
Avaliação da Interpretabilidade em Regressão Simbólica

— GUILHERME SEIDYO IMAI ALDEIA —

Orientador: Prof. Dr. Fabrício Olivetti de França

02 de dezembro de 2021

Agenda

1. Introdução
2. Interpretabilidade no Aprendizado de Máquina
3. Regressão Simbólica: uma alternativa caixa-cinza
4. Medidas de avaliação de explicações
5. Avaliação da Interpretabilidade em Regressão Simbólica
6. Um gerador de relatórios automático
7. Conclusões e considerações finais

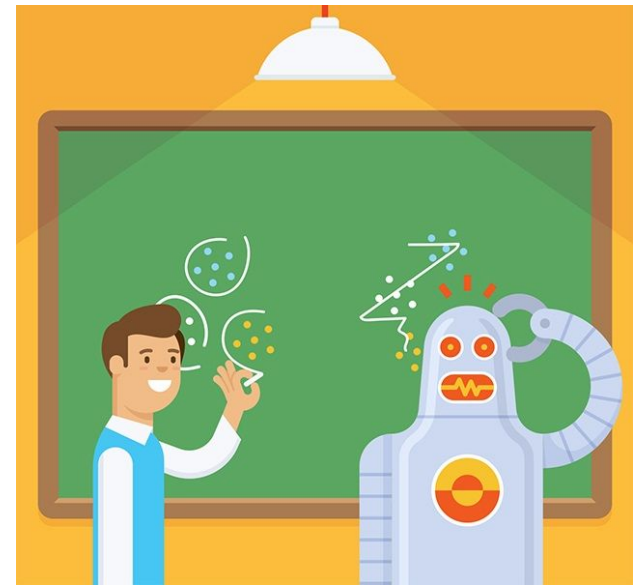
Introdução

- Contextualização
 - Objetivos e justificativa
-

Contextualização

Aprendizado de máquina:

- Aplicação em diversos domínios com um bom desempenho;
- Ajuste de um modelo interno:
 - Guiado por dados;
 - Utilizado para fazer previsões.
- Em alguns casos, pode ser interpretável.



Contextualização

E quando queremos entender um modelo? Interesse em diferentes aspectos:

- Social: Identificar vieses, discriminação, correlações incorretas;
- Científico: Entender o modelo, obter *insights*, descoberta científica.

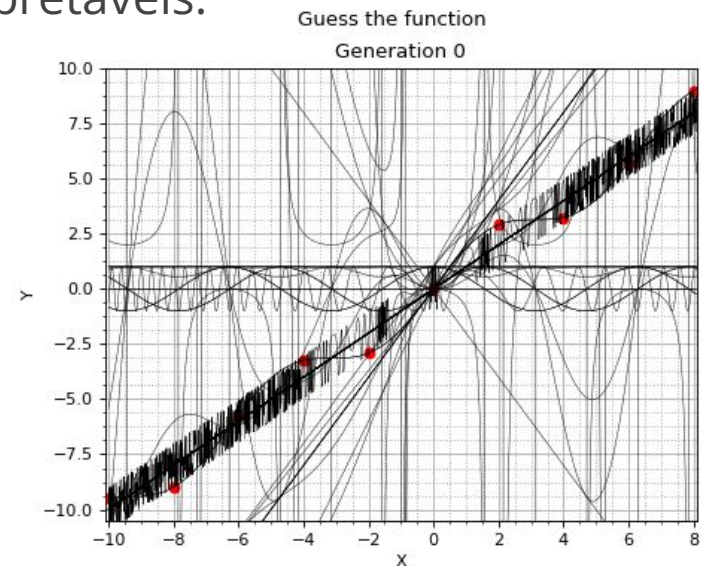
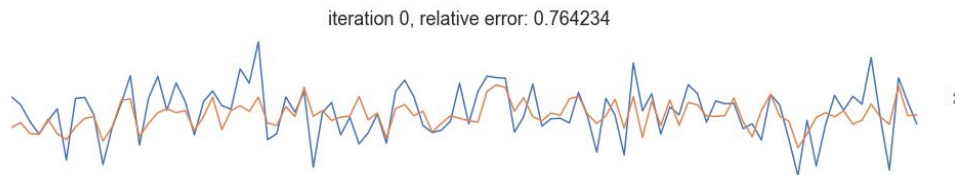
eXplainable Artificial Intelligence (xAI)

Obter explicações do funcionamento dos modelos, através da criação de **modelos mais interpretáveis** ou do uso de **métodos de explicação de modelos**

Contextualização

Regressão simbólica é uma candidata!

- Método que busca por uma expressão matemática para descrever os dados disponíveis;
- Capaz de obter resultados competitivos com métodos de aprendizado de máquina do estado-da-arte;
- Potencial de obter resultados mais interpretáveis.



Objetivos e justificativa

Objetivo principal

Avaliar a **qualidade e robustez de explicações** de importância de variáveis **no contexto da regressão simbólica**, contrastando os resultados obtidos com diferentes modelos utilizados atualmente.

Justificativa

Avaliar a interpretabilidade no contexto da regressão simbólica pode melhorar seus resultados.

Como subproduto, sintetizar os melhores métodos em um gerador de relatórios automático.

Interpretabilidade no Aprendizado de Máquina

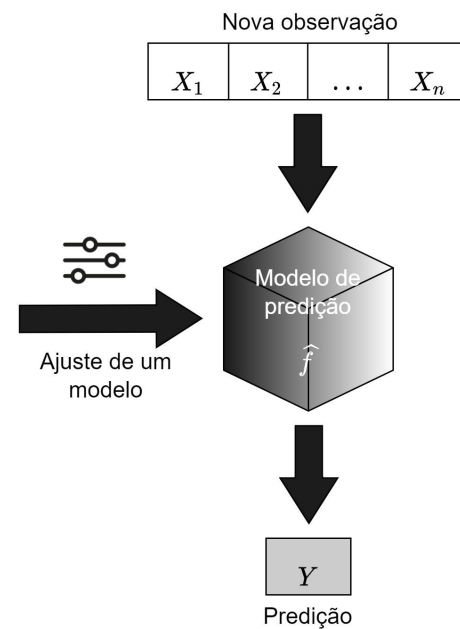
- Aprendizado de máquina
 - O que é interpretabilidade?
 - Caixas brancas, cinzas e pretas
 - Explicando um modelo não interpretável
 - Avaliação de explicações
-

Aprendizado de máquina

Tarefa de construir um modelo, derivado de dados, para realizar novas previsões.

- Diferentes paradigmas:
 - **Supervisionado**, não supervisionado, reforço.
- Diferentes tarefas:
 - **Regressão** ou classificação.

Variáveis independentes				Rótulo
$X_{(1,1)}$	$X_{(1,2)}$...	$X_{(1,n)}$	Y_1
$X_{(2,1)}$	$X_{(2,2)}$...	$X_{(2,n)}$	Y_2
\vdots	\vdots	\ddots	\vdots	\vdots
$X_{(m,1)}$	$X_{(m,2)}$...	$X_{(m,n)}$	Y_m



Esta dissertação avaliou o aprendizado **supervisionado** para a tarefa de **regressão**.

O que é interpretabilidade?

O termo possui diferentes definições na literatura, mas se relaciona com o nível de entendimento do modelo.

Definição de modelo interpretável

Considero um modelo interpretável um modelo que:

- Pode ser inspecionado para entender como as entradas são mapeadas às saídas;
- Pode ser decomposto em componentes que tenham sentido no domínio de aplicação.

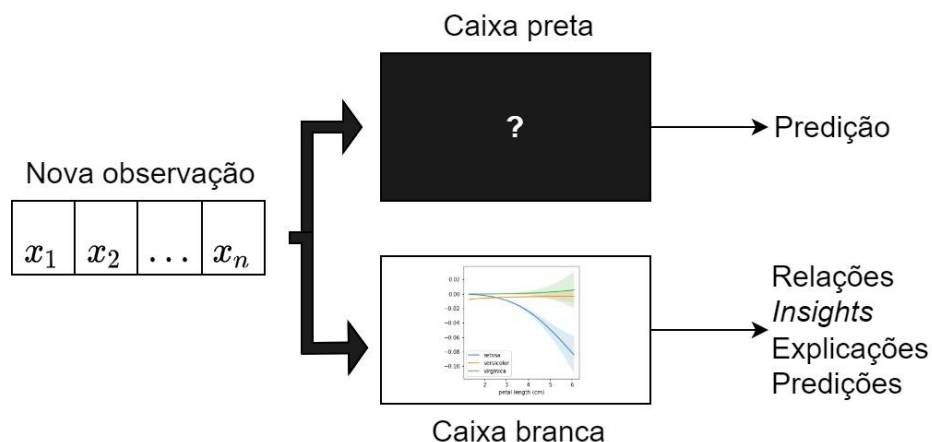
Caixas brancas, cinzas e pretas

Definição de caixas brancas, pretas e cinzas

Caixa-branca: modelo analítico que segue um estudo teórico do sistema, modelo simples considerado interpretável;

Caixa-preta: modelo derivado de dados sem conhecimento sobre seu funcionamento interno;

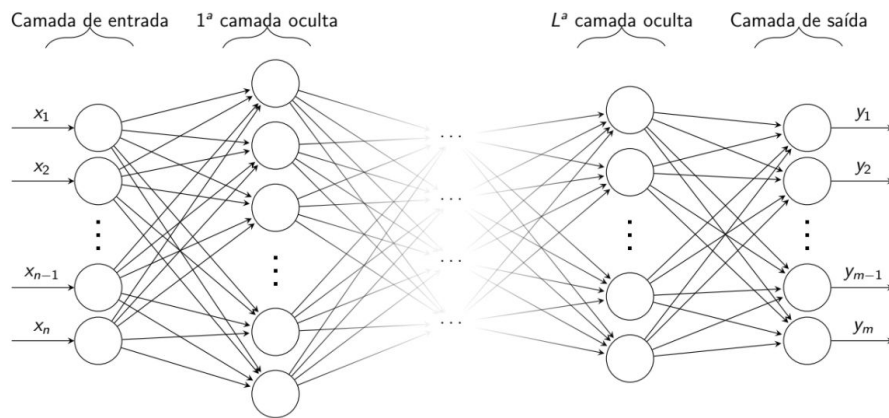
Caixa-cinza: modelo derivado de dados, mas aberto para inspeção (mas não se categoriza como uma caixa branca).



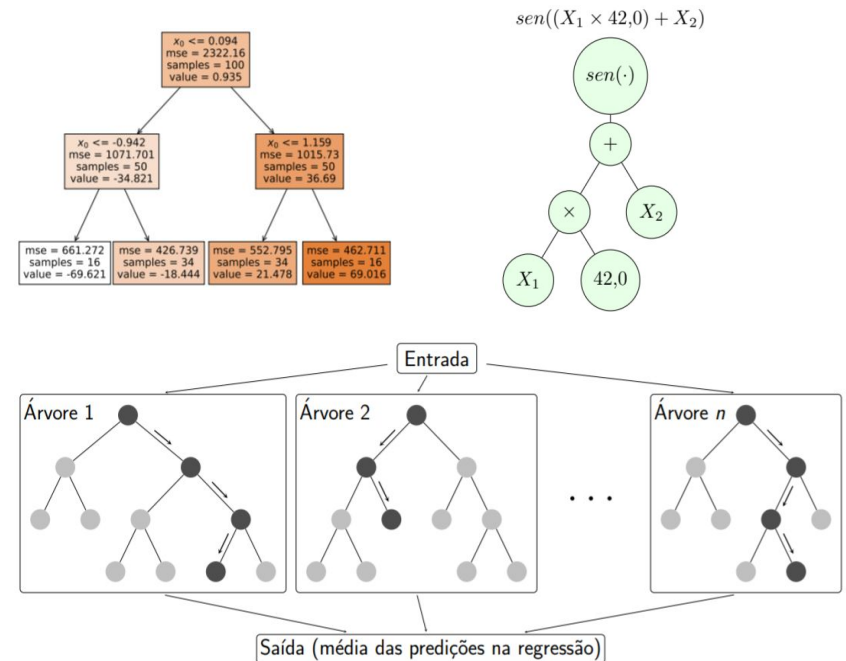
Caixas brancas, cinzas e pretas

O grau de **interpretabilidade** se relaciona com a complexidade do modelo construído.

Modelos mais simples geralmente são ditos caixas-brancas. A regressão simbólica é geralmente vista como uma caixa-cinza.



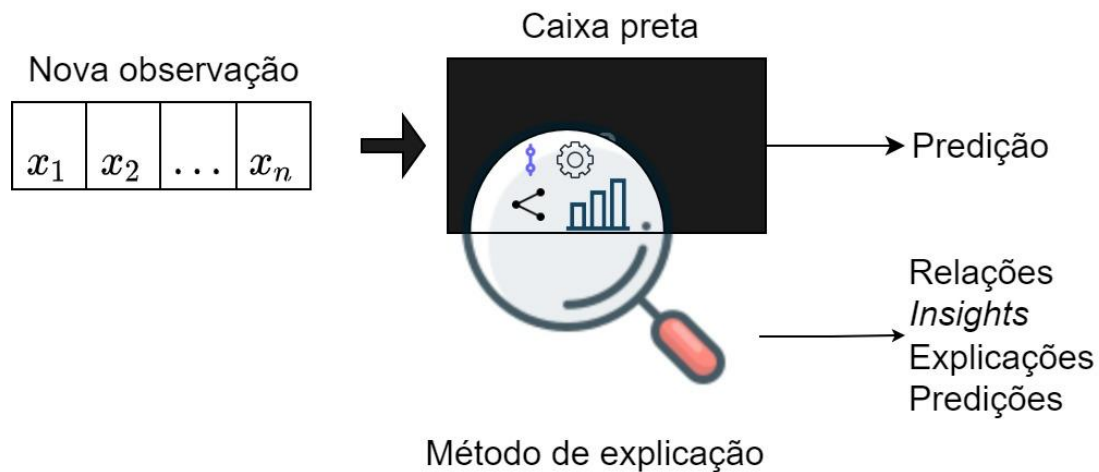
$$\hat{f}(\mathbf{x}) = 42.853x_0 + -1.628$$



Explicando um modelo não interpretável

Quando o modelo não é uma caixa-branca, existem 2 alternativas:

- Substituir por um modelo mais interpretável;
- Usar um **método de explicação** para melhorar o nível de entendimento da caixa-preta.

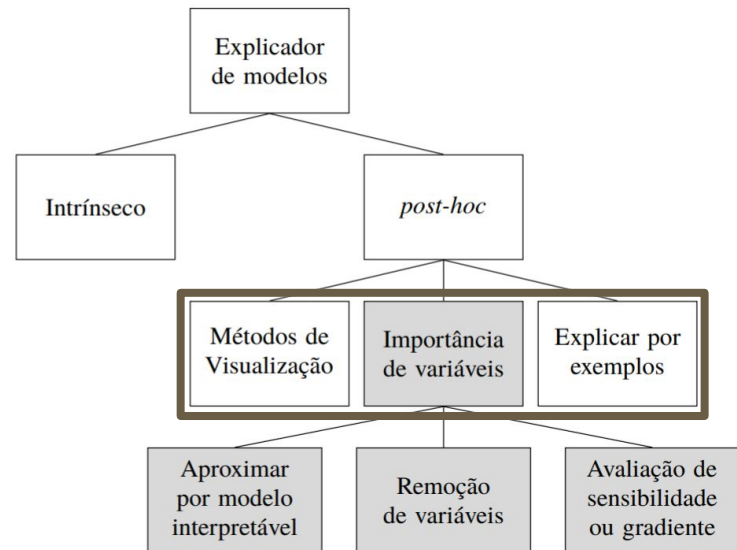


Explicando um modelo não interpretável

Um método de explicação busca fornecer uma explicação para um modelo caixa-preta.

- Diferentes tipos de explicações e diferentes métodos de explicação na literatura:
 - Visualização, **importância de variáveis**, explicar por exemplos;
 - Maior foco é na explicação por importância de variáveis

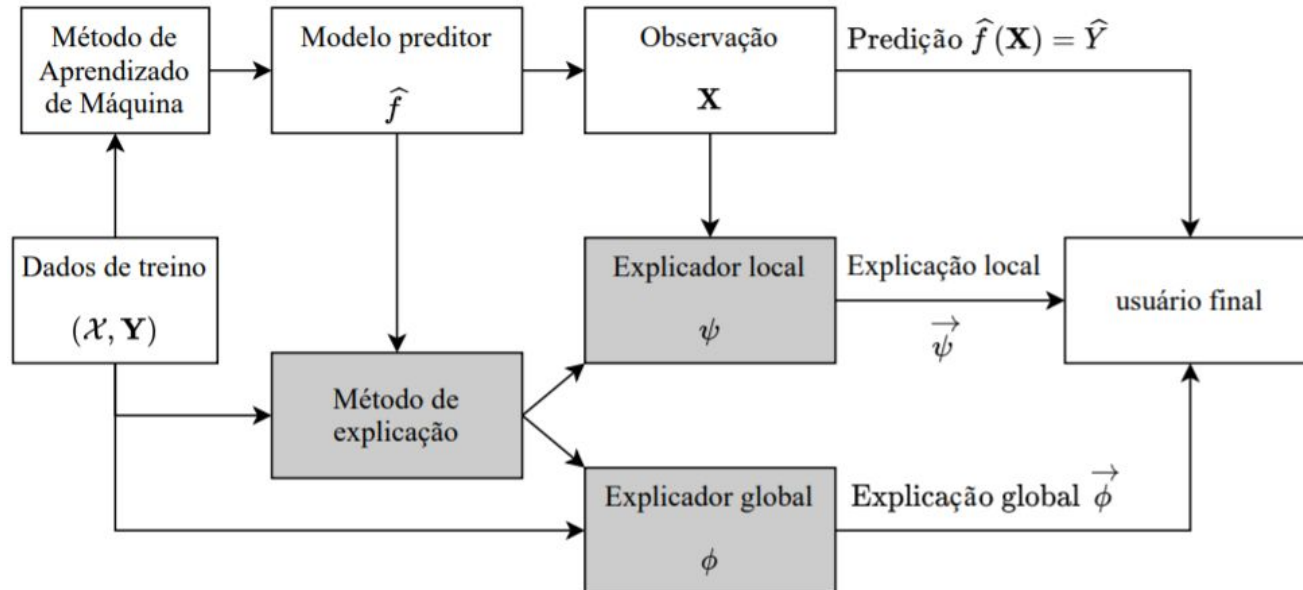
Abordagens	
Específico de modelo	Agnóstico de modelo
Intrínseco	<i>Post-hoc</i>
Local	Global



Explicando um modelo não interpretável

Explicação por importância de variáveis

Gerar uma explicação que atribui uma importância para cada variável do problema.



Avaliação de explicações

Dados podem ser reais ou sintéticos:

- Explicação esperada definida heurísticamente ou utilizando modelos simples conhecidos.

Humanos no processo:

- Especialistas do domínio, pessoas leigas, sem humanos.

Métricas:

- Estabilidade, (in)fidelidade, sensibilidade, estabilidade de Jaccard, robustez-S.

Avaliação ainda é uma **questão em aberto**:

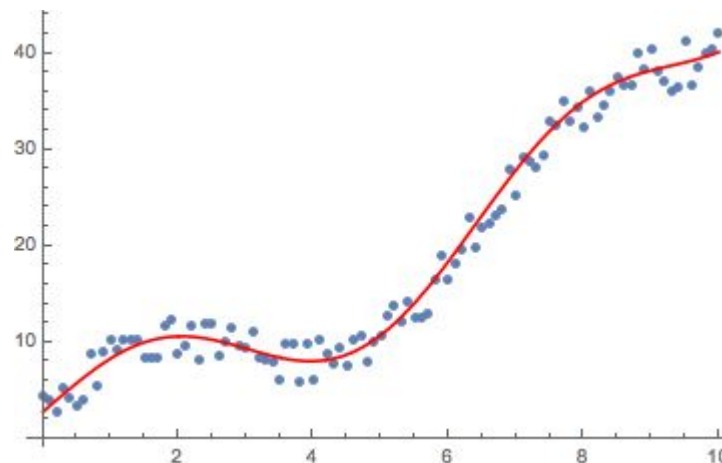
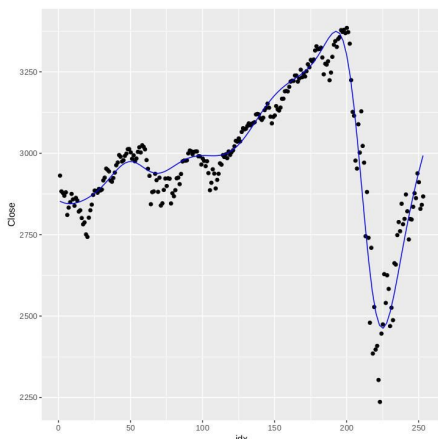
Não existe explicação esperada em vários problemas do mundo real, falta uma definição matemática de interpretabilidade, e é um conceito difícil de medir.

Regressão Simbólica: uma alternativa caixa-cinza

- O que é regressão simbólica
 - Regressão simbólica canônica
 - Estado-da-arte
 - Representação alternativa
-

O que é regressão simbólica?

- Método de regressão não paramétrico;
- Geralmente implementada por meio de algoritmos evolutivos;
- Vista como mais interpretável do que caixas-pretas.

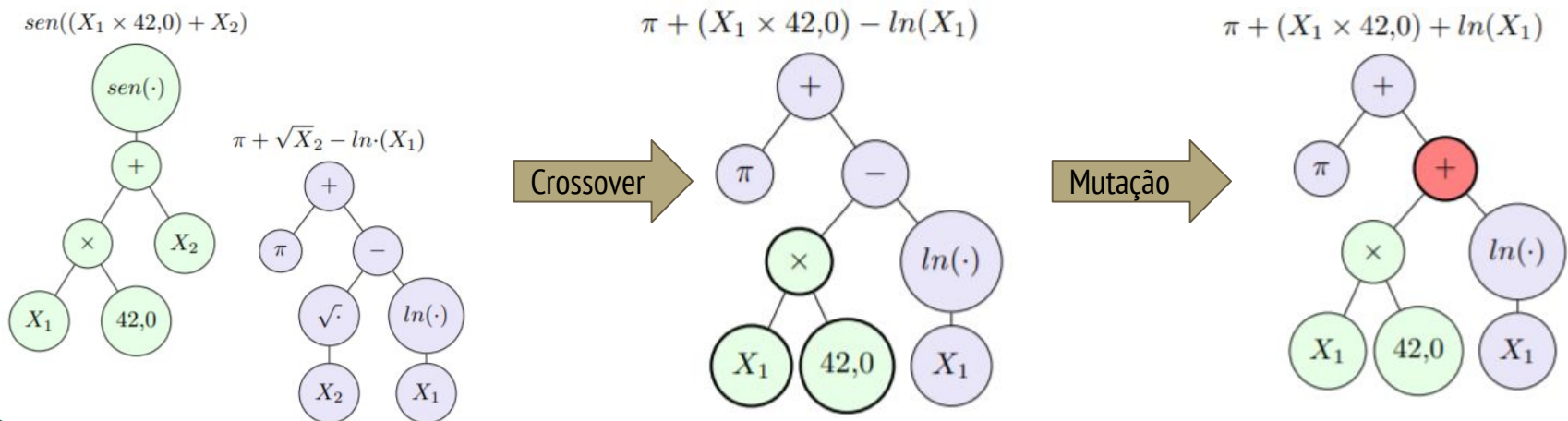


Regressão simbólica canônica (*Genetic Programming*)

Representação por árvores de expressões: raiz é o ponto de entrada, e nós são terminais (constantes/variáveis) ou intermediários (funções n-árias);

Constantes criadas pelo ERC ou sorteando nós pré definidos;

A ideia do algoritmo é simular a evolução sobre uma população de expressões. A evolução é guiada por uma função de aptidão, e operadores genéticos criam descendentes que herdam as características de seus pais.

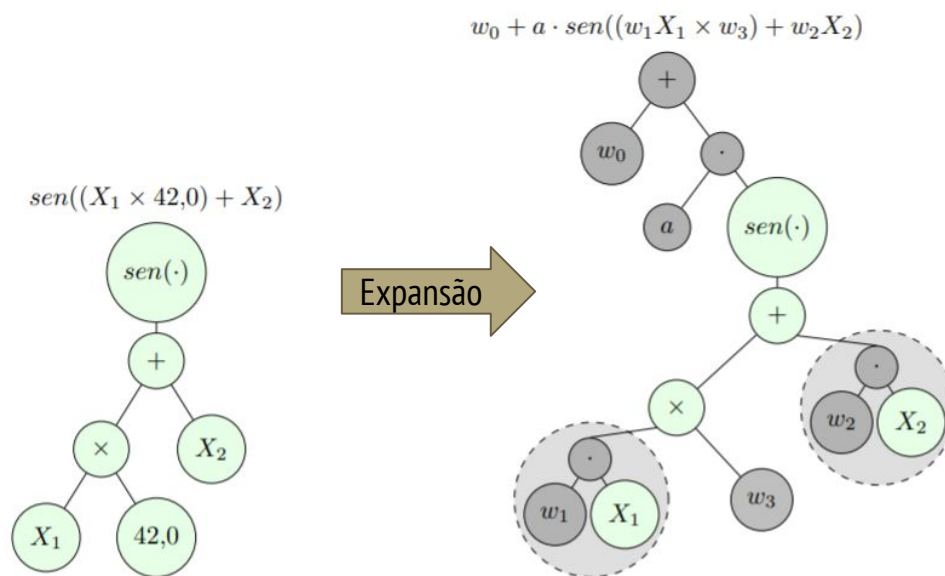


Estado-da-arte (Operon GP-NLS)

Representação por árvores de expressões expandidas: adiciona um intercepto, coeficiente de escala, e um coeficiente para cada variável;

Os coeficientes ajustados pelo método Levenberg-Marquardt, representando um aprendizado local;

A ideia do algoritmo é realizar a regressão simbólica como o GP, mas fazendo um ajuste local dos coeficientes.



Genetic Programming and Evolvable Machines (2020) 21:471–501
<https://doi.org/10.1007/s10710-019-09371-3>



Parameter identification for symbolic regression using nonlinear least squares

Michael Kommenda¹ · Bogdan Burlacu¹ · Gabriel Kronberger¹ · Michael Affenzeller^{1,2}

Received: 20 January 2019 / Revised: 4 November 2019 / Published online: 10 December 2019
© The Author(s) 2019

Abstract

In this paper we analyze the effects of using nonlinear least squares for parameter identification of symbolic regression models and integrate it as local search mechanism in tree-based genetic programming. We employ the Levenberg–Marquardt algorithm for parameter optimization and calculate gradients via automatic differentiation. We provide examples where the parameter identification succeeds and fails and highlight its computational overhead. Using an extensive suite of symbolic regression benchmark problems we demonstrate the increased performance when incorporating nonlinear least squares within genetic programming. Our results are compared with recently published results obtained by several genetic programming variants and state of the art machine learning algorithms. Genetic programming with nonlinear least squares performs among the best on the defined benchmark suite and the local search can be easily integrated in different genetic programming algorithms as long as only differentiable functions are used within the models.

Keywords Genetic programming · Symbolic regression · Parameter identification · Nonlinear least squares · Automatic differentiation

Representação alternativa (ITEA)

Representação Interação-Transformação (IT): descreve a soma de termos com uma função de transformação sobre interações das variáveis do problema;

Os coeficientes são ajustados pelo método dos mínimos quadrados;

A ideia do algoritmo é restringir o espaço de busca (por meio da representação) em um espaço com soluções mais simples, utilizando apenas mutação.

$$\underbrace{\overbrace{42.0}^{\text{Coef}} \cdot \text{id}(X_3)}_{\text{Termo1}} - \underbrace{\overbrace{23.4}^{\text{Coef}} \cdot \text{sen}(X_1 \cdot X_2)}_{\text{Termo2}} + \underbrace{\overbrace{12.3}^{\text{Coef}} \cdot \sqrt{\left(\frac{1}{X_1^2}\right)}}_{\text{Termo3}} + \underbrace{100.0}_{\text{Intercepto}}$$

```
intercepto      = 100.0
coeficientes    = [42.0, -23.4, 12.3]
transformações  = [id, sen, sqrt]
expo_interações = [[0, 0, 1],
                   [1, 1, 0],
                   [-2, 0, 0]]
```

Interaction-Transformation Evolutionary Algorithm for Symbolic Regression

F. O. de Franca folivetti@ufabc.edu.br
Center for Mathematics, Computation and Cognition, Heuristics, Analysis and Learning Laboratory, Federal University of ABC, Santo Andre, Brazil –
G. S. I. Aldeia guilherme.aldeia@ufabc.edu.br
Center for Mathematics, Computation and Cognition, Heuristics, Analysis and Learning Laboratory, Federal University of ABC, Santo Andre, Brazil

Abstract

Interaction-Transformation (IT) is a new representation for Symbolic Regression that reduces the space of solutions to a set of expressions that follow a specific structure. The potential of this representation was illustrated in prior work with the algorithm called SymTree. This algorithm starts with a simple linear model and incrementally introduces new transformed features until a stop criteria is met. While the results obtained by this algorithm were competitive with the literature, it had the drawback of not scaling well with the problem dimension. This paper introduces an mutation only Evolutionary Algorithm, called ITEA, capable of evolving a population of IT expressions. One advantage of this algorithm is that it enables the user to specify the maximum number of terms in an expression. In order to verify the competitiveness of this approach, ITEA is compared to linear, nonlinear and Symbolic Regression models from the literature. The results indicate that ITEA is capable of finding equal or better approximations than other Symbolic Regression models while being competitive to state-of-the-art non-linear models. Additionally, since this representation follows a specific structure, it is possible to extract the importance of each original feature of a data set as an analytical function, enabling us to automate the explanation of any prediction. In conclusion, ITEA is competitive when comparing to regression models with the additional benefit of automating the extraction of additional information of the generated models.

Keywords

Symbolic Regression, Interaction-Transformation, evolutionary algorithms.

Medidas de avaliação de explicações

- Robustez e qualidade de explicações
 - Medidas de robustez
 - Medidas de qualidade
-

Robustez e qualidade de explicações

Robustez: observações similares devem ter explicações similares:

- Explicação robusta deve ter pouca variação!

Qualidade: se temos explicação esperada, a explicação dada deve se aproximar dela:

- As explicações comparadas devem ter o mesmo significado!
- Explicações de qualidade deve ser parecidas com a explicação esperada!

Medidas de robustez

Estabilidade: grau em que a explicação local muda em relação à vizinhos com variações mínimas. Essa medida só leva em consideração a explicação.

$$S(\hat{f}, \psi, \mathbf{X}) = \mathbb{E}_{\mathbf{X}' \in N_{\mathbf{X}}} \left[\|\psi(\hat{f}, \mathbf{X}) - \psi(\hat{f}, \mathbf{X}')\|_2^2 \right]$$

(In)fidelidade: considera a magnitude da perturbação aplicada, a explicação dada e a diferença da predição do modelo.

$$\text{INFD}(\hat{f}, \psi, \mathbf{X}) = \mathbb{E}_{\mathbf{I} \in \mu_I} \left[(\mathbf{I}^T \psi(\hat{f}, \mathbf{X}) - (\hat{f}(\mathbf{X}) - \hat{f}(\mathbf{X} - \mathbf{I})))^2 \right]$$

Estabilidade de Jaccard: o quão similar o subconjunto de k variáveis mais importantes é para vizinhos próximos.

$$J(\hat{f}, \psi, \mathbf{X}) = \mathbb{E}_{\mathbf{X}' \in N_{\mathbf{X}}} \left[\frac{|\tilde{\psi}(\mathbf{X})_k \cap \tilde{\psi}(\mathbf{X}')_k|}{|\tilde{\psi}(\mathbf{X})_k \cup \tilde{\psi}(\mathbf{X}')_k|} \right]$$

$$N_{\mathbf{X}} \sim \mathcal{N}_n(\mathbf{X}, \lambda \cdot \text{cov}(\mathcal{X}))$$

$$\mu_I \sim \mathbf{X} - \mathcal{N}_d(\mathbf{X}, \lambda \cdot \text{cov}(\mathcal{X}))$$

Medidas de qualidade

Similaridade de cossenos: cosseno entre dois vetores. Tendo uma explicação esperada, calculamos o cosseno entre ela e a explicação dada.

$$\cos(\vec{e}, \vec{\phi}) = \frac{\vec{e} \cdot \vec{\phi}}{\|\vec{e}\|_2 \|\vec{\phi}\|_2}$$

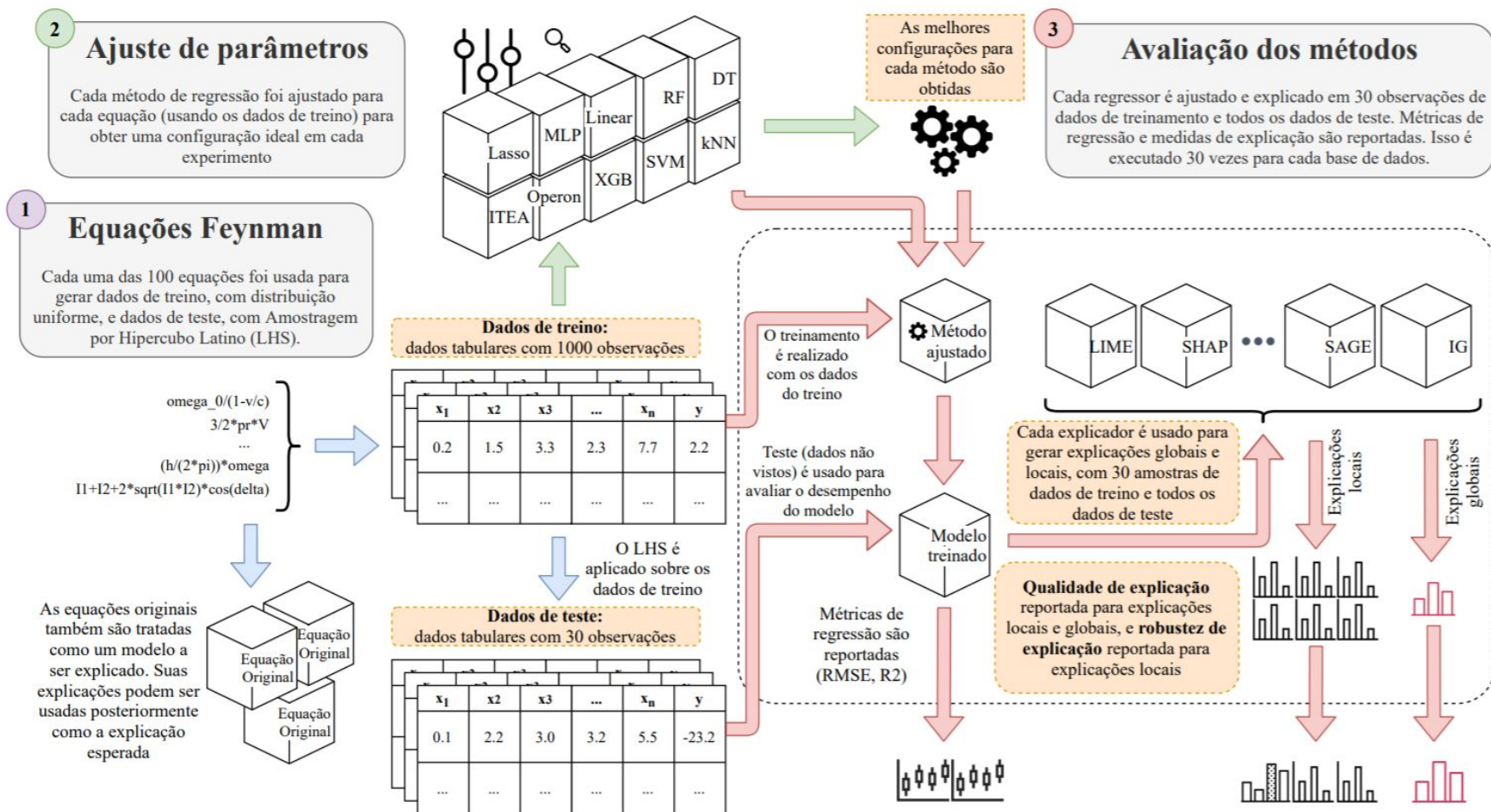
RMSE: desvio da explicação obtida com uma explicação observada, apresentando o quão distante o modelo está de dar a importância certa, na média entre as importâncias dadas para diferentes variáveis.

$$\text{RMSE}(\vec{e}, \vec{\phi}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - \phi_i)^2}$$

Avaliação de métodos de interpretabilidade

- Metodologia
 - Resultados e discussão
 - Conclusões dos experimentos
-

Metodologia: procedimentos dos experimentos



Metodologia: métodos considerados

Foram considerados 10 métodos de aprendizado de máquina para regressão:

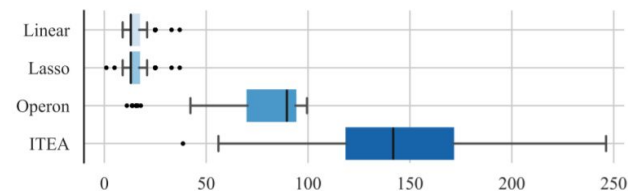
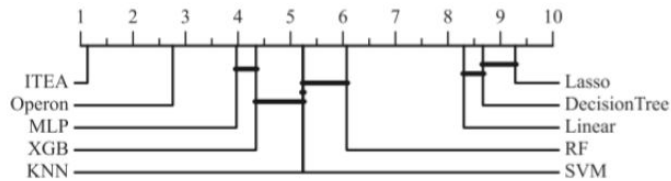
Método de regressão	Nível de interpretabilidade	Tipo
XGB*	Caixa-preta	<i>Tree boosting</i>
RF*	Caixa-preta	<i>Tree bagging</i>
MLP*	Caixa-preta	Rede neural sem realimentação
SVM	Caixa-preta	Máquina de vetores
Operon*	Caixa-cinza	Regressão simbólica
ITEA*	Caixa-cinza	Regressão simbólica
kNN	Caixa-branca	Método de instância
Linear regression	Caixa-branca	Modelagem de regressão
LASSO regression	Caixa-branca	Modelagem de regressão
Single Decision Tree*	Caixa-branca	Árvore de decisão

E 8 métodos de explicação com diferentes características:

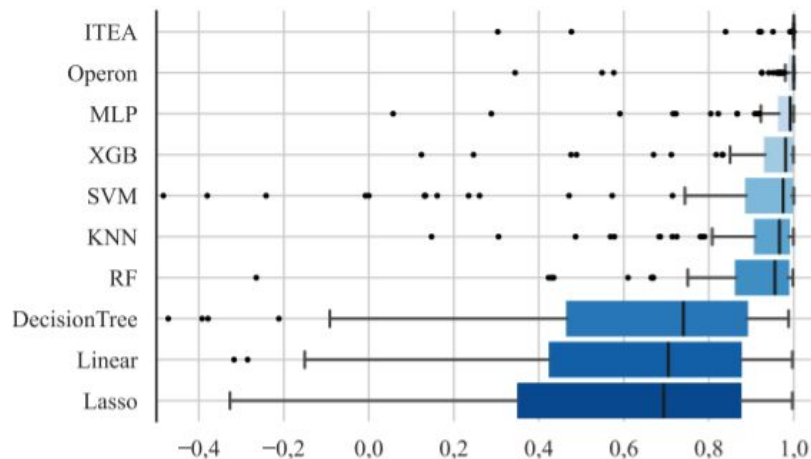
Método de explicação	Agnóstico	Local	Global	Tipo
<i>Permutation Importance</i>	S	N	S	Remoção de variáveis
SHAP	S	S	S	Remoção de variáveis
SAGE	S	N	S	Remoção de variáveis
<i>Random Importance</i>	S	S	S	Atribuição aleatória
<i>Morris sensitivity</i>	S	N	S	Análise de sensibilidade
<i>Integrated Gradients</i>	S	S	N	Análise de gradiente
<i>Partial Effects</i>	N	S	S	Análise de gradiente
LIME	S	S	N	Aproximar por modelo interpretável

Metodologia: Análise dos resultados

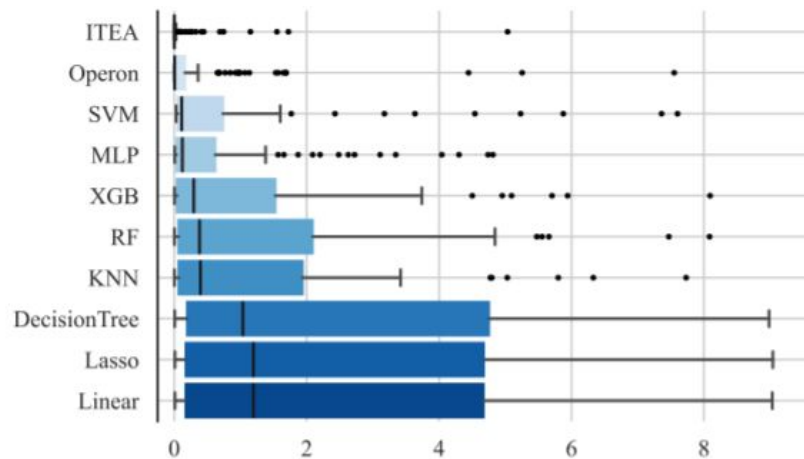
- Resultados são reportados utilizando uma partição de testes;
- 18,500 execuções no total:
 - Estocásticos executados 30x, e a mediana das execuções foi utilizada
- Visualização dos resultados:
 - Heatmaps;
 - Boxplots;
 - Diagramas críticos.



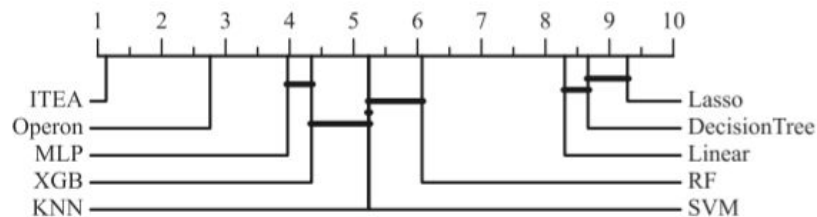
Resultados: erro de predição



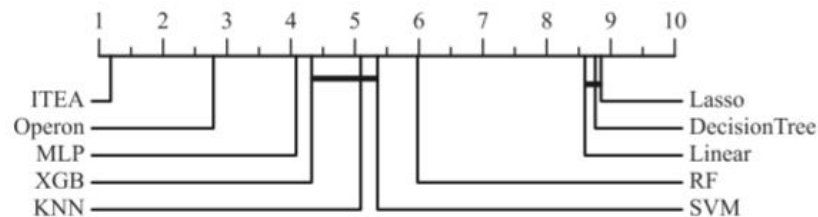
(a) Pontuação R^2 no teste



(b) Métrica RMSE no teste

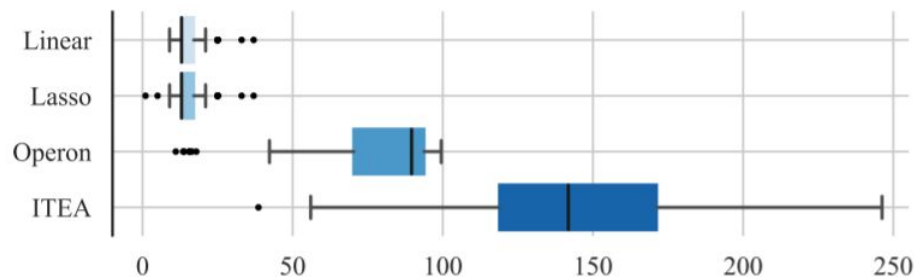


(c) Diagrama crítico do R^2 no teste



(d) Diagrama crítico do RMSE no teste

Resultados: complexidade de expressões



Método	Menor expressão na base com mais variáveis (9)
Linear	$0.196 \cdot x_0 + 0.187 \cdot x_1 + 0.193 \cdot x_2 + -0.095 \cdot x_3 + 0.097 \cdot x_4 + -0.105 \cdot x_5 + 0.099 \cdot x_6 + -0.094 \cdot x_7 + 0.094 \cdot x_8 + 0.014$
Lasso	$0.182 \cdot x_0 + 0.177 \cdot x_1 + 0.179 \cdot x_2 + -0.084 \cdot x_3 + 0.084 \cdot x_4 + -0.091 \cdot x_5 + 0.089 \cdot x_6 + -0.081 \cdot x_7 + 0.082 \cdot x_8 + -0.011$
Operon	$((-0.00135) + (3.69096 * ((((((1.42081 * x_5) + 1.44003) / \exp((-0.34371) * x_9))) / (((0.48698 * x_8) + (0.22187 * x_4)) * \sqrt{(\sqrt{(0.49086 * x_6)) + \dots$
ITEA	$0.022 \cdot \log(x_0^4 * x_1^3 * x_2^3 * x_3^{-1} * x_4^{-3} * x_5 * x_6 * x_7^{-2}) + 672.83 \cdot \exp(x_0^2 * x_1^2 * x_2^2 * x_3^{-4} * x_4 * x_5^{-3} * x_7^{-4} * x_8) + 0.665 \cdot \sqrt{x_0^4 * x_1^4 * x_2^4 * x_3^{-3} * x_4^2 * x_5^{-4} * x_6^4 * x_7^{-4} * x_8^3} + \dots$

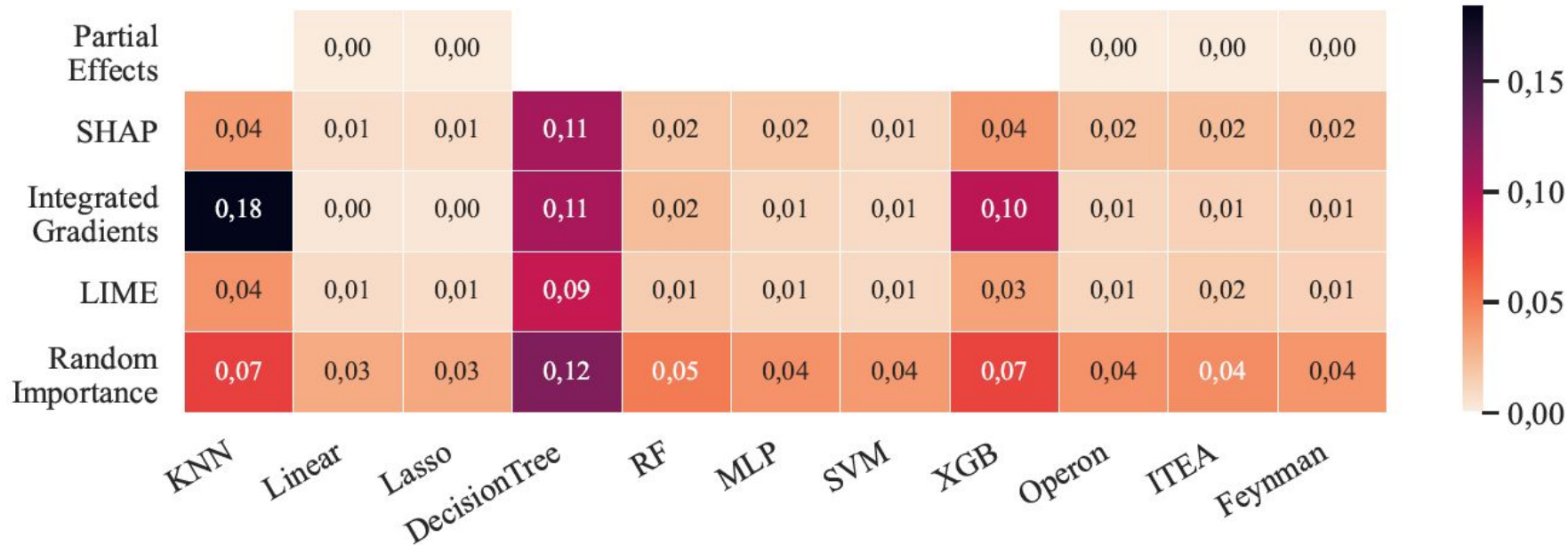
Resultados: robustez de explicações locais



(a) Mapa de calor da mediana da estabilidade de Jaccard (maiores valores são melhores) para cada combinação de regressor-explicador sobre os conjuntos de dados de Feynman

$$J(\hat{f}, \psi, \mathbf{X}) = \mathbb{E}_{\mathbf{X}' \in N_{\mathbf{X}}} \left[\frac{|\tilde{\psi}(\mathbf{X})_k \cap \tilde{\psi}(\mathbf{X}')_k|}{|\tilde{\psi}(\mathbf{X})_k \cup \tilde{\psi}(\mathbf{X}')_k|} \right]$$

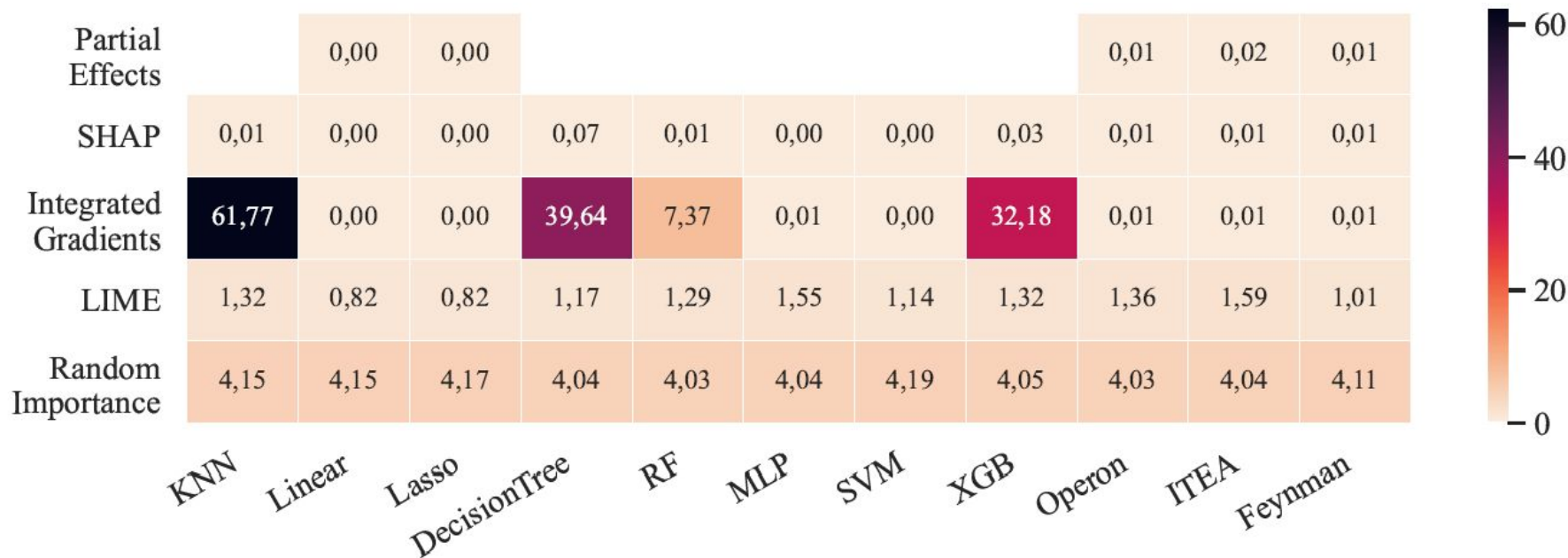
Resultados: robustez de explicações locais



(b) Mapa de calor da mediana da infidelidade (menores valores são melhores) para cada combinação de regressor-explicador sobre o conjunto de bases de dados de Feynman

$$\text{INFD}(\hat{f}, \psi, \mathbf{X}) = \mathbb{E}_{\mathbf{I} \in \mu_I} \left[(\mathbf{I}^T \psi(\hat{f}, \mathbf{X}) - (\hat{f}(\mathbf{X}) - \hat{f}(\mathbf{X} - \mathbf{I})))^2 \right]$$

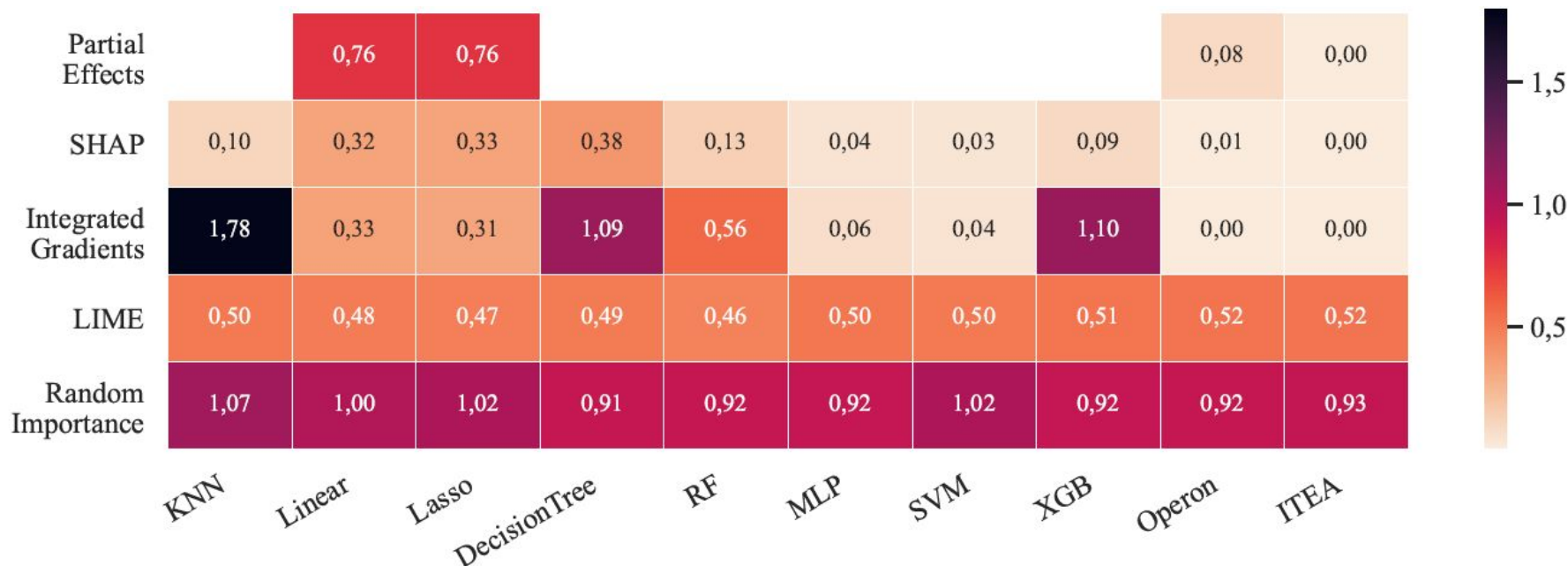
Resultados: robustez de explicações locais



(c) Mapa de calor da mediana da estabilidade (menores valores são melhores) para cada combinação de regressor-explicador sobre o conjunto de bases de dados de Feynman

$$S(\hat{f}, \psi, \mathbf{X}) = \mathbb{E}_{\mathbf{X}' \in N_{\mathbf{X}}} \left[\|\psi(\hat{f}, \mathbf{X}) - \psi(\hat{f}, \mathbf{X}')\|_2^2 \right]$$

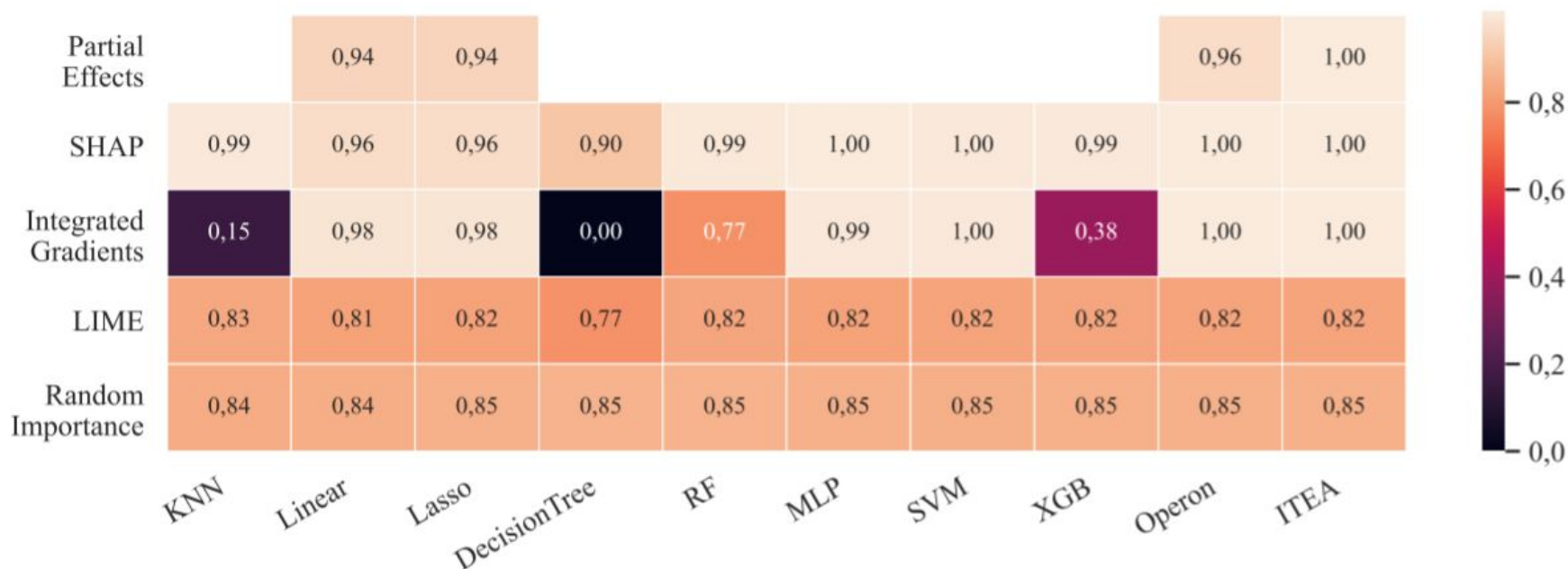
Resultados: qualidade de explicações locais



(a) Mapa de calor da mediana do RMSE (menores valores são melhores) calculado entre as explicações locais obtidas sobre o regressor e as explicações obtidas sobre a equação de original

$$\text{RMSE}(\vec{e}, \vec{\phi}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - \phi_i)^2}$$

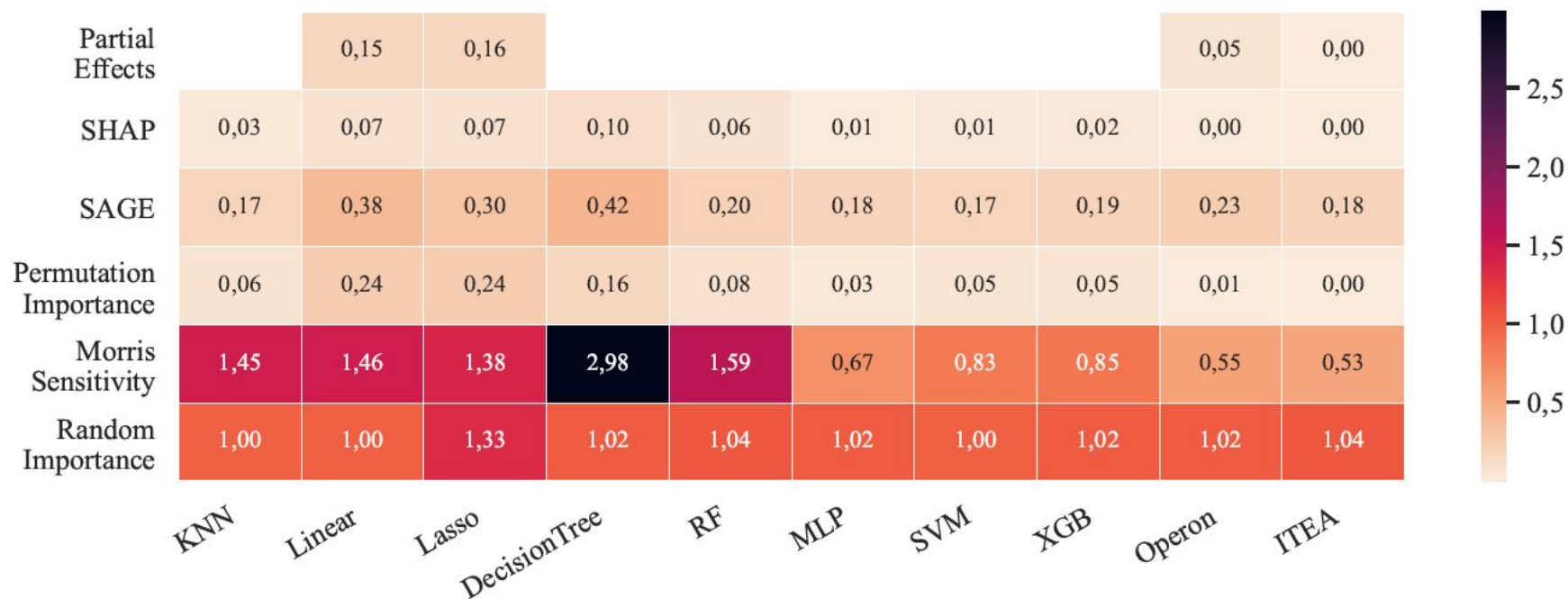
Resultados: qualidade de explicações locais



(b) Mapa de calor da mediana da similaridade de cossenos (maiores valores são melhores) calculado entre as explicações locais obtidas sobre o regressor e as explicações obtidas sobre a equação de original

$$\cos(\vec{e}, \vec{\phi}) = \frac{\vec{e} \cdot \vec{\phi}}{\|\vec{e}\|_2 \|\vec{\phi}\|_2}$$

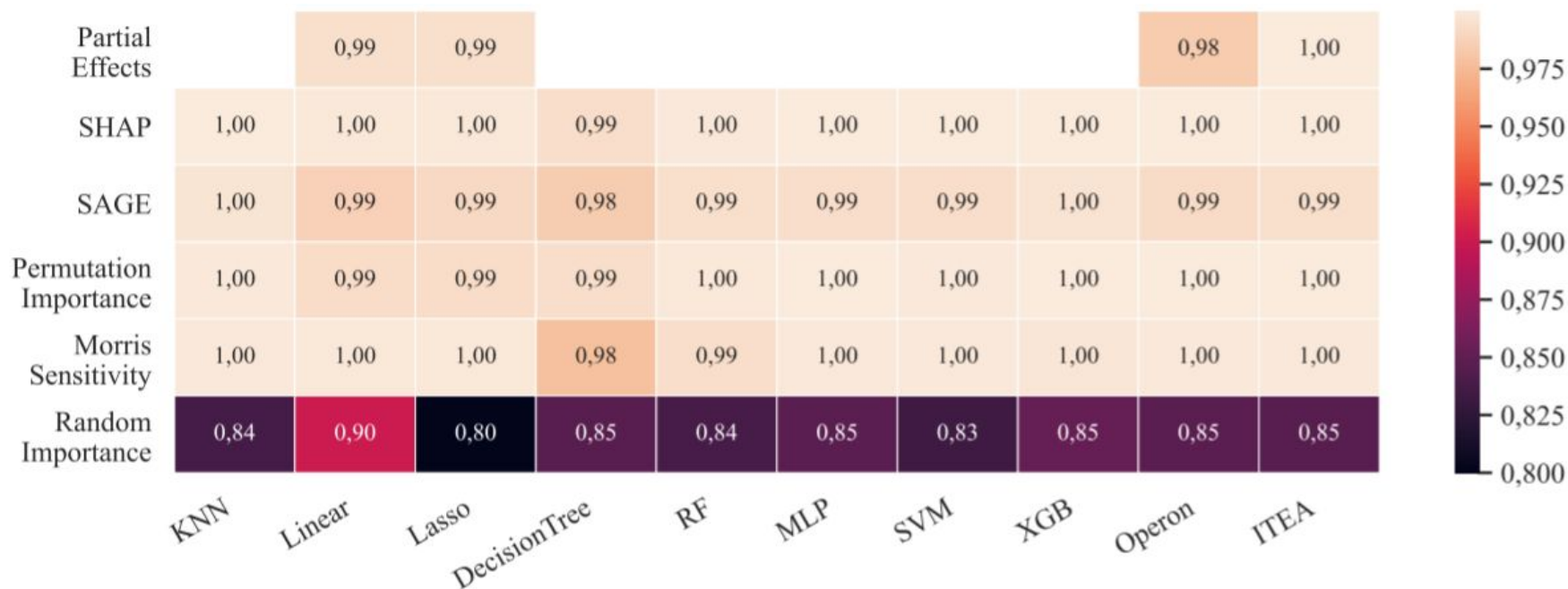
Resultados: qualidade de explicações globais



(a) Mapa de calor da mediana do RMSE (menores valores são melhores) calculado entre as explicações globais obtidas sobre o regressor e as explicações obtidas sobre a equação de original

$$\text{RMSE}(\vec{e}, \vec{\phi}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - \phi_i)^2}$$

Resultados: qualidade de explicações globais



(b) Mapa de calor da mediana da similaridade de cossenos (maiores valores são melhores) calculado entre as explicações globais obtidas sobre o regressor e as explicações obtidas sobre a equação de original

$$\cos(\vec{e}, \vec{\phi}) = \frac{\vec{e} \cdot \vec{\phi}}{\|\vec{e}\|_2 \|\vec{\phi}\|_2}$$

Conclusões dos experimentos

- Os resultados indicam que o *Partial Effects* (específico de modelo) pode providenciar melhores explicações para a regressão simbólica que métodos agnósticos;
 - Indicativo de que a informação extra por ser específico de modelo leva à explicações mais robustas
 - O problema de ser específico de modelo é a dificuldade de implementação para diferentes casos
- Mesmo se considerando apenas métodos agnósticos, a regressão simbólica apresenta resultados competitivos, estando entre os melhores resultados obtidos;
- A representação do modelo de aprendizado de máquina influencia no nas medidas de explicações

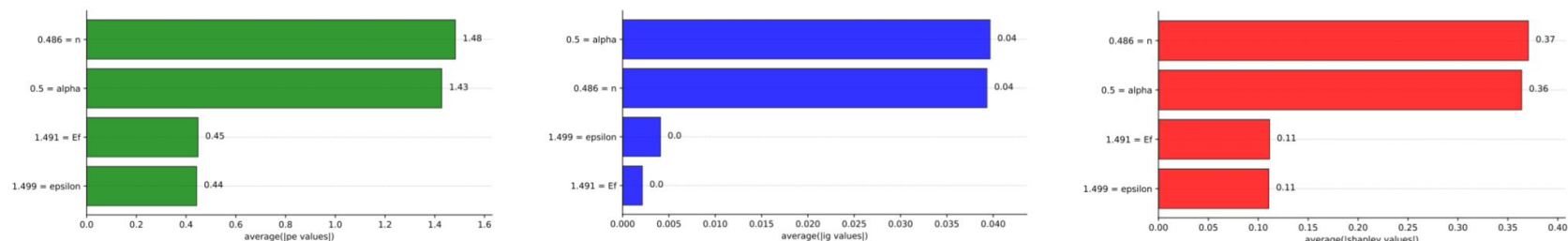
Gerador de relatórios automático

- Métodos utilizados
 - Gerando um relatório
 - Saiba mais
-

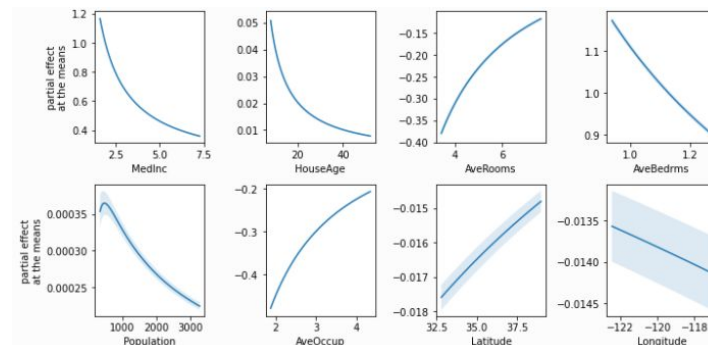
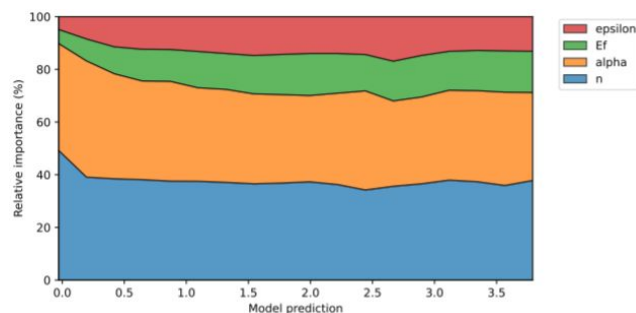
Métodos utilizados

O método de regressão simbólica utilizado foi o ITEA, explicado por:

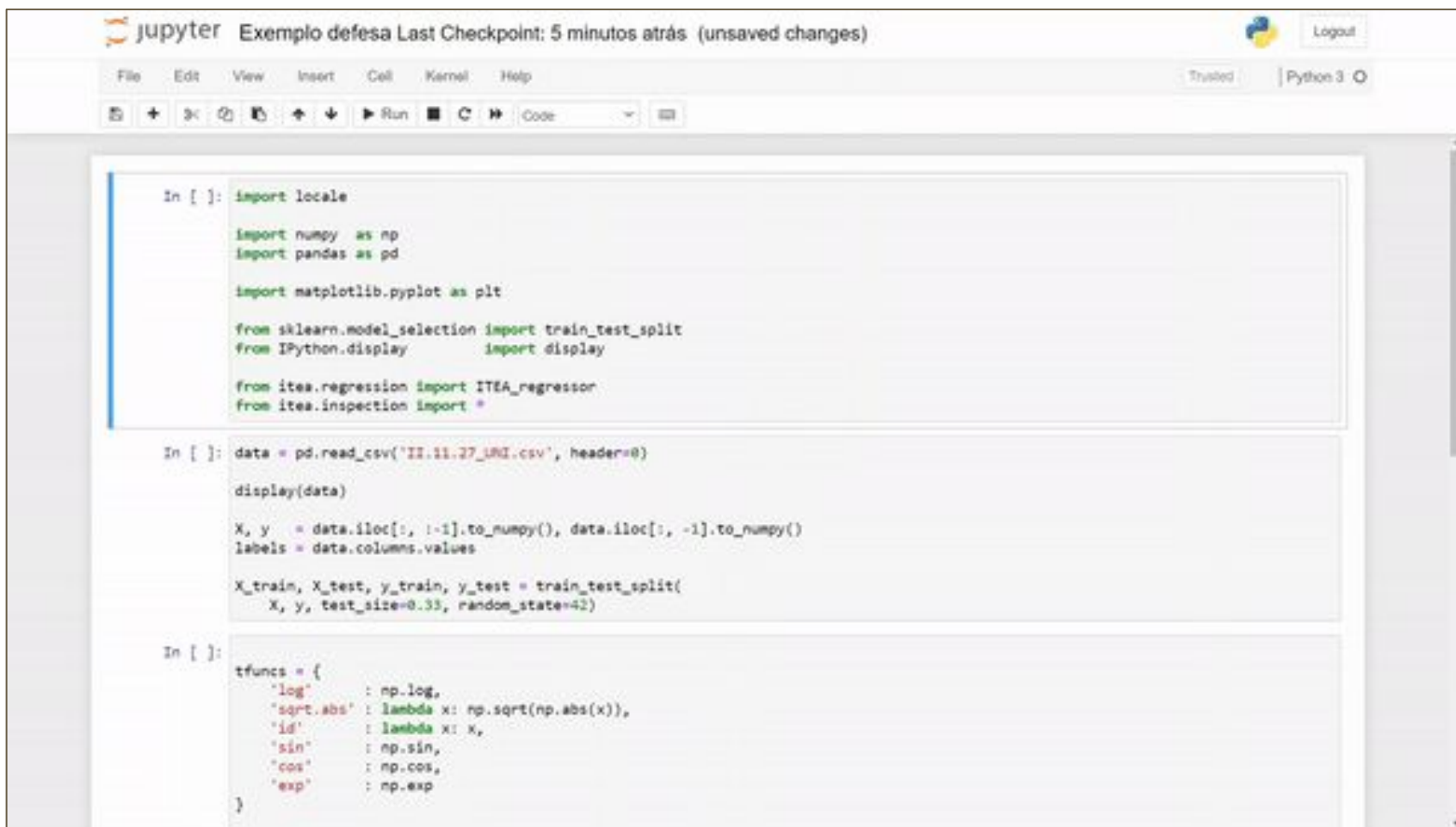
- Atribuição de importância: *Integrated Gradients*, SHAP e *Partial Effects*;



- Visualização: *Partial Effects at the means* e gráfico de importâncias empilhadas.



Gerando um relatório



The screenshot shows a Jupyter Notebook window titled "Exemplo defesa Last Checkpoint: 5 minutos atrás (unsaved changes)". The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Help), a toolbar with icons for file operations and execution, and a status bar indicating "Trusted" and "Python 3". The notebook contains three code cells:

```
In [ ]: import locale

import numpy as np
import pandas as pd

import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split
from IPython.display import display

from itea.regression import ITEA_regressor
from itea.inspection import *
```

```
In [ ]: data = pd.read_csv('II.11.27_UR1.csv', header=0)

display(data)

X, y = data.iloc[:, :-1].to_numpy(), data.iloc[:, -1].to_numpy()
labels = data.columns.values

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.33, random_state=42)
```

```
In [ ]: tfuncs = {
    'log' : np.log,
    'sqrt.abs' : lambda x: np.sqrt(np.abs(x)),
    'id' : lambda x: x,
    'sin' : np.sin,
    'cos' : np.cos,
    'exp' : np.exp
}
```

Saiba mais



Documentação:

<https://itea-python.readthedocs.io>

PyPi:

<https://pypi.org/project/itea/>

Código fonte:

<https://github.com/gAldeia/itea-python>

Experimentos:

<https://github.com/gAldeia/iirsBenchmark>

Conclusões e considerações finais

- Conclusões
 - Contribuições
 - Considerações finais
-

Conclusões

- Foi elaborada uma metodologia experimental para avaliar a qualidade e robustez de explicações de importância de atributos;
- Experimentos no contexto da regressão simbólica foram introduzidos;
- A ideia do efeito parcial foi introduzida como explicador específico da regressão simbólica;
- Foram feitas várias implementações de métodos de explicação;
- Foi possível comparar o desempenho da regressão simbólica e atingir todos os objetivos;
- A própria função retornada pela regressão simbólica pode ser uma forma de explicação, e possibilita extrair mais formas de explicar o modelo.

Contribuições

- Revisão bibliográfica;
- Formalização de explicações por importância de atributos;
- Apresentação de métricas para avaliação de explicações;
- *Partial effects* como explicador da regressão simbólica;
- Experimentos disponibilizados como biblioteca;
- ITEA e *Partial Effects* disponibilizados como biblioteca.

Considerações finais

Possíveis vieses:

- Avaliação apenas com bases de dados do domínio da física;
- Generalização para todos os domínios de aplicação pode não ser real.

Trabalhos futuros:

- Expandir a quantidade de modelos de regressão simbólica;
- Aumentar o conjunto de bases de dados dos experimentos.

Artigos publicados

ALDEIA, G. S. I.; de FRANÇA, F. O. A parametric study of interaction-transformation evolutionary algorithm for symbolic regression. In: **2020 IEEE Congress on Evolutionary Computation (CEC)**. New York: IEEE, 2020. Disponível em:

<<https://doi.org/10.1109%2Fcec48606.2020.9185521>>.

de FRANÇA, F. O.; ALDEIA, G. S. I. Interaction-transformation evolutionary algorithm for Symbolic Regression. **Evolutionary Computation**, v. 29, n. 3, p. 367–390, 2021. ISSN 1063-6560. Disponível em: <https://doi.org/10.1162/evco_a_00285>.

ALDEIA, G. S. I.; de FRANÇA, F. O. Measuring Feature Importance using Marginal Effects with Symbolic Regression. In: **2021 ACM Genetic and Evolutionary Computation Conference (GECCO)**. Junho. p. 750–758, 2021. Disponível em:

<<https://doi.org/10.1145/3449639.3459302>>.

Até o momento da defesa desta dissertação, o seguinte artigo encontrava-se em avaliação:

ALDEIA, G. S. I.; de FRANÇA, F. O. Interpretability in Symbolic Regression: a benchmark of Explanatory Methods using the Feynman data set. **Genetic Programming and Evolvable Machines**. (Submetido para avaliação).

Obrigado!

Avaliação da Interpretabilidade em Regressão Simbólica

GUILHERME SEIDYO IMAI ALDEIA

02 de dezembro de 2021