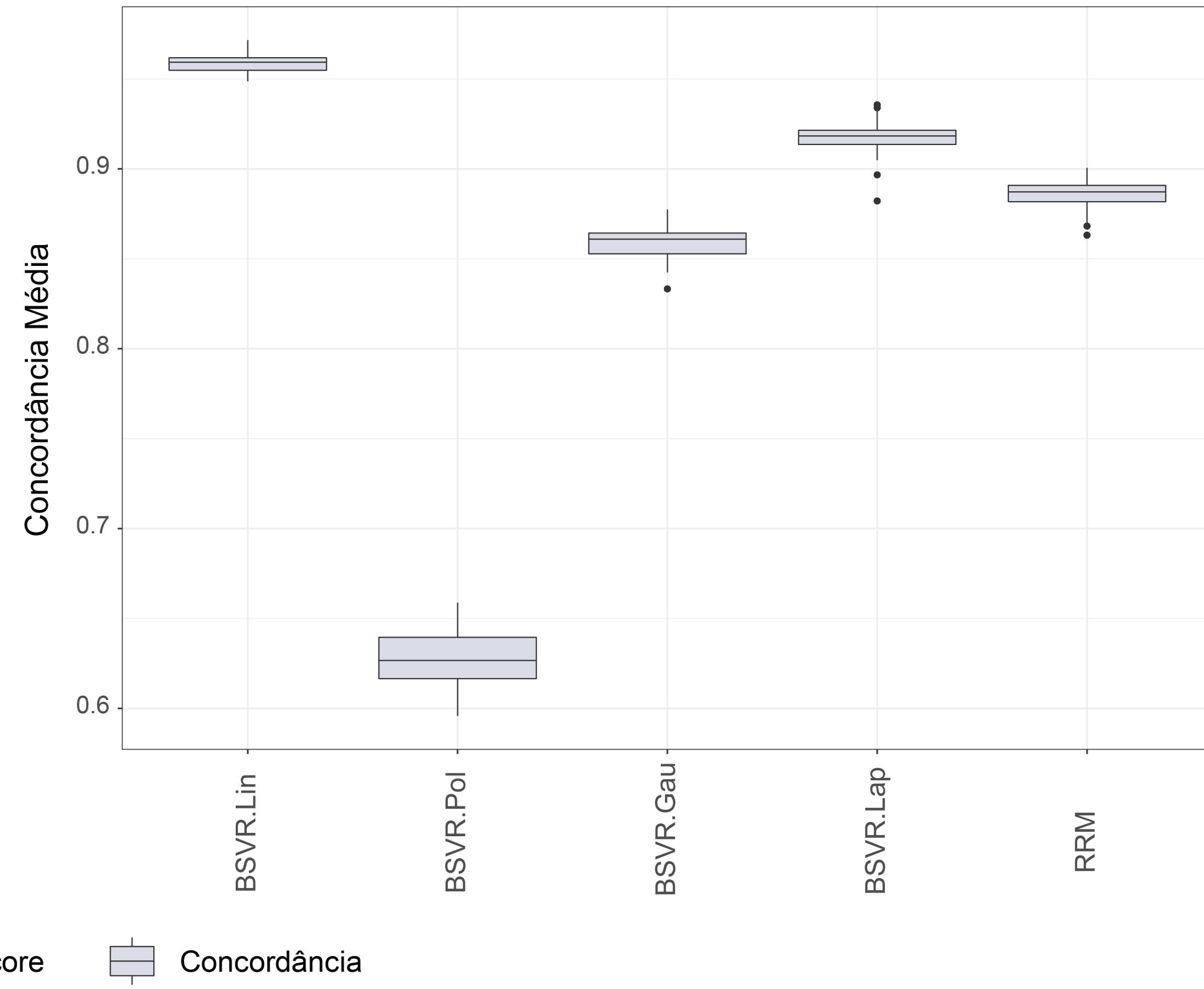
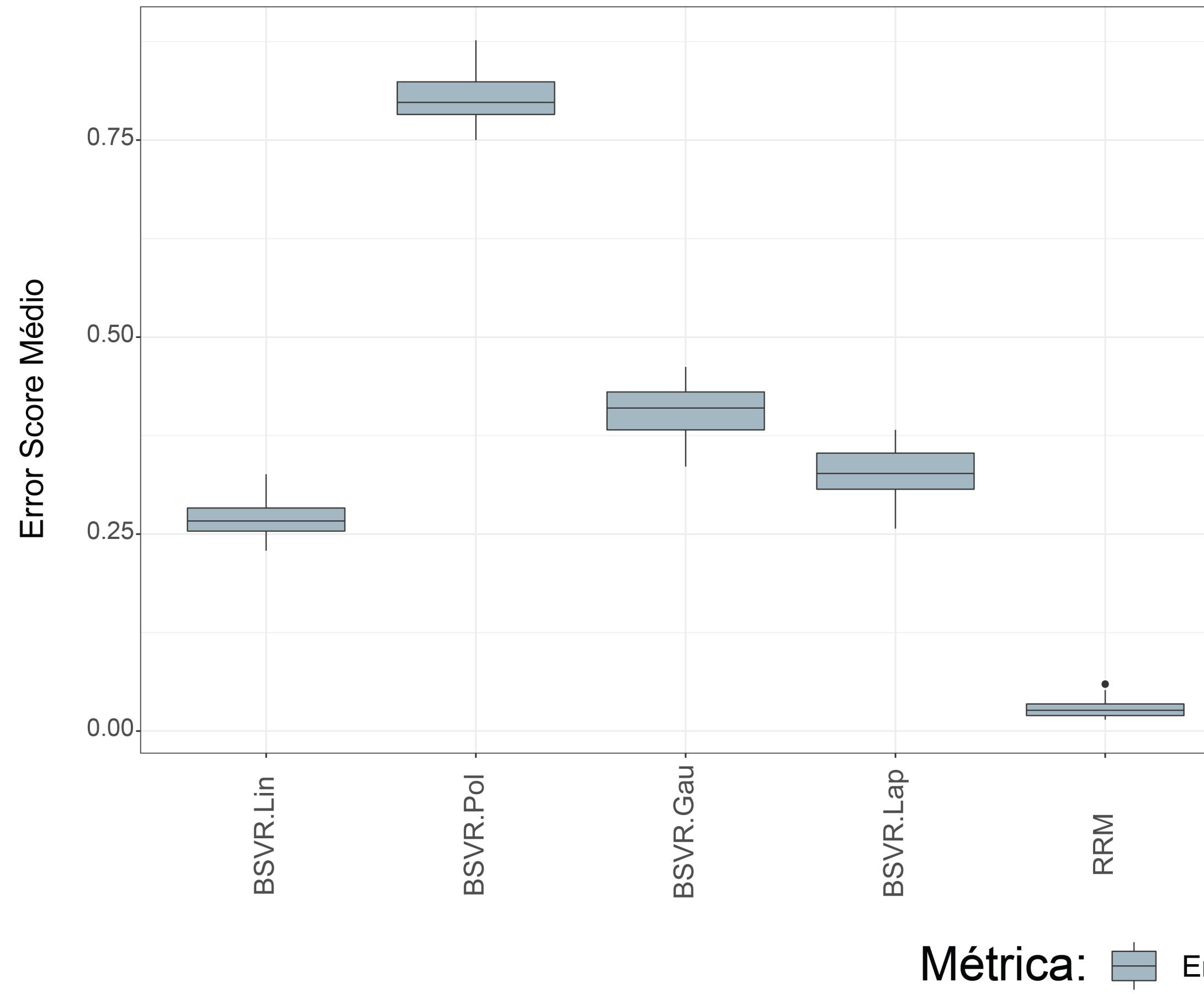
Máquinas Aleatórias de Regressão

Performance e Concordância



Máquinas Aleatórias de Regressão

Performance e concordância



Uma comparação empírica entre Máquinas Aleatórias e Florestas Aleatórias

Máquinas Aleatórias X Florestas Aleatórias

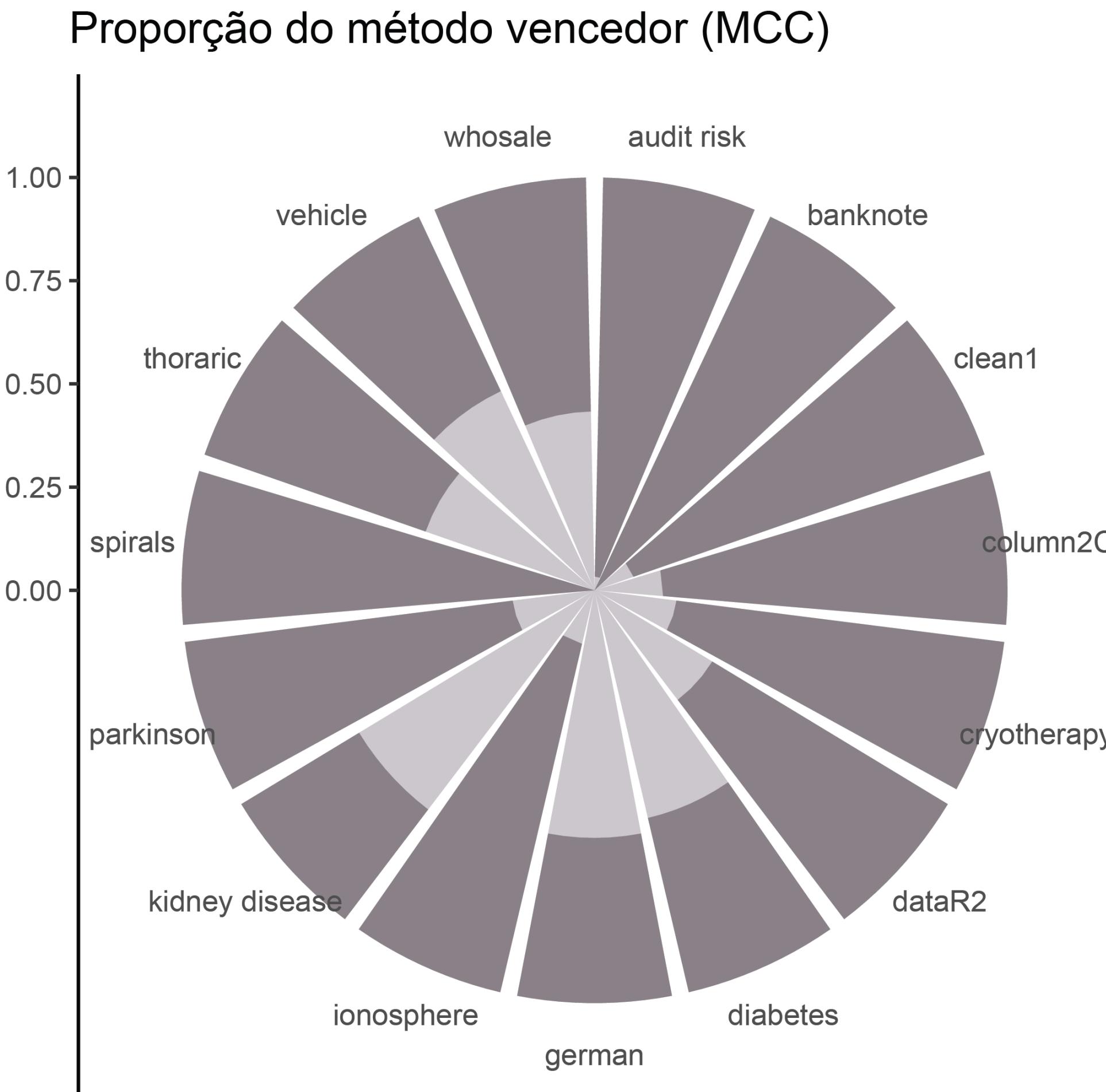
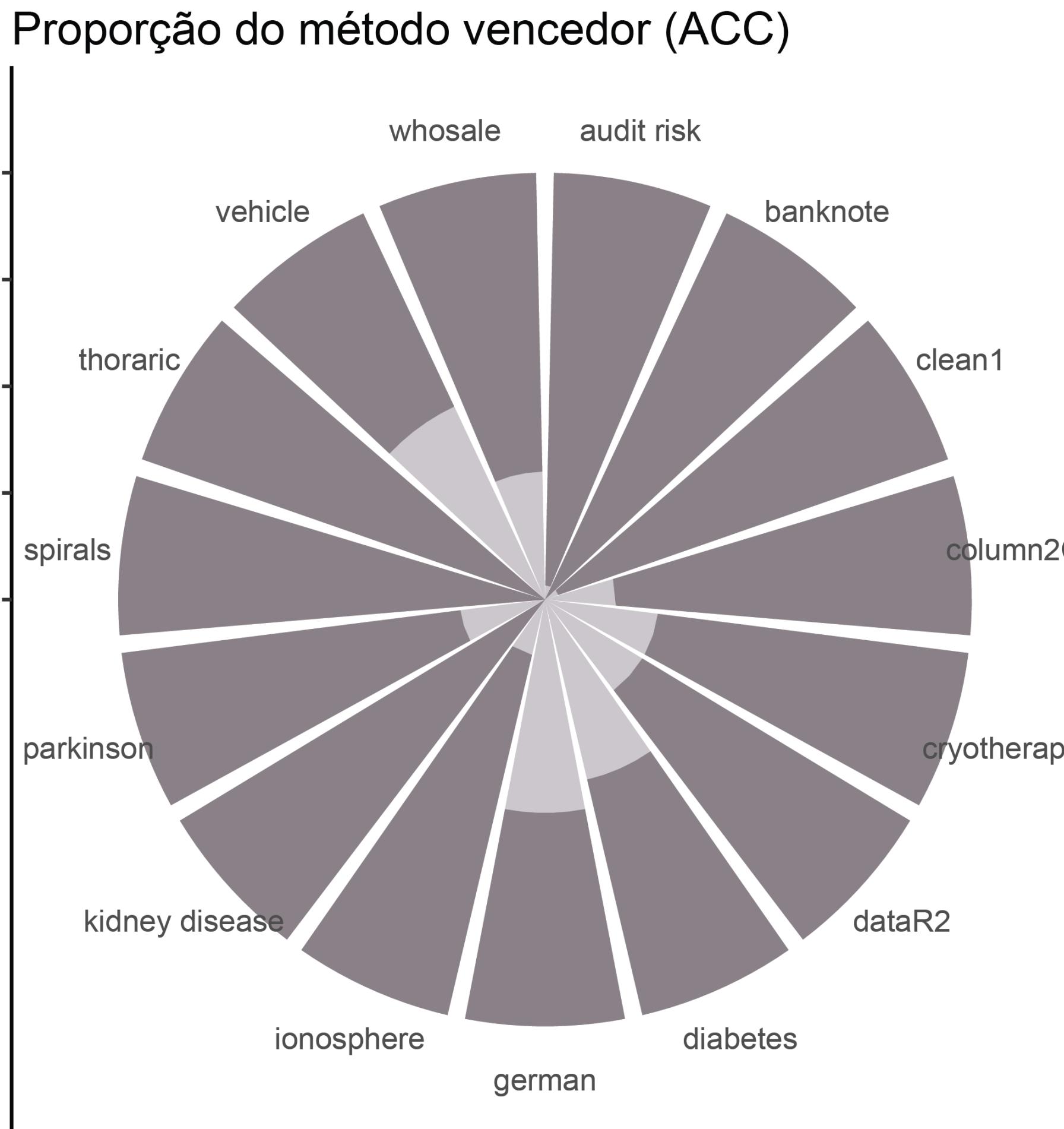
- A performance das Florestas Aleatórias (RF) é notável. Fernando Delgado *et al.* (2017) fizeram uma extensiva comparação utilizando um família de 17 classificadores e 179 bases de dados, e chegou à conclusão de que as Árvores Aleatórias, foi considerando o método mais forte, atingindo 94% da precisão máxima, superando 90% nos 84% dos conjuntos de dados.
- A proposta, então, é de comparar as Máquinas Aleatórias com este método de ensemble consolidado, através de diversas bases simuladas e reais.

Highlight

- RM apresentou performance melhor, ou competitiva quando comparado com o RF em tarefas de classificação e regressão
- Foram utilizadas bases de simulações e benchmarking
- Foram utilizadas três “novas bases” de dados reais.

Máquinas Aleatórias X Florestas Aleatórias

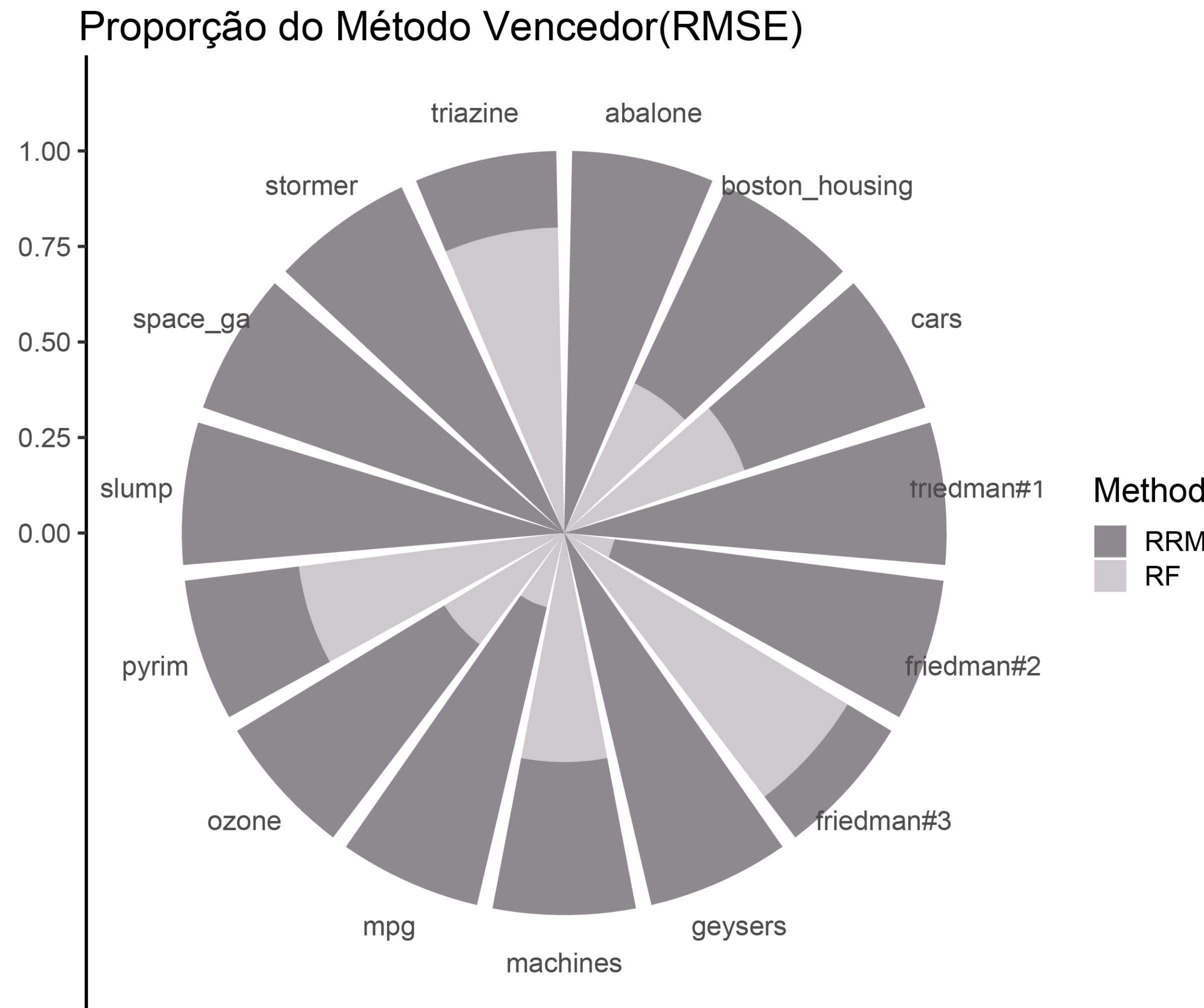
Classificação: Benchmarking



Método: RM RF

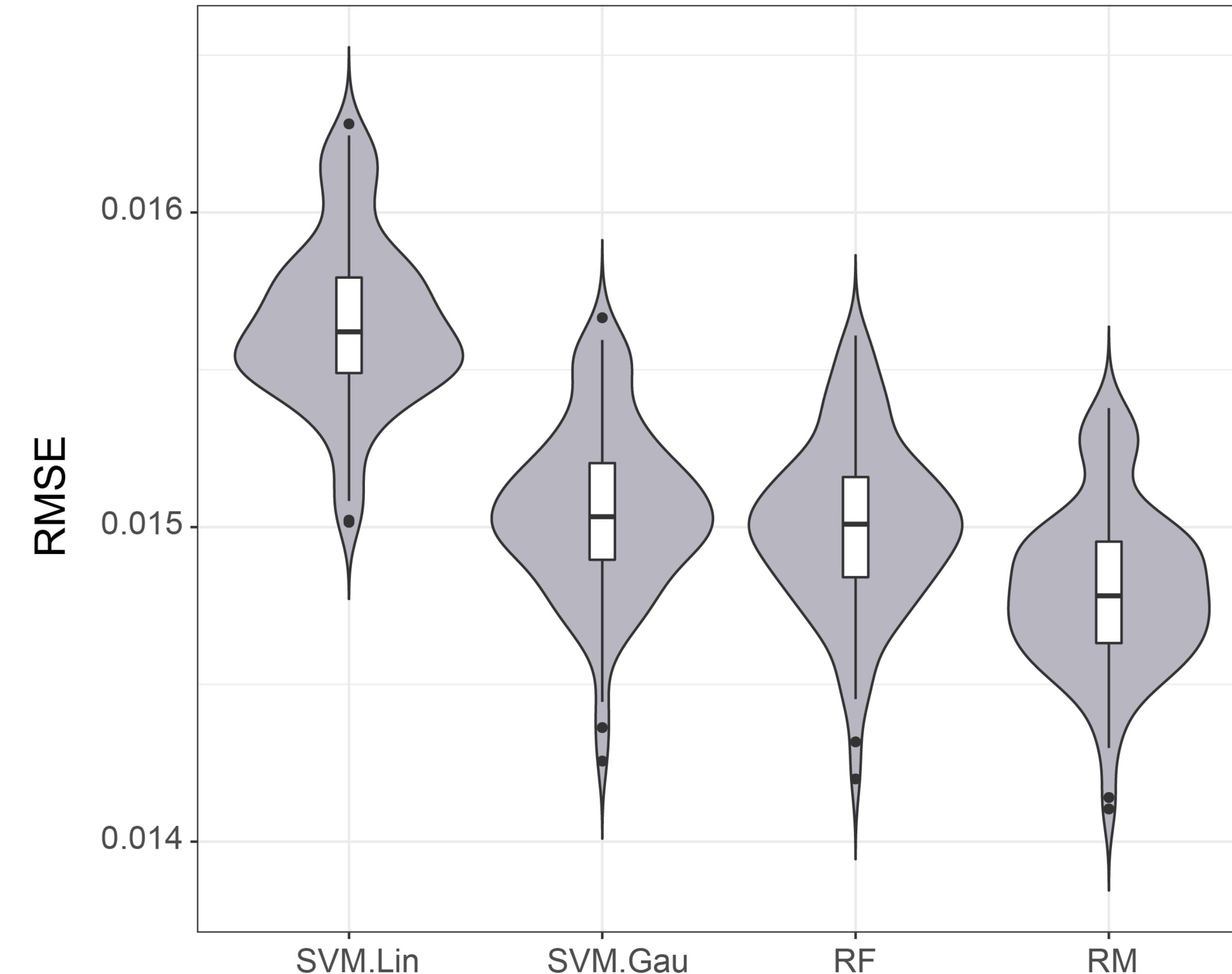
Máquinas Aleatórias de Regressão

Regressão: Benchmarking



Máquinas Aleatórias de Regressão

Novas Bases de dados Reais: Bolsa Família



Resultados da Pesquisa

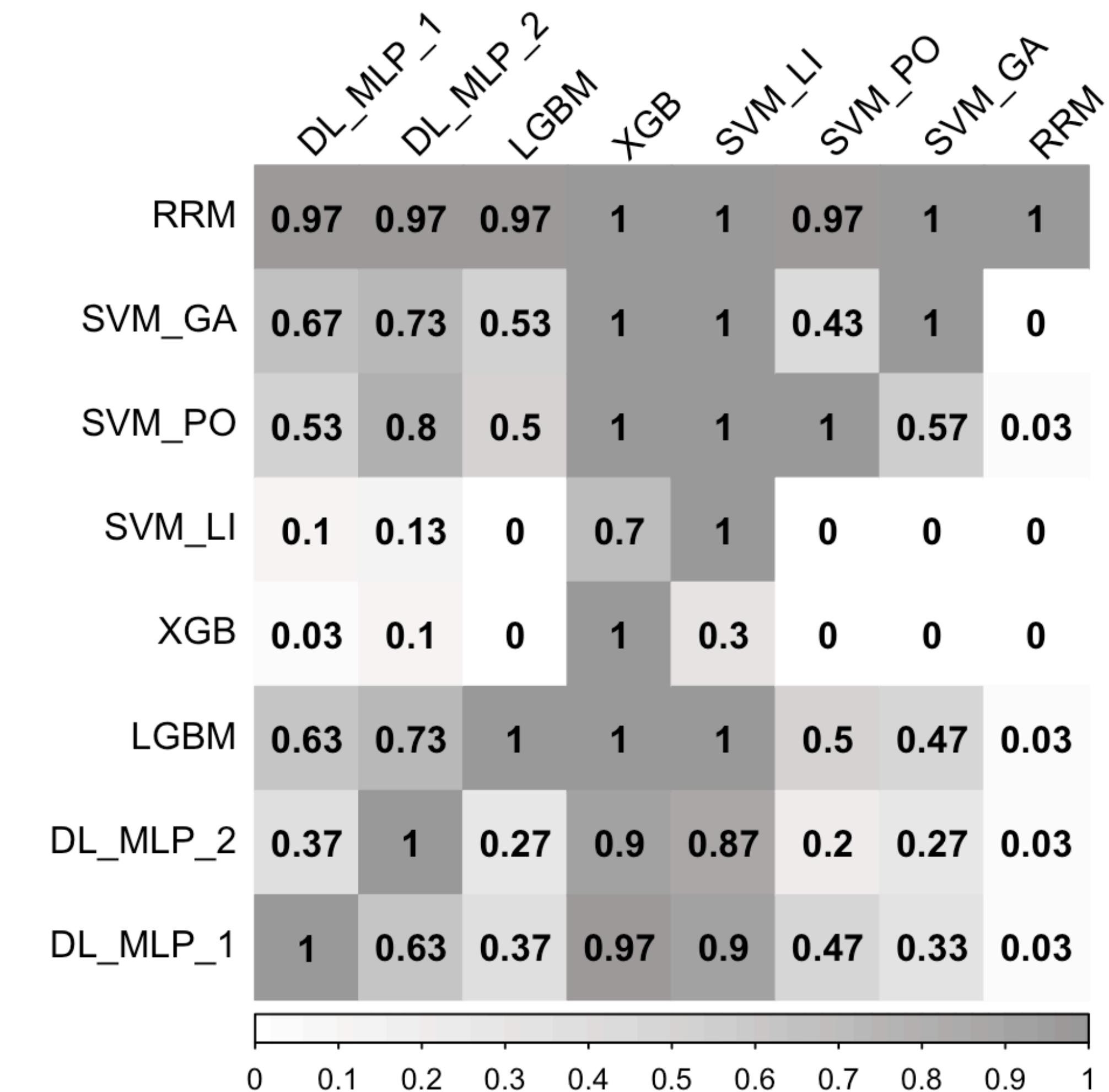
Resultados da Pesquisa

Artigos publicados

1. Maia, M., Pimentel, J. S., Pereira, I. S., Gondim, J., Barreto, M. E., & Ara, A. (2020). Convolutional support vector models: prediction of coronavirus disease using chest X-rays. *Information*, 11(12), 548.
2. Ara, Anderson, et al. "Random machines: A bagged-weighted support vector model with free kernel choice." *Journal of Data Science* 19.3 (2021): 409-428.
3. Ara, Anderson, et al. "Regression random machines: An ensemble support vector regression model with free kernel choice." *Expert Systems with Applications* 202 (2022): 117107.
4. Maia, Mateus, Arthur R. Azevedo, and Anderson Ara. "Predictive Comparison Between Random Machines and Random Forests." *Journal of Data Science* 19.4 (2021).

Resultados da Pesquisa

"Regression random machines: An ensemble support vector regression model with free kernel choice."



Resultados da Pesquisa

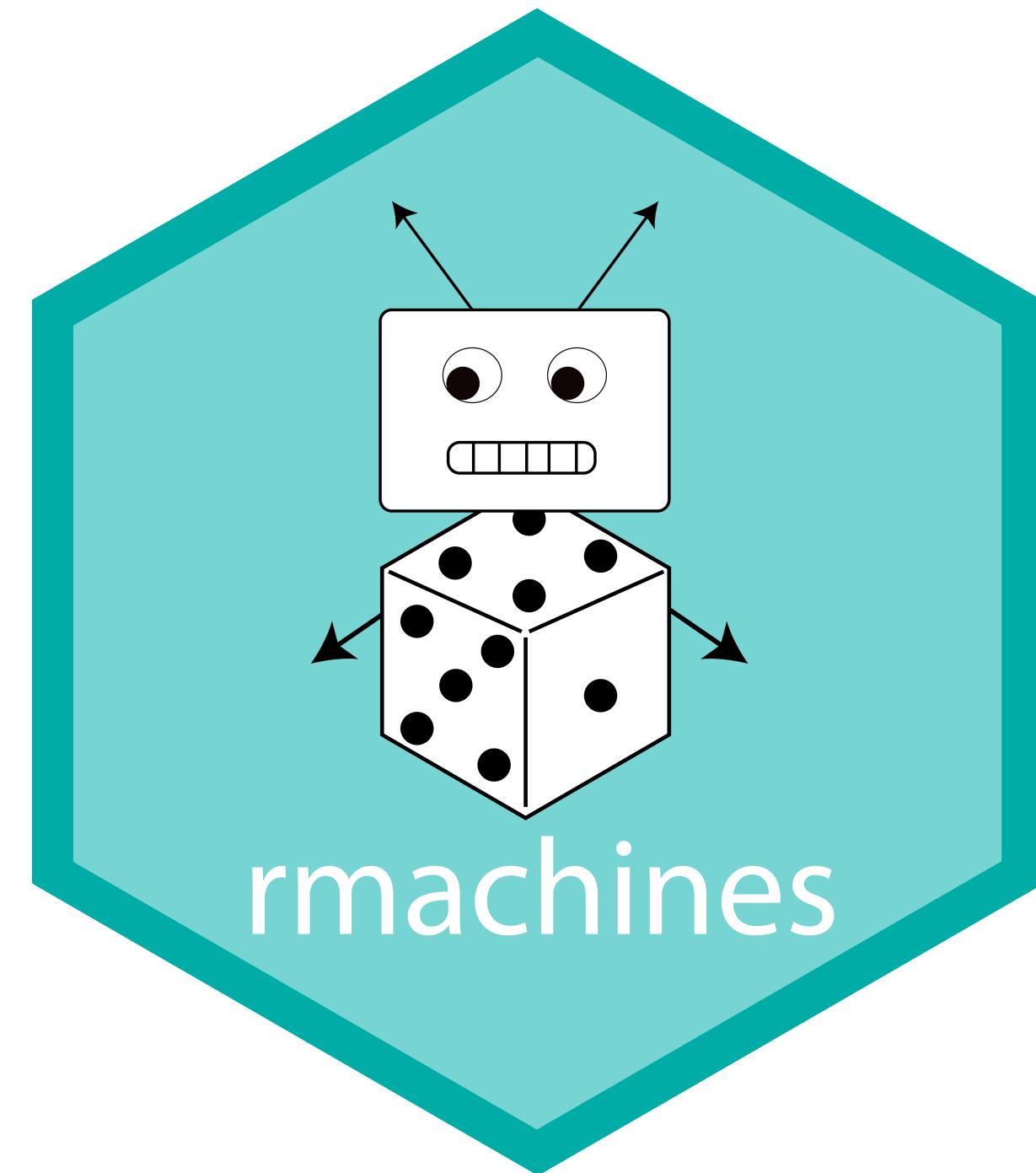
Implementação em R

Installation

To install the **Random Machines** package `rmachines` in the development version from GitHub, just use the command:

```
# install.packages("devtools")
devtools::install_github("MateusMaiaDS/rmachines")
```

Link: <https://github.com/MateusMaiaDS/rmachines>



Resultados da Pesquisa

Artigos em construção:

1. **Tutorial on Ensemble methods:** revisão bibliográfica dos métodos de ensemble o estado-da-arte dessa abordagem de modelagem.
2. **Parallel Random Machines:** paralelização do algoritmo para maiores optimizar o custo computacional.
3. **Extensões ao modelo:** variação das funções de kernel, importância de variáveis, outras aplicações.

Obrigado.

Agradecimentos.



*Conselho Nacional de Desenvolvimento
Científico e Tecnológico*



Mateus Maia's work was supported by Science Foundation Ireland Career Development Award grant number 17/CDA/4695 and SFI research centre award 12/RC/2289P2.

Referências

1. Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." *Machine learning* 20.3 (1995): 273-297.
2. Breiman, Leo. "Bagging predictors." *Machine learning* 24.2 (1996): 123-140.
3. Buciu, Ioan, Constantine Kotropoulos, and Ioannis Pitas. "Demonstrating the stability of support vector machines for classification." *Signal Processing* 86.9 (2006): 2364-2380.
4. Stamp, Mark. *Introduction to machine learning with applications in information security*. Chapman and Hall/CRC, 2017.
5. Fernández-Delgado, Manuel, et al. "Do we need hundreds of classifiers to solve real world classification problems?." *The journal of machine learning research* 15.1 (2014): 3133-3181.