SymReg
JOSEF RESSEL CENTER FOR SYMBOLIC REGRESSION

https://symreg.at

Genetic Programming and Symbolic Regression

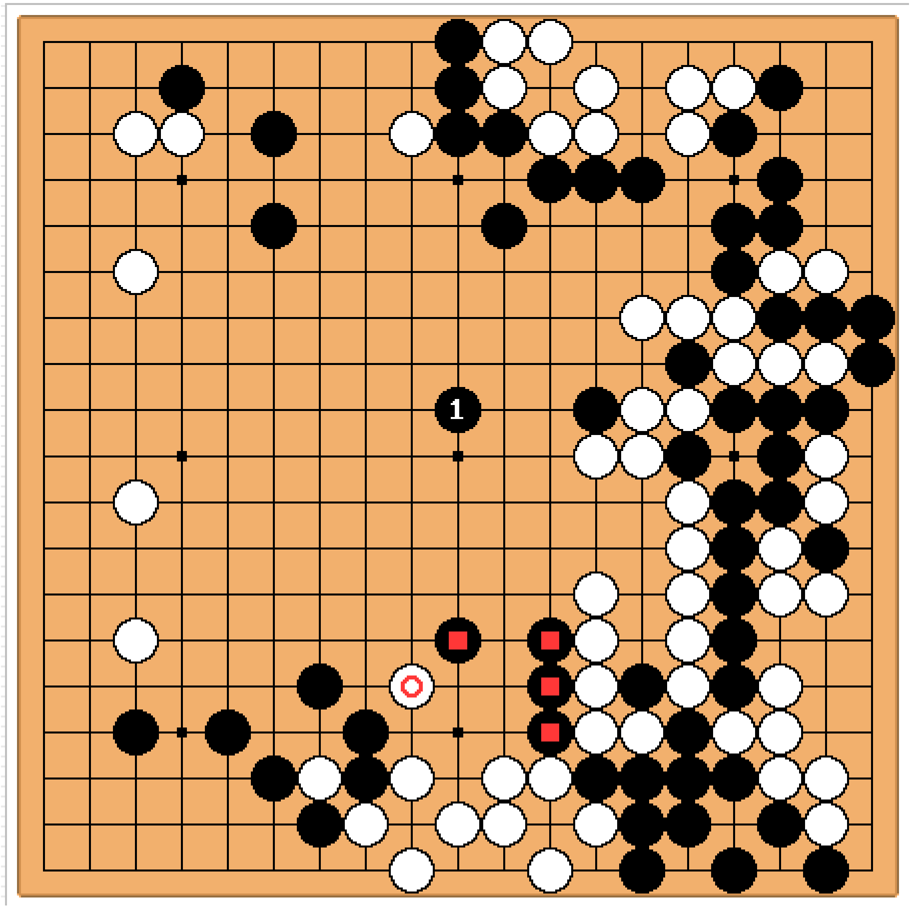
AI Meetup Graz

Gabriel Kronberger, Fachhochschule OÖ, Campus Hagenberg

12. Dezember 2020

FH OBERÖSTERREICH

Explain!

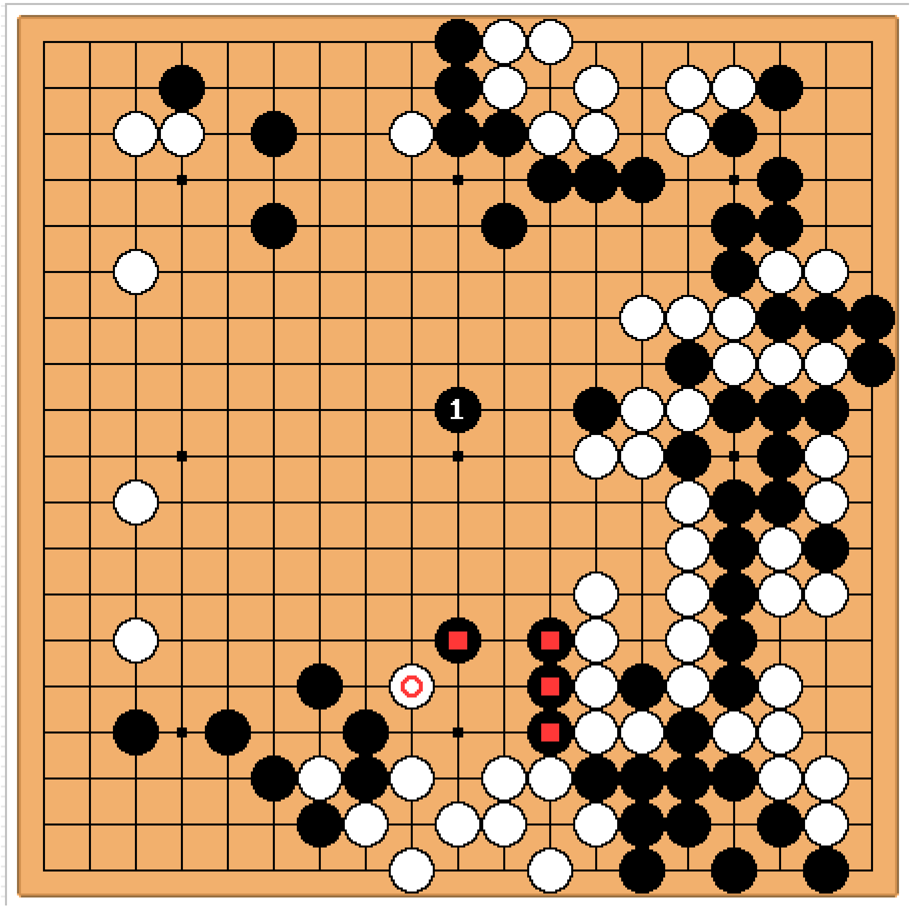


The „ear-reddening move“.

Shusaku (B) vs. Gennan (W)

1846

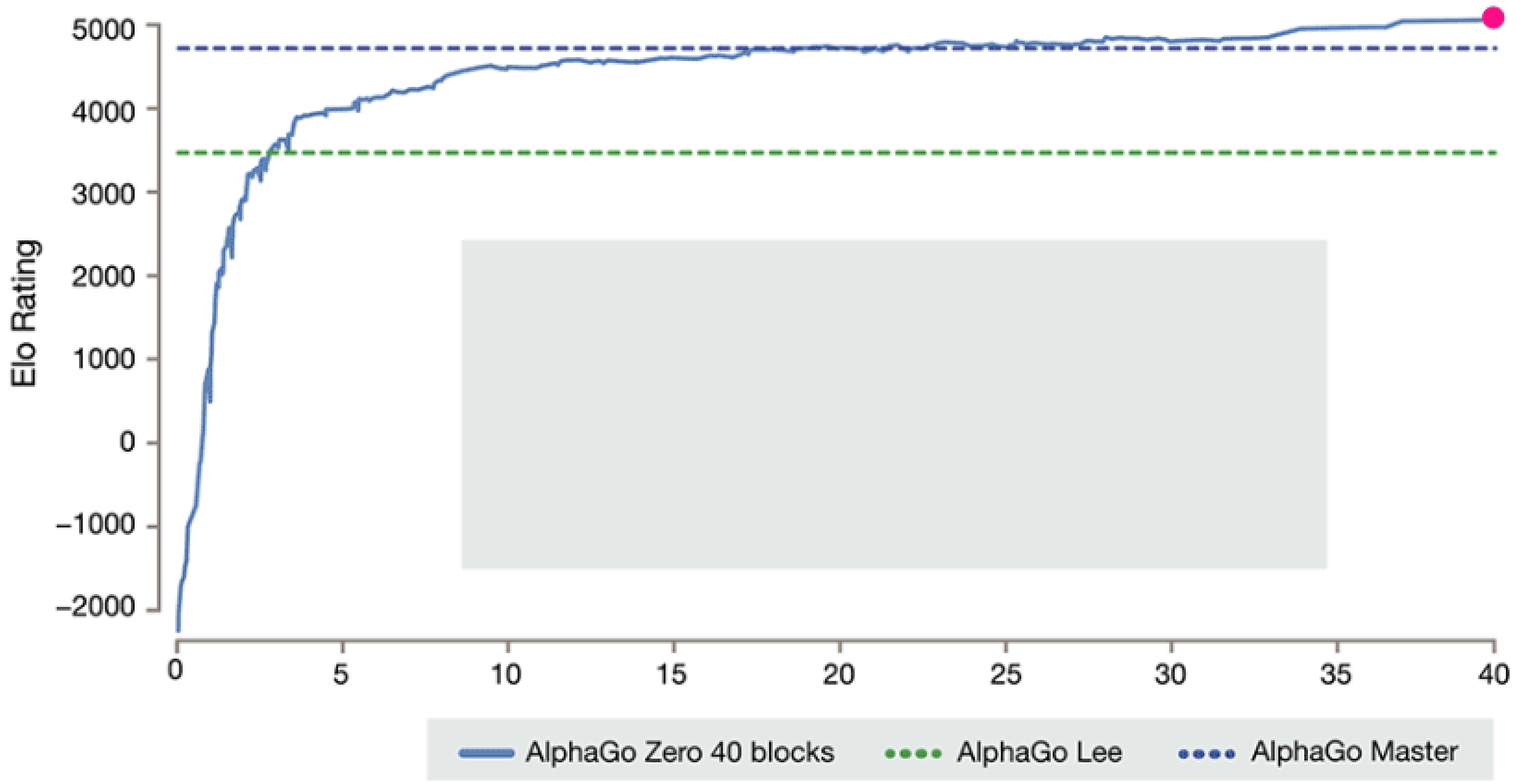
Explain!



B1 has different objectives.

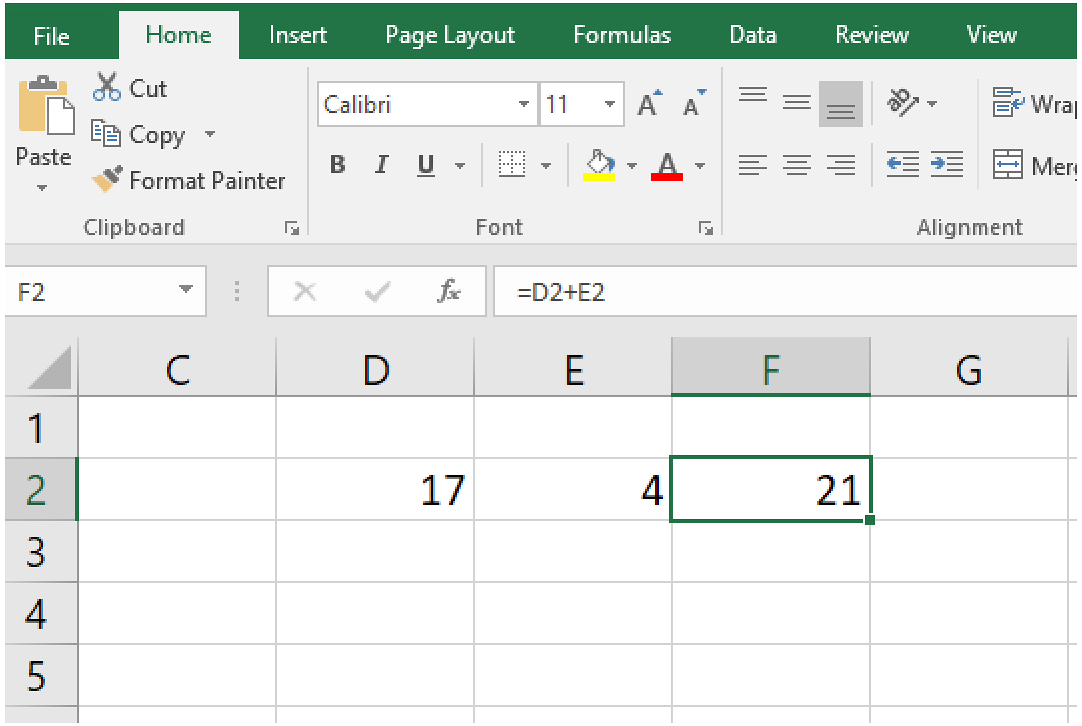
- It expands Black's moyo at the top,
- it helps the four black stones marked,
- it reduces the influence of White's strong position to the right, and it also has an eye on White's moyo on the left side.

Reliability / Trust



Source:
https://deepmind.com/research/alphago/

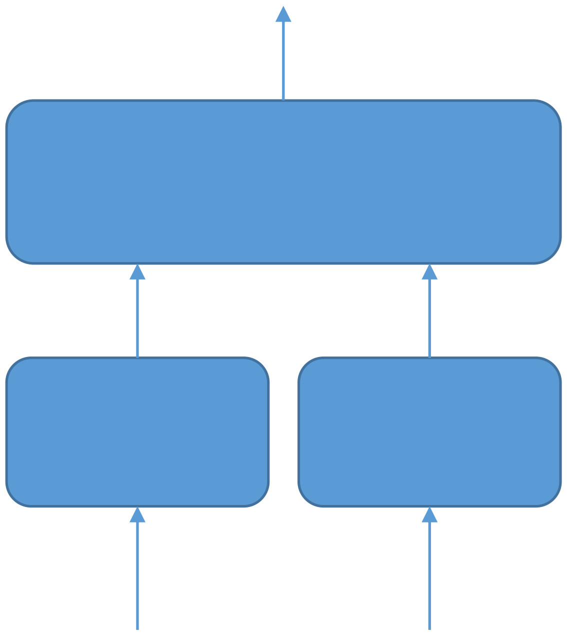
Why do I trust the result?



Btw.: I cannot fully understand what happens to produce 21 even though I studied computer science

Requirements for trust: modularity

- Abstraction of complexity
- Interacting smaller components
- Each component can be trusted
- Communication protocols can be trusted
- Good track record



Automatic programming – State-of-the-art

Example: Excel „Flash Fill“ - Feature

	A	B
1	Email	Column 2
2	Nancy.FreeHafer@fourthcoffee.com	nancy freehafer
3	Andrew.Cencici@northwindtraders.com	andrew cencici
4	Jan.Kotas@litwareinc.com	jan kotas
5	Mariya.Sergienko@gradicdesigninstitute.com	mariya sergienko
6	Steven.Thorpe@northwindtraders.com	steven thorpe
7	Michael.Neipper@northwindtraders.com	michael neipper
8	Robert.Zare@northwindtraders.com	robert zare
9	Laura.Giussani@adventure-works.com	laura giussani
10	Anne.HL@northwindtraders.com	anne hl
11	Alexander.David@contoso.com	alexander david
12	Kim.Shane@northwindtraders.com	kim shane
13	Manish.Chopra@northwindtraders.com	manish chopra
14	Gerwald.Oberleitner@northwindtraders.com	gerwald oberleitner
15	Amr.Zaki@northwindtraders.com	amr zaki
16	Yvonne.McKay@northwindtraders.com	yvonne mckay
17	Amanda.Pinto@northwindtraders.com	amanda pinto

Paper:
Gulwani, S.; José Hernández-Orallo; Kitzelmann, E.; Muggleton, SH.; Schmid, U.; Zorn, B. (2015). Inductive programming meets the real world. Communications of the ACM. 58(11):90-99. doi:10.1145/2736282

https://riunet.upv.es/handle/10251/64984

Automatic programming – State-of-the-art
Solving programming exercises

5. **Double Letters (P 4.1)** Given a string, print the string, doubling every letter character, and tripling every exclamation point. All other non-alphabetic and non-exclamation characters should be printed a single time each.

13. **Vector Average (Q 7.7.11)** Given a vector of floats, return the average of those floats. Results are rounded to 4 decimal places.

14. **Count Odds (Q 7.7.12)** Given a vector of integers, return the number of integers that are odd, without use of a specific `even` or `odd` instruction (but allowing instructions such as `mod` and `quotient`).

Thomas Helmuth, *General Program Synthesis from Examples Using Genetic Programming with Parent Selection Based on Random Lexicographic Orderings of Test Cases,*
University of Massachusetts - Amherst, PhD Thesis, 2015
https://web.cs.umass.edu/publication/docs/2015/UM-CS-PhD-2015-005.pdf

Automatic programming – State-of-the-art

	Lexicase				Tournament			
Problem	**100%**	**75%**	**50%**	**25%**	**100%**	**75%**	**50%**	**25%**
Double Letters	6	1	1	0	0	0	0	0
Replace Space with Newline	51	46	20	24	8	13	11	9
String Lengths Backwards	66	47	17	17	7	6	12	10
Vector Average	16	*33	*49	25	14	11	5	8
Count Odds	8	3	0	1	0	0	0	0
Mirror Image	78	78	67	48	46	41	34	44
X-Word Lines	8	17	4	0	0	0	0	0
Negative To Zero	45	28	19	9	10	5	10	7
Syllables	18	13	10	8	1	2	1	3

Thomas Helmuth, *General Program Synthesis from Examples Using Genetic Programming with Parent Selection Based on Random Lexicographic Orderings of Test Cases*,
University of Massachusetts - Amherst, PhD Thesis, 2015
https://web.cs.umass.edu/publication/docs/2015/UM-CS-PhD-2015-005.pdf

Automatic programming – State-of-the-art

Herbie: Automatically improving floating point accuracy of expressions

$$\frac{1}{2}\sqrt{2\left(\sqrt{x \cdot x + y \cdot y} + x\right)}$$

$$\frac{1}{2}\sqrt{2\frac{y^2}{\sqrt{x \cdot x + y \cdot y} - x}}$$

This improvement was implemented as a patch to Math.js, accepted by the Math.js developers, and released with version 0.27.0 of Math.js [32].

https://herbie.uwplse.org/
P. Panchekha et al.
Automatically improving accuracy for floating point expressions, PLDI '15 Proceedings of the 36th ACM SIGPLAN Conference on Programming Language Design and Implementation
Pages 1-11, 2015

Automatic programming – State-of-the-art

Herbie: Automatically improving floating point accuracy of expressions

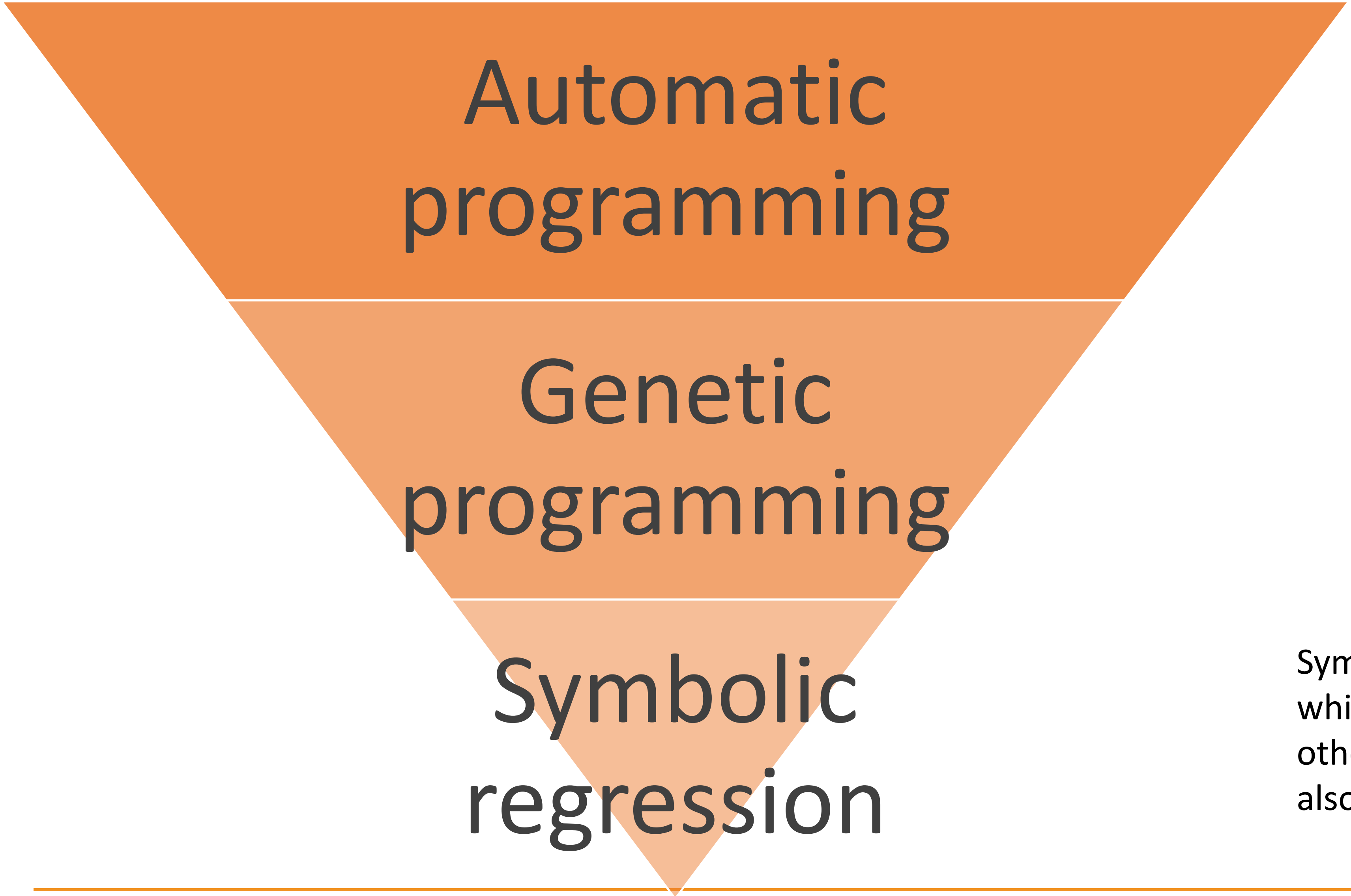
$$\frac{1}{2}(\sin x)(e^{-y} - e^{y})$$

$$-(\sin x)\left(y + \frac{1}{6}y^{3} + \frac{1}{120}y^{5}\right)$$

https://herbie.uwplse.org/
P. Panchekha et al.
Automatically improving accuracy for floating point expressions, PLDI '15 Proceedings of the 36th ACM SIGPLAN Conference on Programming Language Design and Implementation
Pages 1-11, 2015

SymReg

JOSEF RESSEL CENTER FOR SYMBOLIC REGRESSION

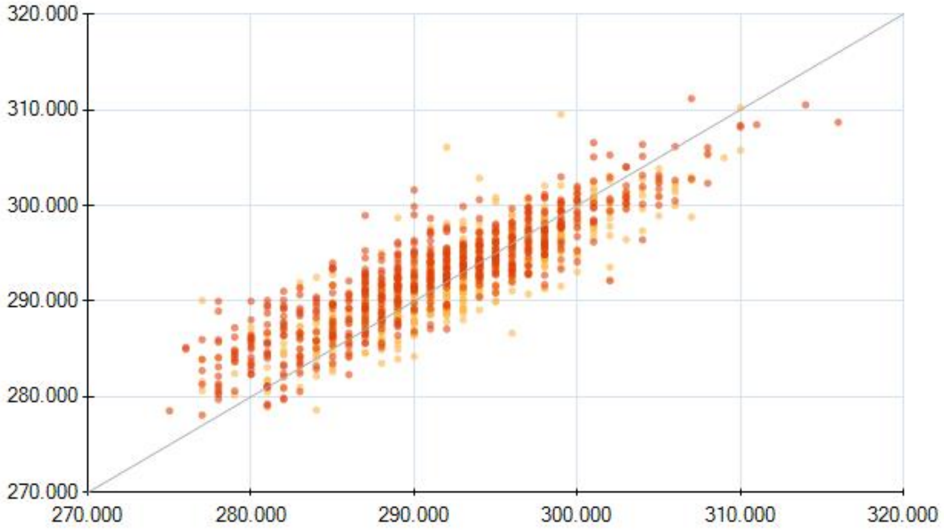


SymReg can be solved using GP which is a form of AP. However, other solution methods are also possible (see below)

Symbolic regression

Learning of models as mathematical expressions

$$f(x_1, x_2) = \frac{0.0651\, x_2 + 1.316}{1.5156\, x_1 + 17.619}$$

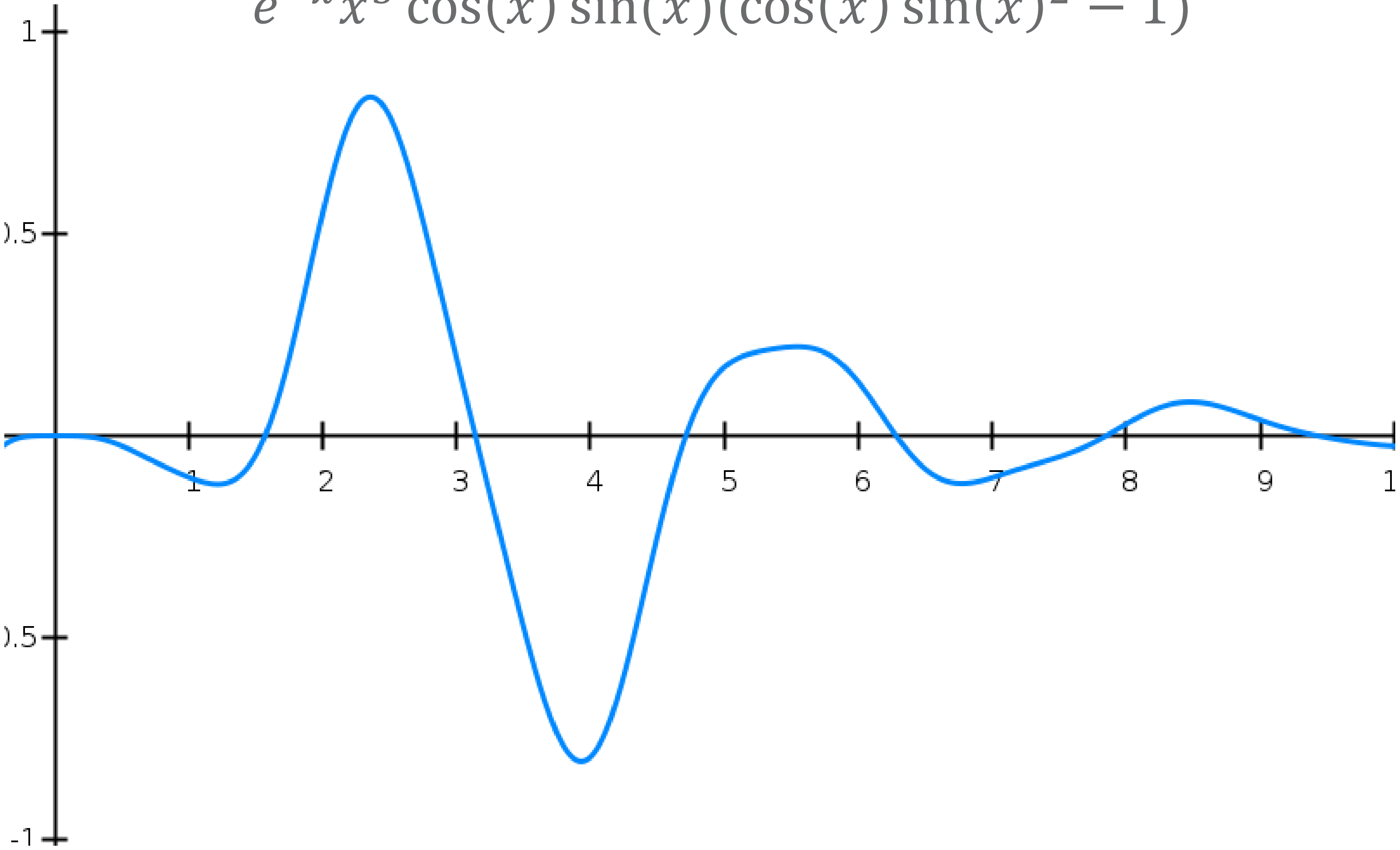


Properties

- Nonlinear Models
- Smooth Response Functions
- Integration of Prior Knowledge

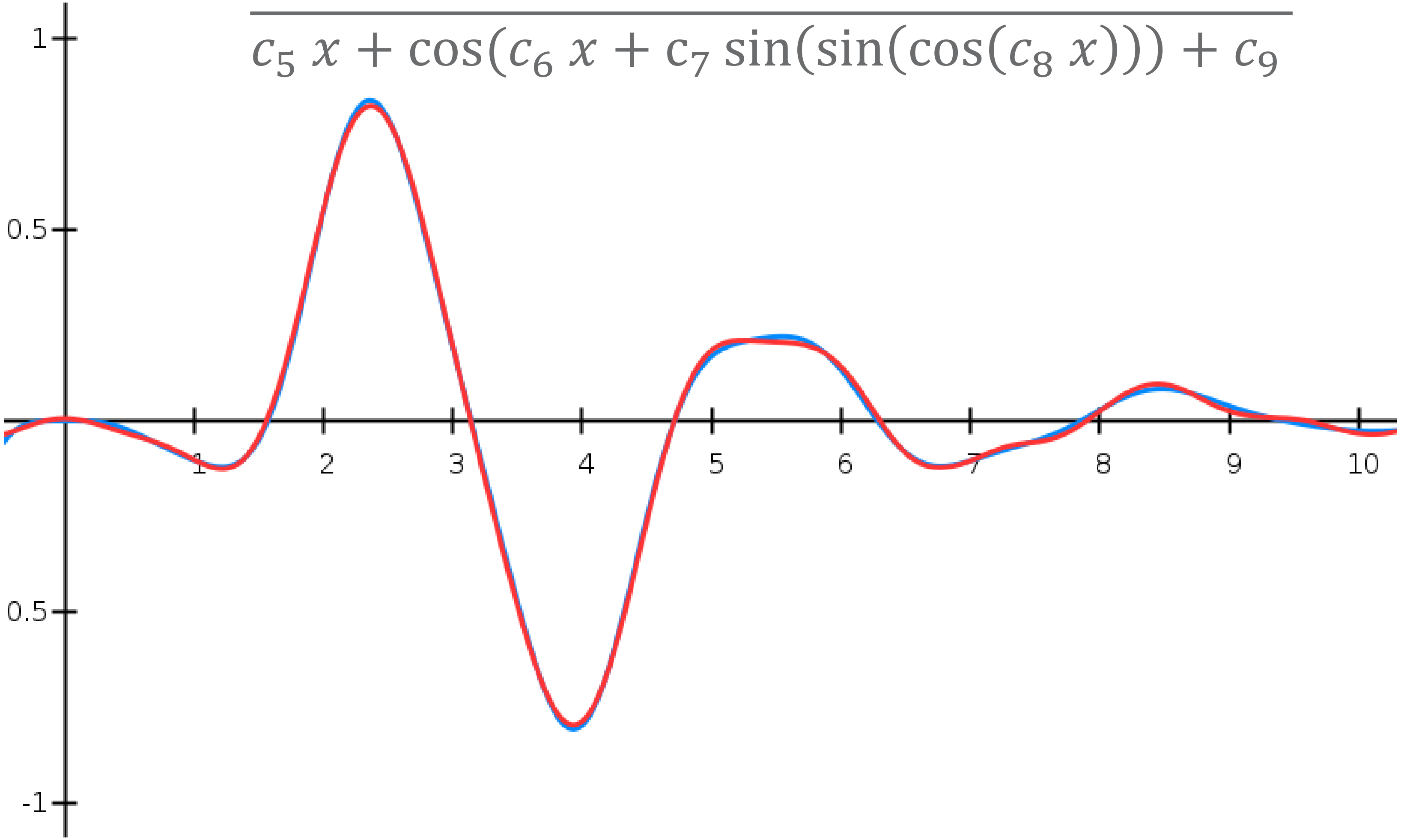
Symbolic regression

$$e^{-x}x^3\cos(x)\sin(x)(\cos(x)\sin(x)^2-1)$$

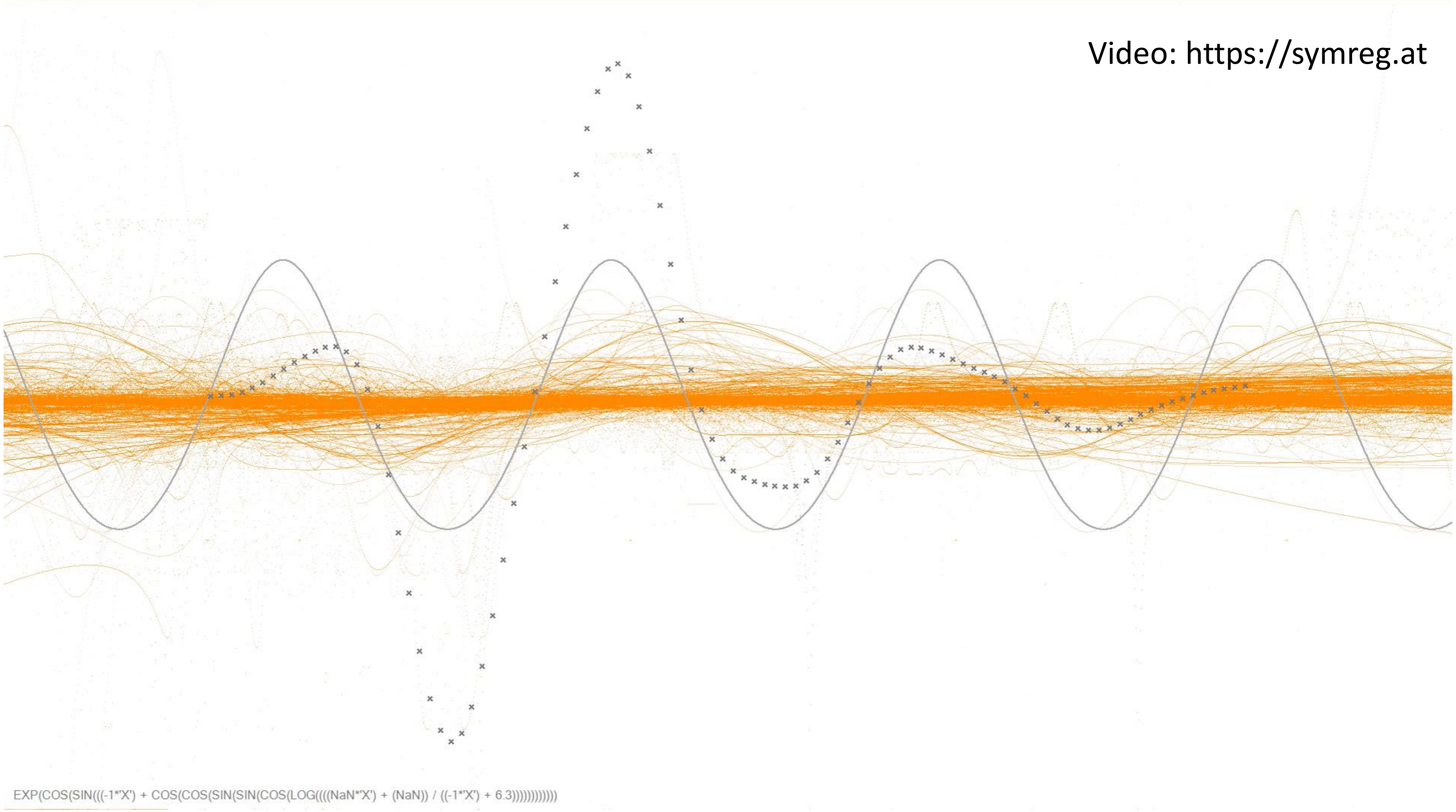


Symbolic regression

$$\frac{c_0 \sin(c_1 \cos(c_2\, x) + c_3 \cos(c_4\, x)}{c_5\, x + \cos(c_6\, x + c_7 \sin(\sin(\cos(c_8\, x))) + c_9}$$



Video: https://symreg.at



https://symreg.at

Symbolic regression

Pros

- Analytical model
- Fast evaluation
- Implicit feature selection
- No assumption about the model structure
- Simple integration in other software

Cons

- Computationally expensive
- Algorithm is hard to configure
- Bloated models
- Non-deterministic

Symbolic regression using genetic programming

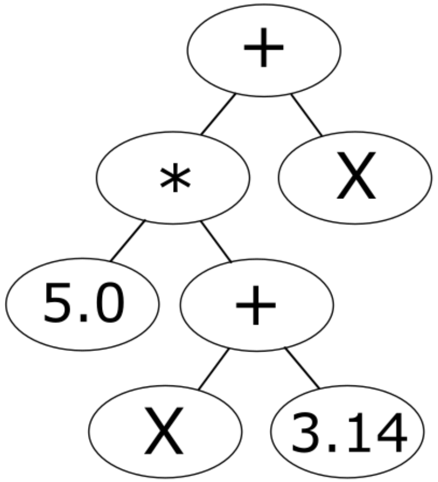
Symbolic expression trees

- Encode regression models
- Easily manipulated

Objective function

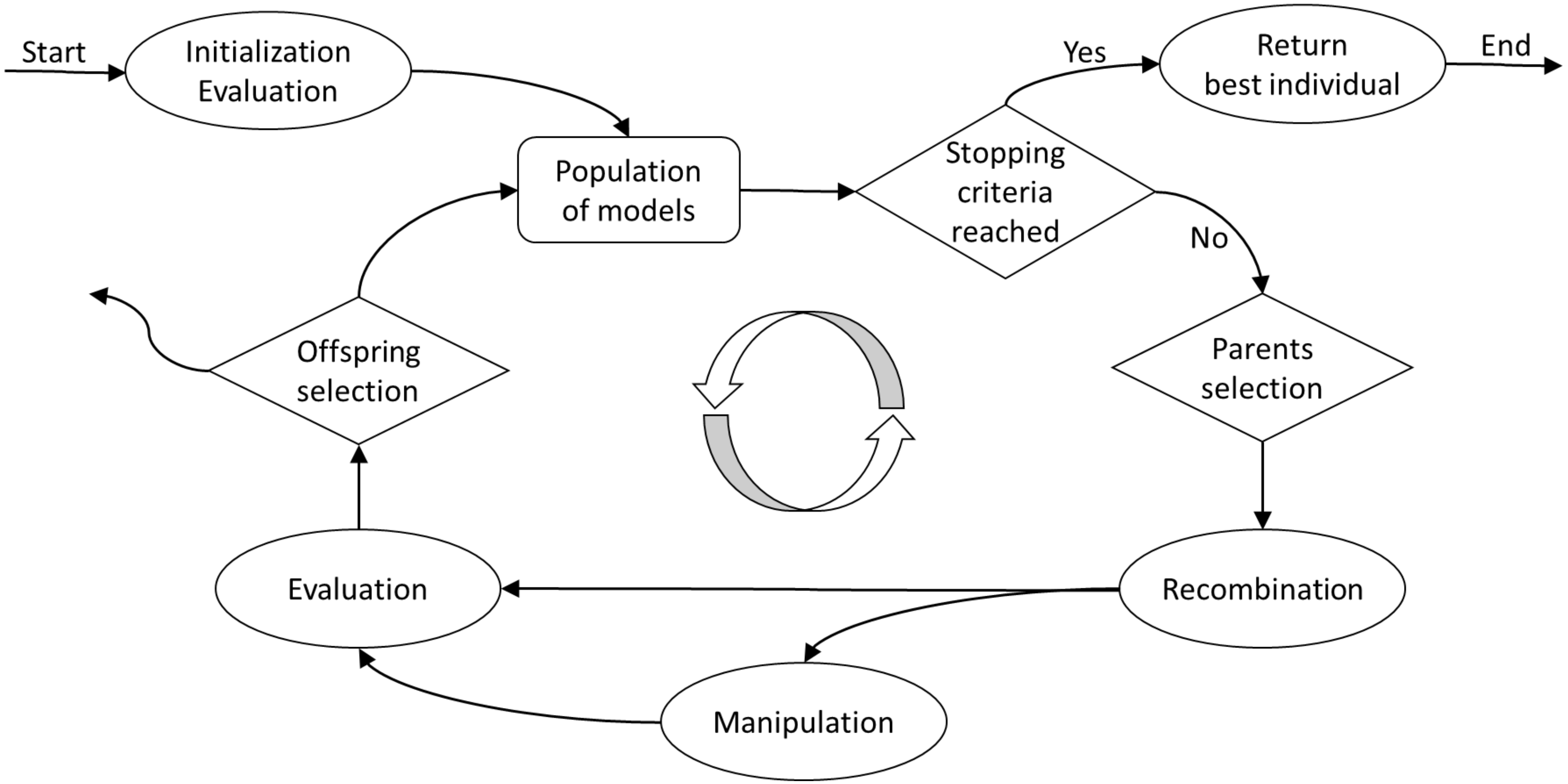
- Minimize error between model estimations and presented data

$$y = f(x) + \epsilon$$

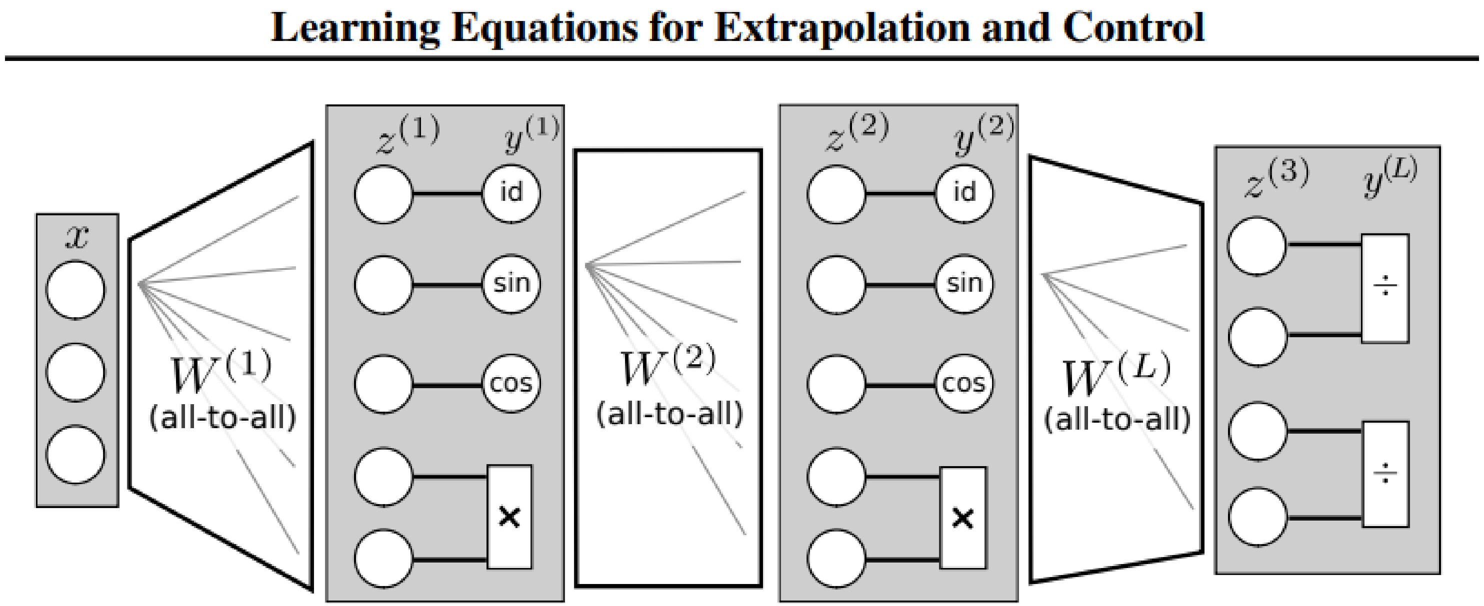


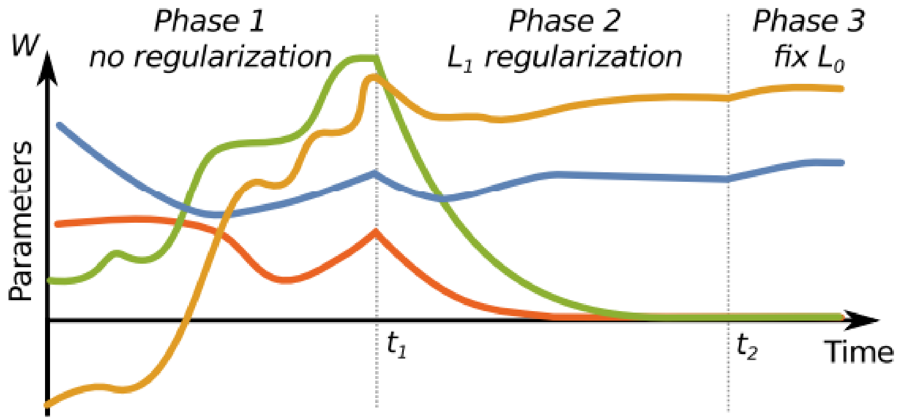
$$f(x) = 5 * (x + 3.14) + x$$

Genetic programming



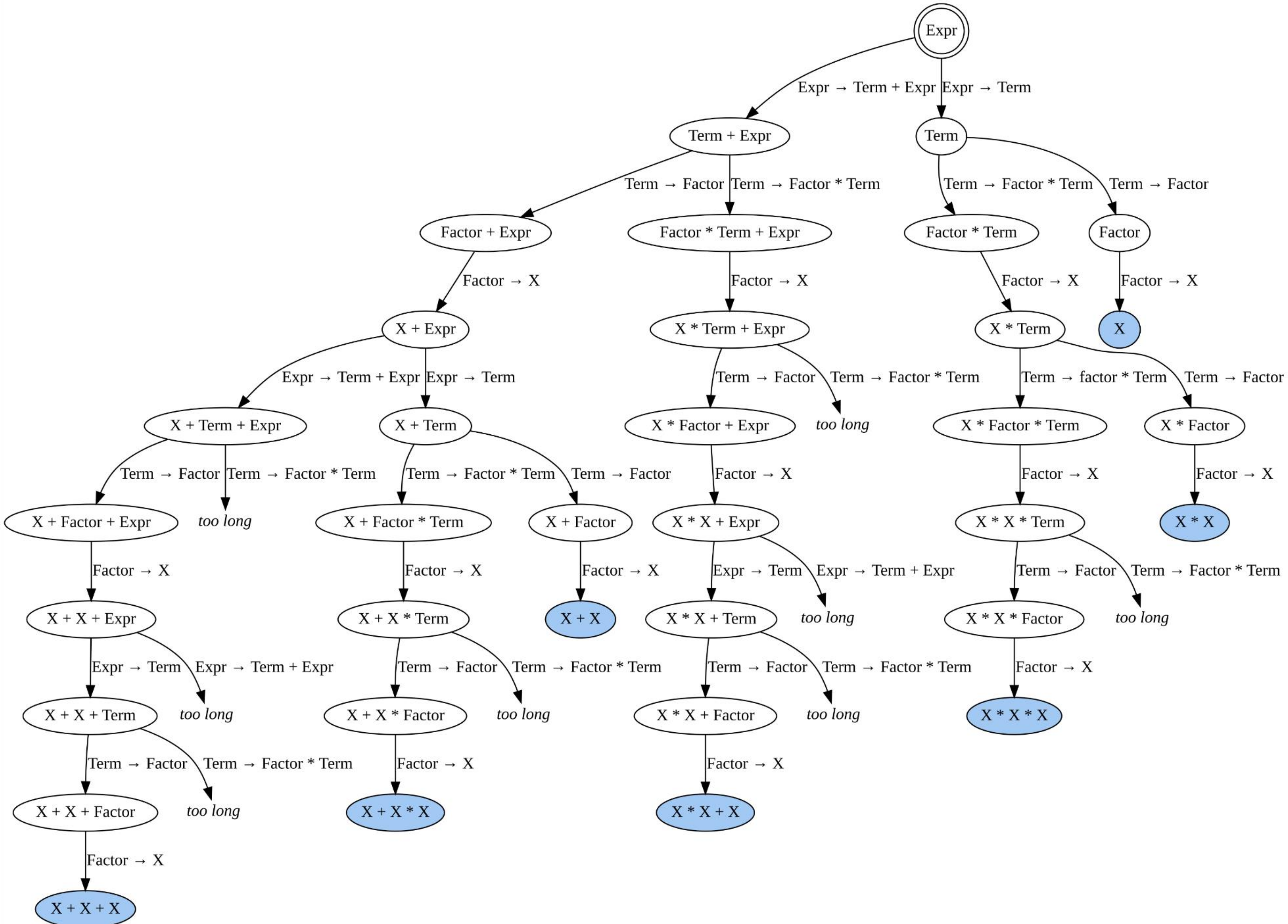
Symbolic regression with neural networks





Sahoo, S. S., Lampert, C. H., and Martius, G. Learning equations for extrapolation and control. Proceedings of the 35 th International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

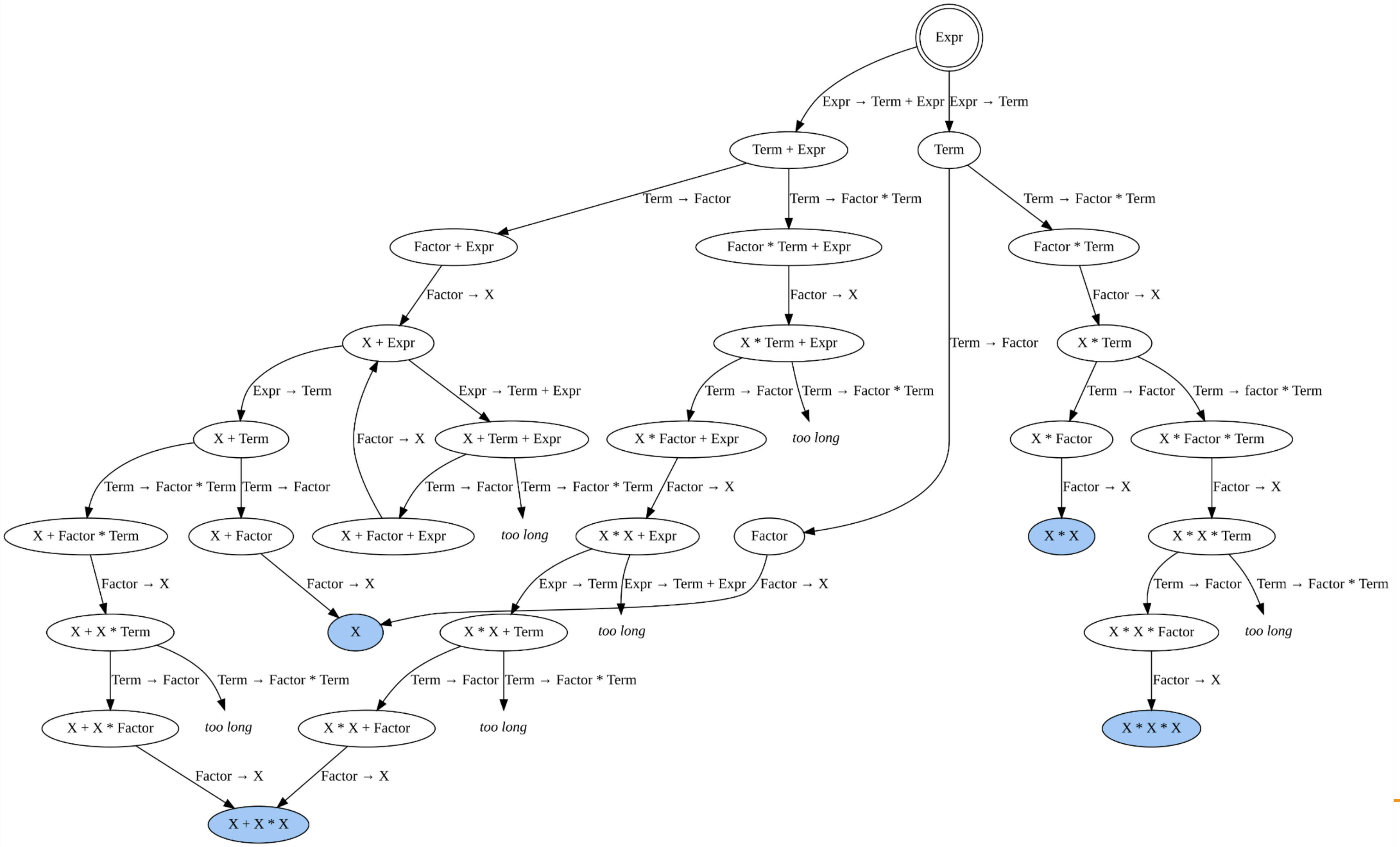
Symbolic regression as a graph search problem




```
G(Expr):
Expr → Term |
       Term  + Expr
Term → Fact  |
        Fact * Term
Fact → X
```

L. Kammerer, G. Kronberger, B. Burlacu, S. Winkler, M. Kommenda, M. Affenzeller, *Symbolic Regression by Exhaustive Search,* In Genetic Programming in Theory and Practice Springer, 2019

Search space reduction through deduplication



SymReg

Symbolic regression algorithms compared to state-of-the-art algorithms

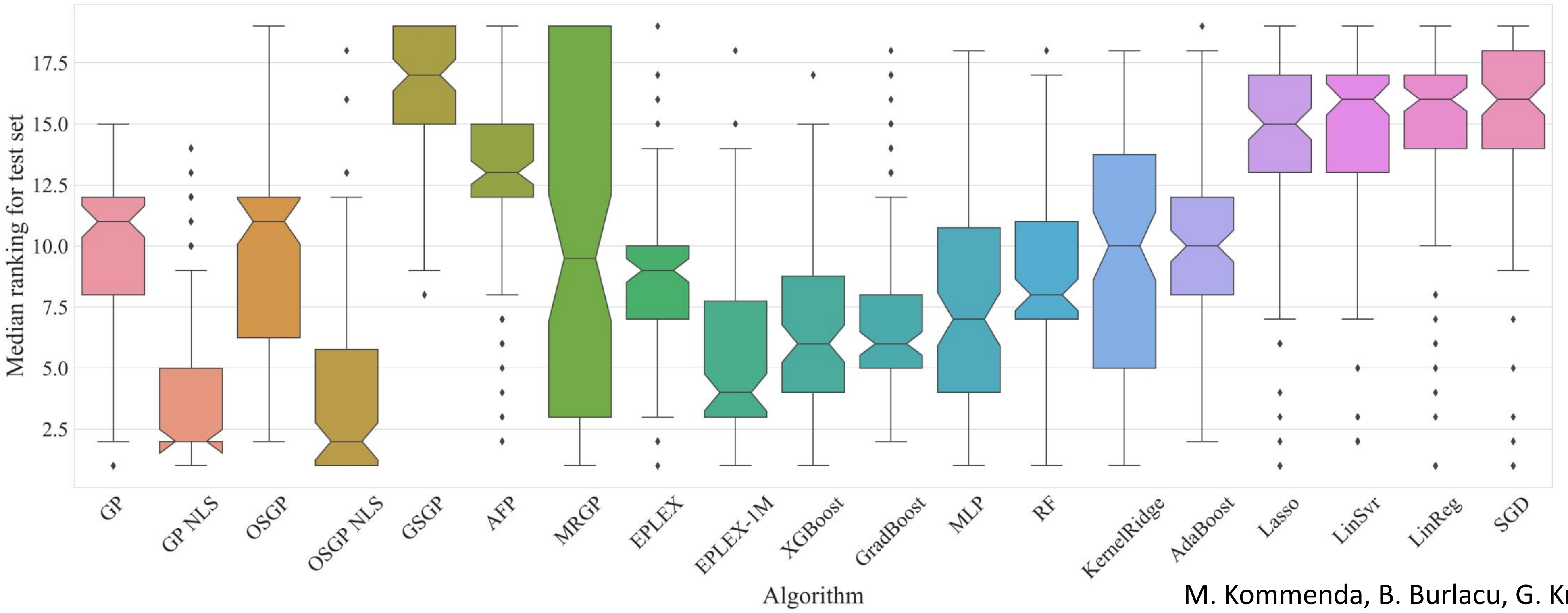
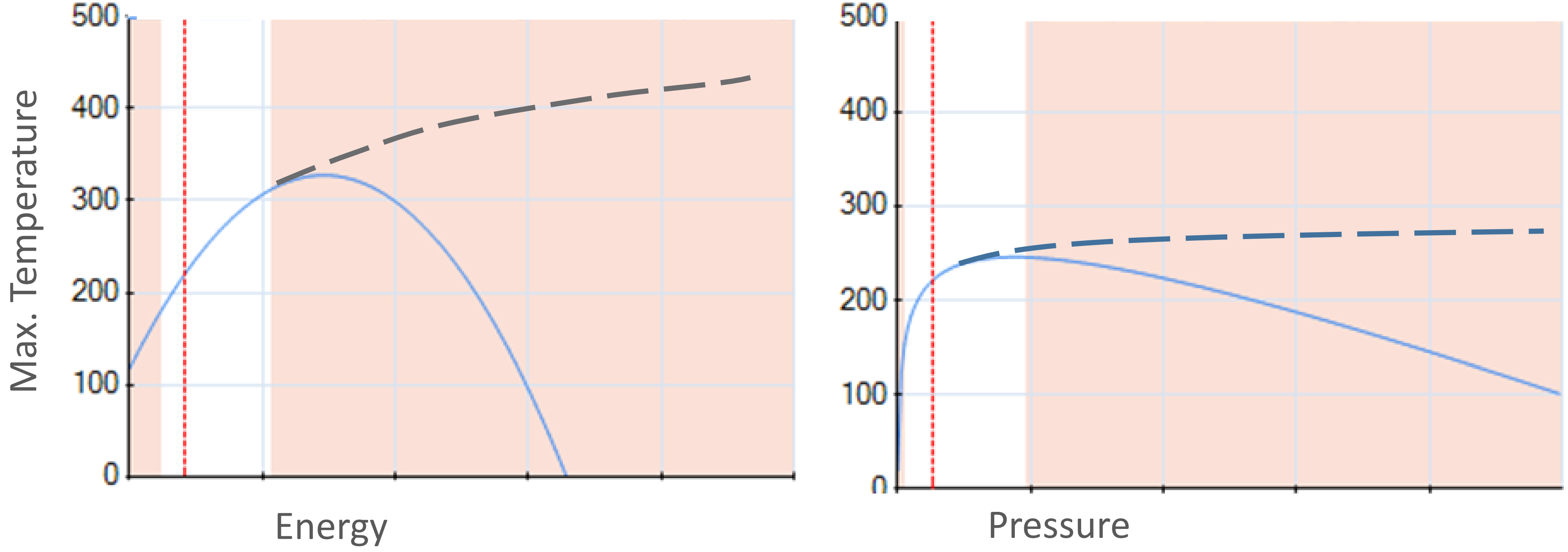


Fig. 7 Algorithm ranking based on the MSE scores on the test set.

M. Kommenda, B. Burlacu, G. Kronberger, M. Affenzeller, *Integrating Numerical Optimization Methods with Genetic Programming,* Genetic Programming and Evolvable Machines, to appear 2020

Knowledge integration

How can we enforce monotonicity?



SymReg

The concept of shape-constrained regression

$$f^* = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \mathcal{L}(f, X, \boldsymbol{y})$$

$s.t.$:

$$\begin{matrix} l_f & & f(x_f) & & u_f \\ l_{Jac} & \leq & \nabla f(x_{Jac}) & \leq & u_{Jac} \\ l_{Hess} & & \nabla^2 f(x_{Hess}) & & u_{Hess} \end{matrix}$$

$$\forall x_f, x_{Jac}, x_{Hess} \in \mathbb{R}^d,$$

$$\begin{matrix} l_{x_f} & & x_f & & u_{x_f} \\ l_{x_{Jac}} & \leq & x_{Jac} & \leq & u_{x_{Jac}} \\ l_{x_{Hess}} & & x_{Hess} & & u_{x_{Hess}} \end{matrix}$$

$\mathcal{L}(f, X, y)$ is the loss function (e.g. sum of squared errors)

$\mathcal{F}$ is a model class e.g.:

- polynomials of given degree
- neural network architecture

∇f is the vector of partial derivatives of f over all inputs

f must be differentiable

SymReg

SymReg
JOSEF RESSEL CENTER FOR
SYMBOLIC REGRESSION

https://symreg.at

Symbolic Regression

Gabriel Kronberger, Fachhochschule OÖ, Campus Hagenberg
28. November 2019