UNIVERSITY
OF TRIESTE

Improving Geometric Semantic Genetic Programming with Gradient Descent

Speaker : *Gloria Pietropolli*

Mail : *gloria.pietropolli@phd.units.it*

OUTLINE

GENETIC PROGRAMMING

GEOMETRIC **SEMANTIC GENETIC PROGRAMMING**

GEOMETRIC **SEMANTIC GENETIC PROGRAMMING** HYBRIDIZED WITH **GRADIENT DESCENT**

GENETIC PROGRAMMING

GP is a technique to **stochastically evolve a population** (multiset) of individuals **encoding computer programs.**

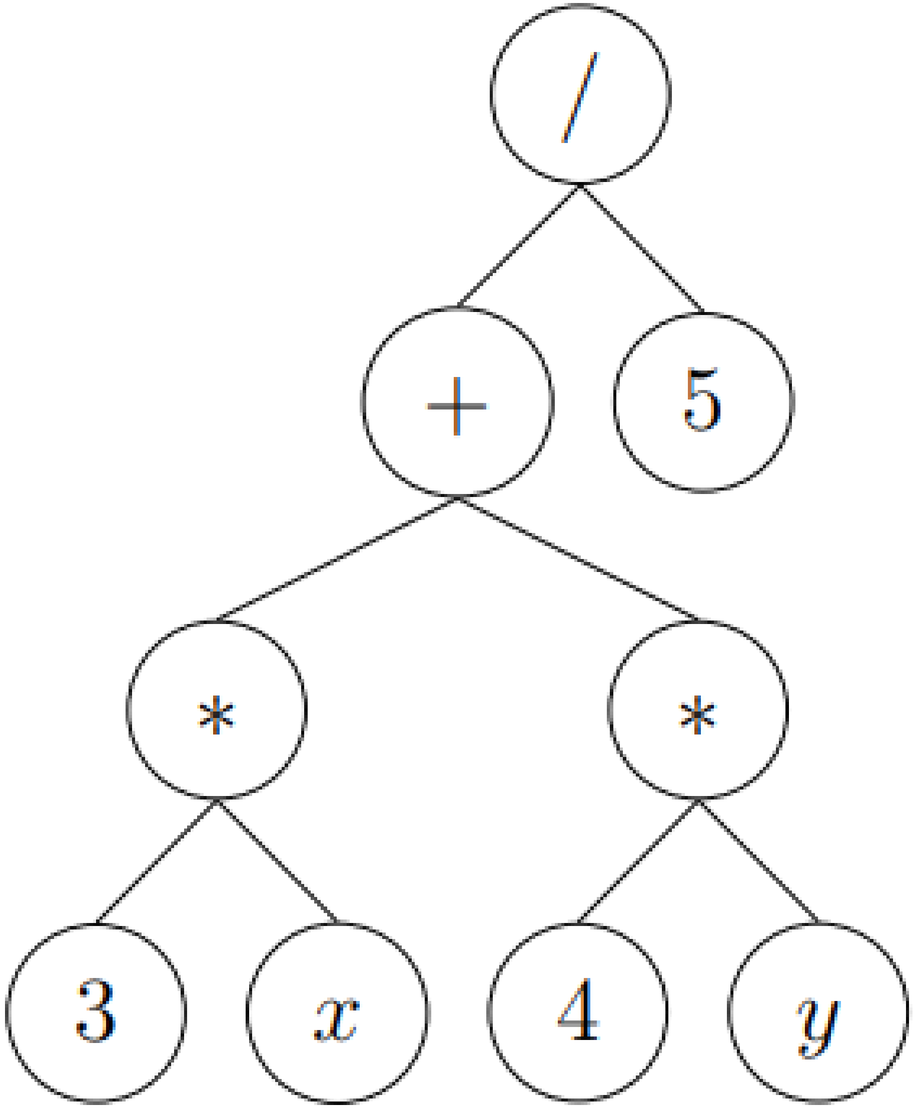
<u>**GOAL**</u> : use stochastic methods to search for and optimize small *computer programs:*

- Finding team soccer robot behaviour
- Fitting arbitrary mathematical equations
- Finding finite-state automata which match a given language

PROGRAM REPRESENATION

As computer programs are variable in size, also the representations used must be variable in size:

- **TREE**
- LIST



$$\boxed{\frac{3x+4y}{5}}$$

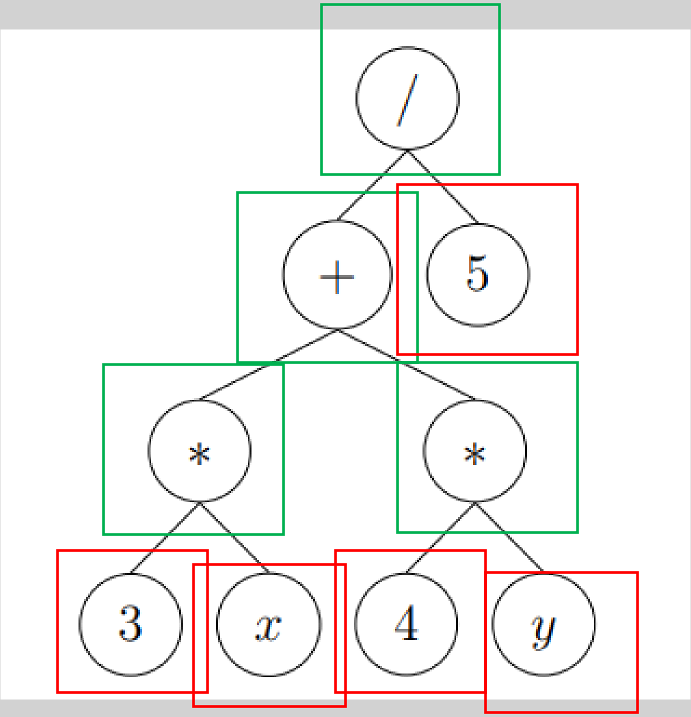
TERMINAL AND FUNCTION SET

The TERMINAL SET contains all the possible **leaves**, for example:

- Constants :
  {0,1,2,3,4,…}
  {true, false}
  {e, π, − 1,…}
- Input Variables : {x0, x1, …}

The FUNCTION SET contains the possible **inner nodes**, for example:

- Arithmetical operations : { + , − , × , ÷ }
- Trigonometric functions : {sin, cos, tan}
- Boolean operators : { ∧ , ∨ , ¬}
- Choice/conditional : {if … then … else …}



SUFFICIENCY

To find a solution we must be able to **represent it**, which means that the primitives must be sufficient to write a solution

For example: with real constants, variables, and {+, -, *,} we can represent any polynomial…

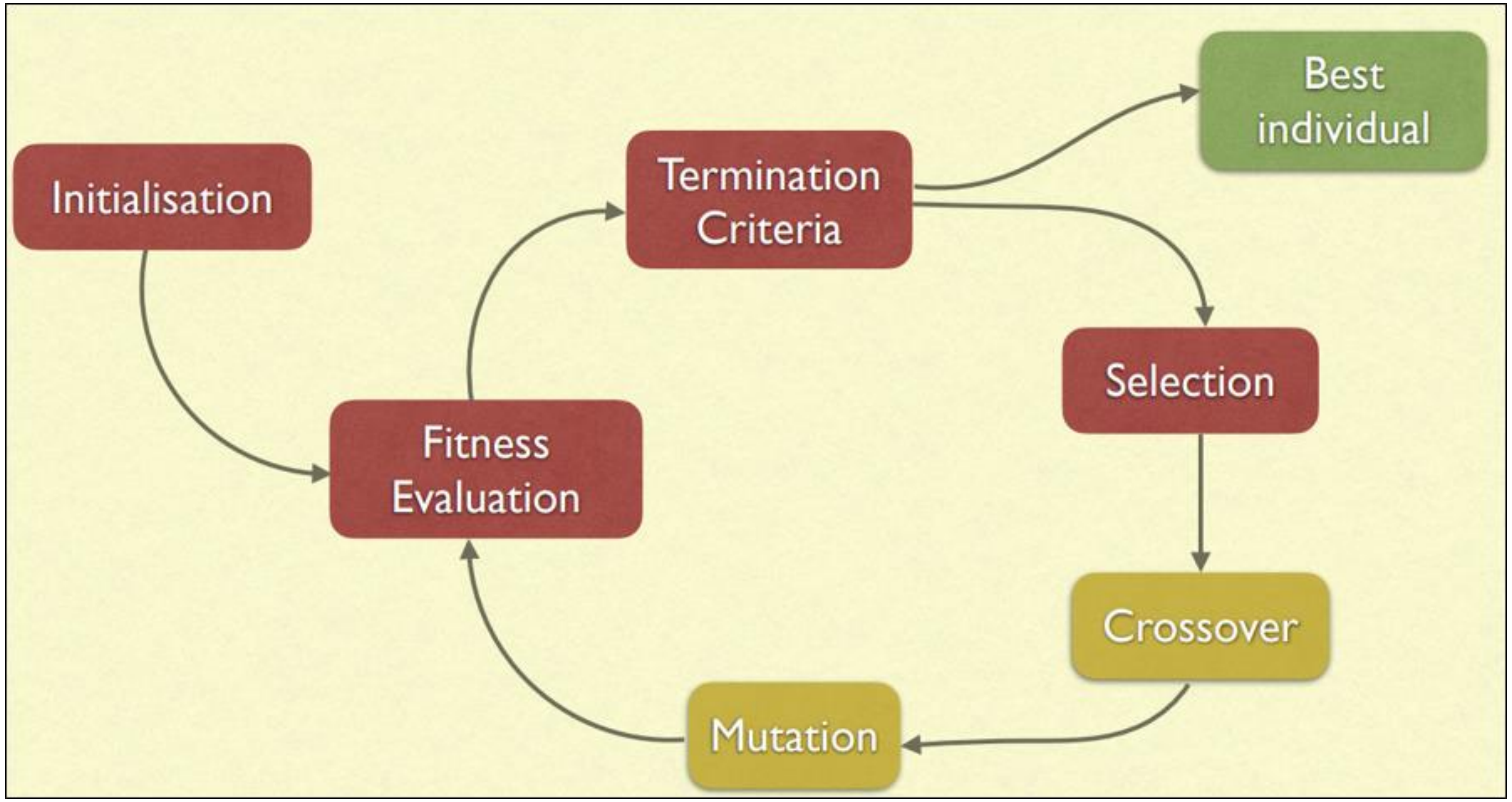
…which is not very useful if the function that we must fit is an exponential

Usually **we cannot assure sufficiency**, but we might still obtain solutions that are good approximations

GENETIC PROGRAMMING DICTIONARY

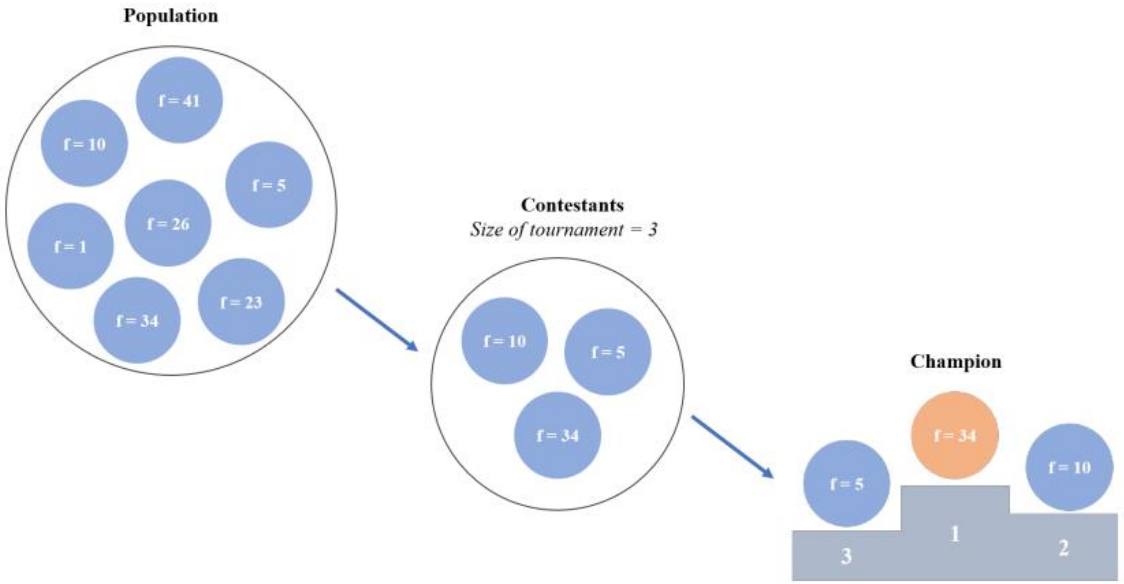
INDIVIDUAL	A candidate solution
CHILD AND PARENT	A child is the tweaked copy of a candidate solution (its parent)
POPULATION	Set of candidate solution
FITNESS	Quality
GENOTYPE/GENOME	Individual's data structure
CHROMOSOME	Genotype in the form of a fixed-lenght vector
GENE	Particular slot position in a chromosome
ALLELE	Particular setting of gene
GENERATION	One cycle of fitness assessment, breeding and population reassembly

EVOLUTION CYCLE



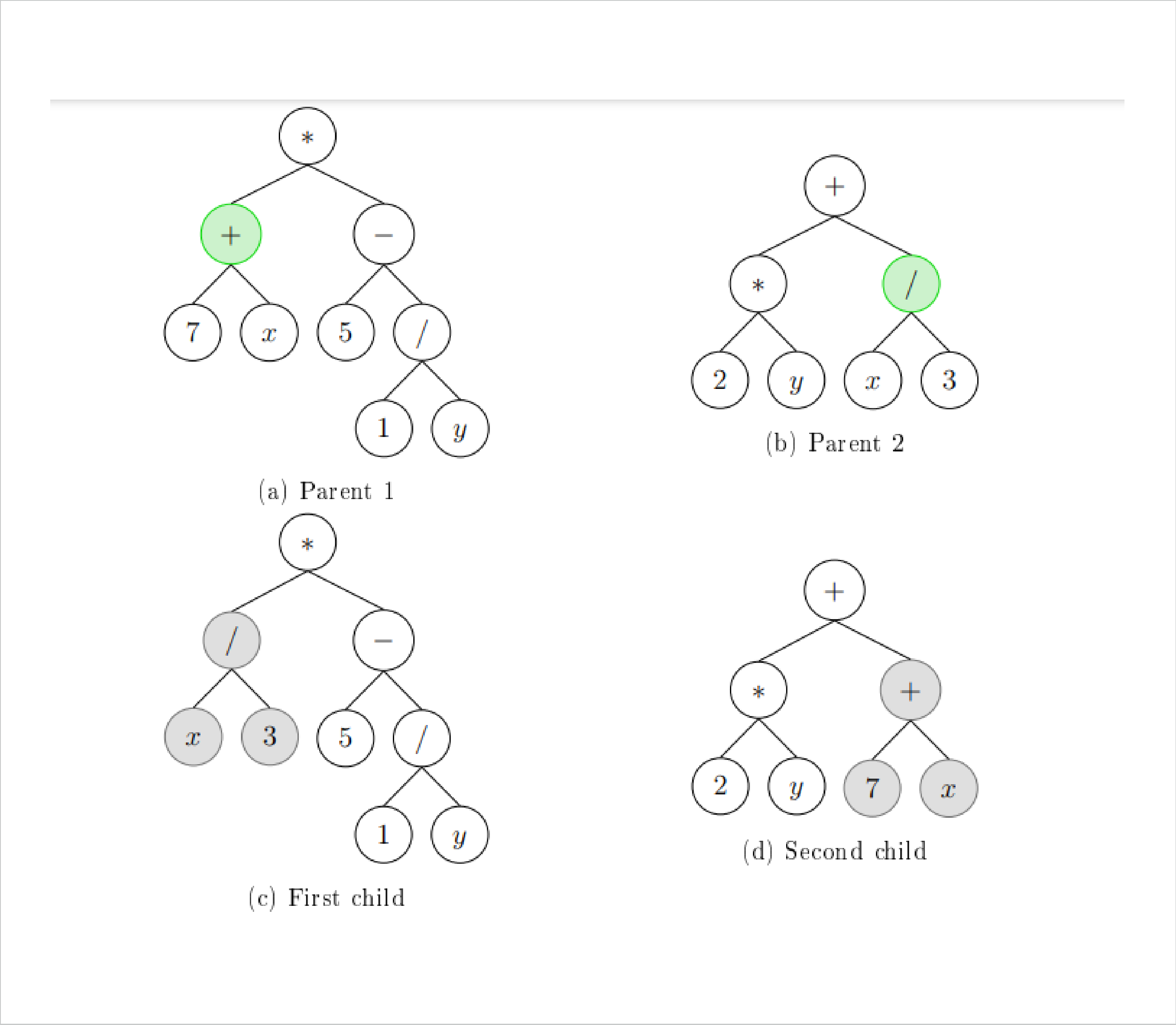
SELECTION

Picking individuals based on their fitness.



CROSSOVER

Tweak which takes two parents, swap section of them and produces (usually 2) children.

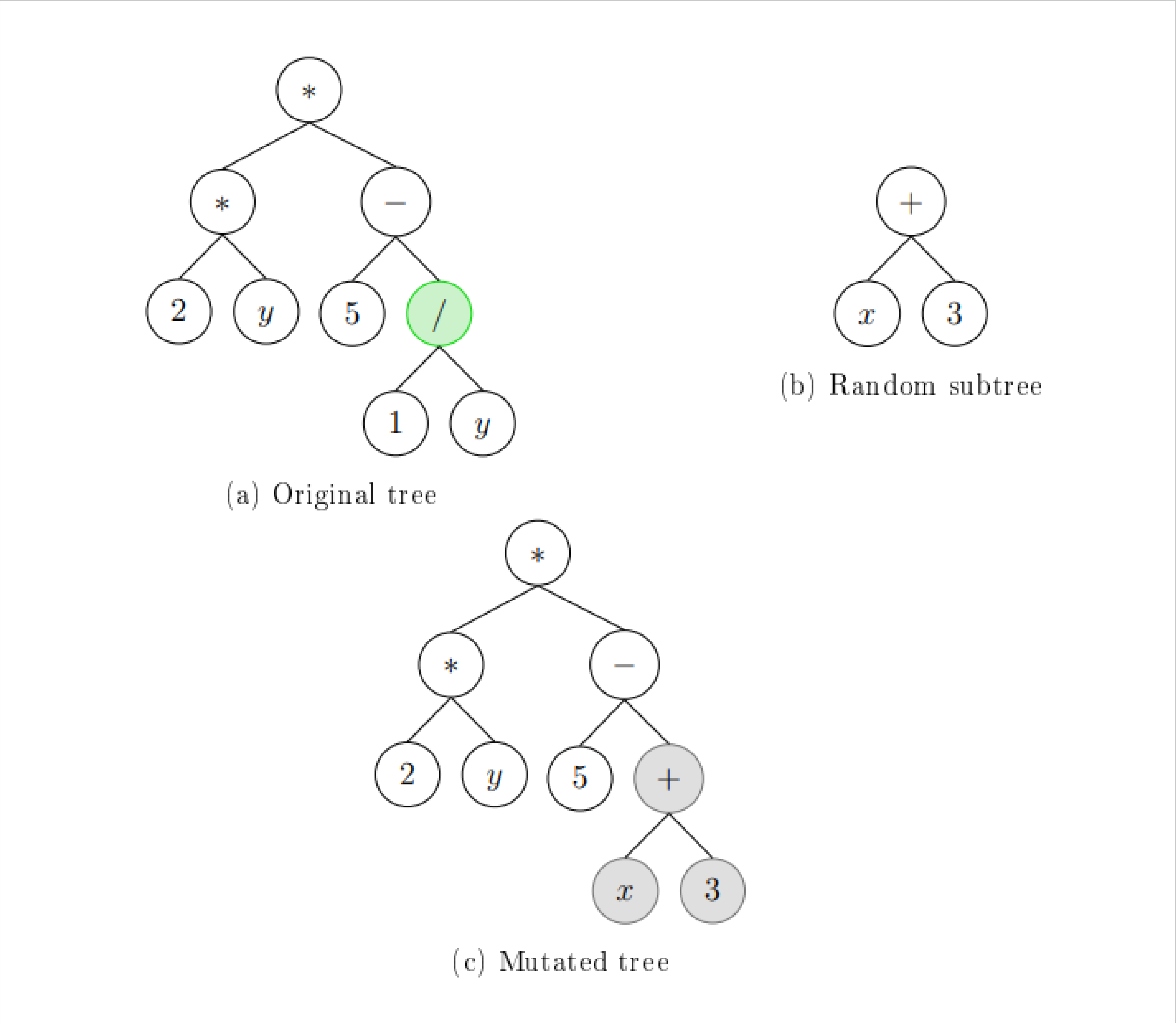


MUTATION

1) Replacement of a **randomly selected subtree** with a **new random subtree**.

2) Replacement of a **randomly selected node** with a **compatible randomly selected node**.

3) Replacement of the **entire tree** with **one of its subtree**.



WAIT.. WHY GEOMETRIC?

WAIT.. WHY SEMANTIC?

$\boxed{\text{GEOMETRIC}}$ $\boxed{\text{SEMANTIC}}$ GENETIC PROGRAMMING

WHY SEMANTIC?

	<u>SYNTAX</u> = STRUCTURE	<u>SEMANTIC</u> = MEANING
GENERAL DEFINITION	Set of rules, principles and processes th at govern the structure of sentences.	Meaning of the sentence.
GP	**Syntactically well-formed individuals are guarantees.** (As programs are represented as syntax trees)	• Fitness of the program • **Set outputs values on input training data**

TRADITIONAL <u>GENETIC PROGRAMMING</u> **IGNORES THE MEANING OF PROGRAMS**

↓

Crossover and **Mutation** operators **act on their syntactical representation**, regardless of their semantic.

WHY GEOMETRIC?

The **semantic** of a solution can be identified by the **vector** of its output values calculated on the training data.

⇩

GP individual can be represented as a **point** in a **real finite-dimension vector space**, the so-called **Semantic Space**.

GEOMETRIC SEMANTIC GENETIC PROGRAMMING

Geometric Semantic Genetic Programming (GSGP) is an evolutionary technique originating from GP that directly searches the semantic space of the programs.

GSGP has been introduced toghether with the definition of the correspondent Geometric Semantic Operators (GSOs).

→

Perform search directly in the **Semantic Space**.

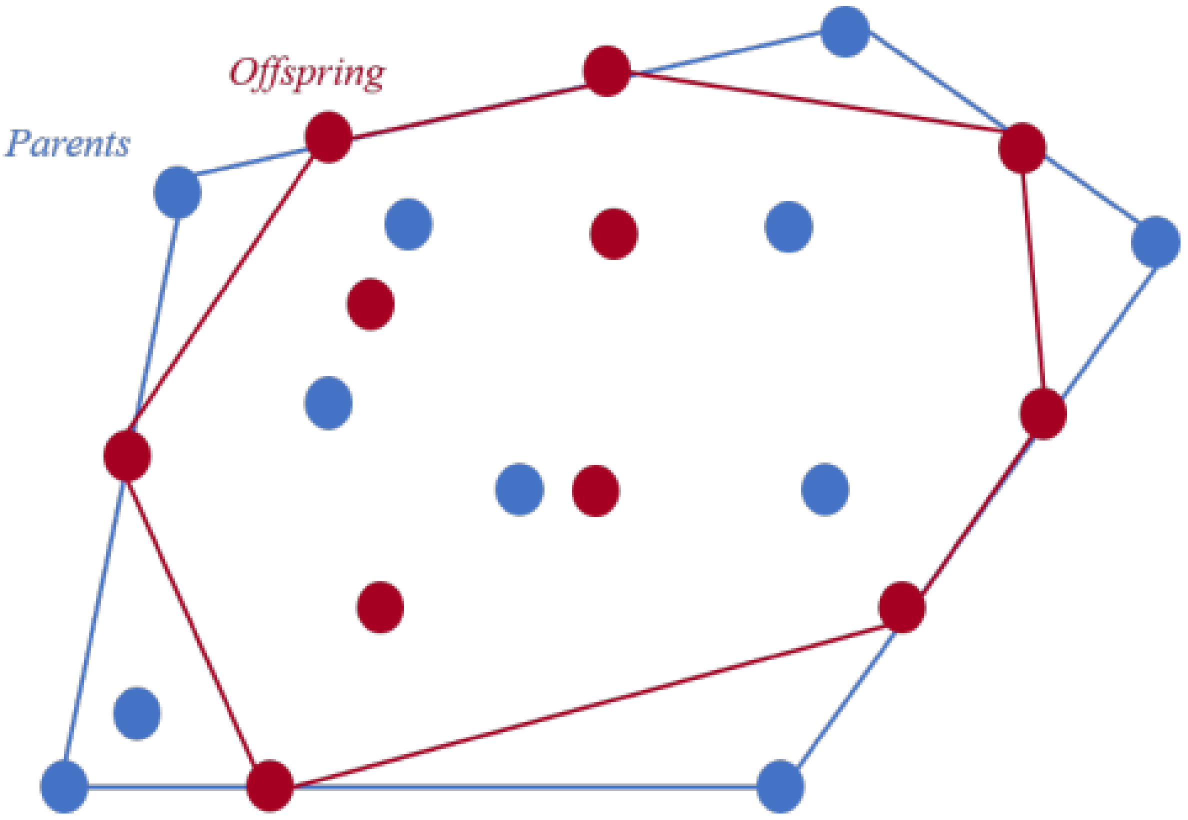
Geometric Crossover

A search operator $\mathcal{C}X : S \times S \rightarrow S$ is a *Geometric Crossover* w.r.t. the metric d if for any choice of parents p_1 and p_2 any of their offspring $o = \mathcal{C}X(p_1, p_2)$ is in the segment between parents.

Geometric Semantic Crossover

Given two parent functions T_1 and $T_2 : \mathrm{R}^n \rightarrow \mathrm{R}$ the *Geometric Semantic Crossover* returns the real function $T_{X0} = (T_1 \times T_R) + ((1 - T_R) \times T_2)$ where T_R is a random real function whose output range in the interval $[0, 1]$.

An important consequence is that crossover is limited by the **convex hull** of the initial population.



Geometric Mutation

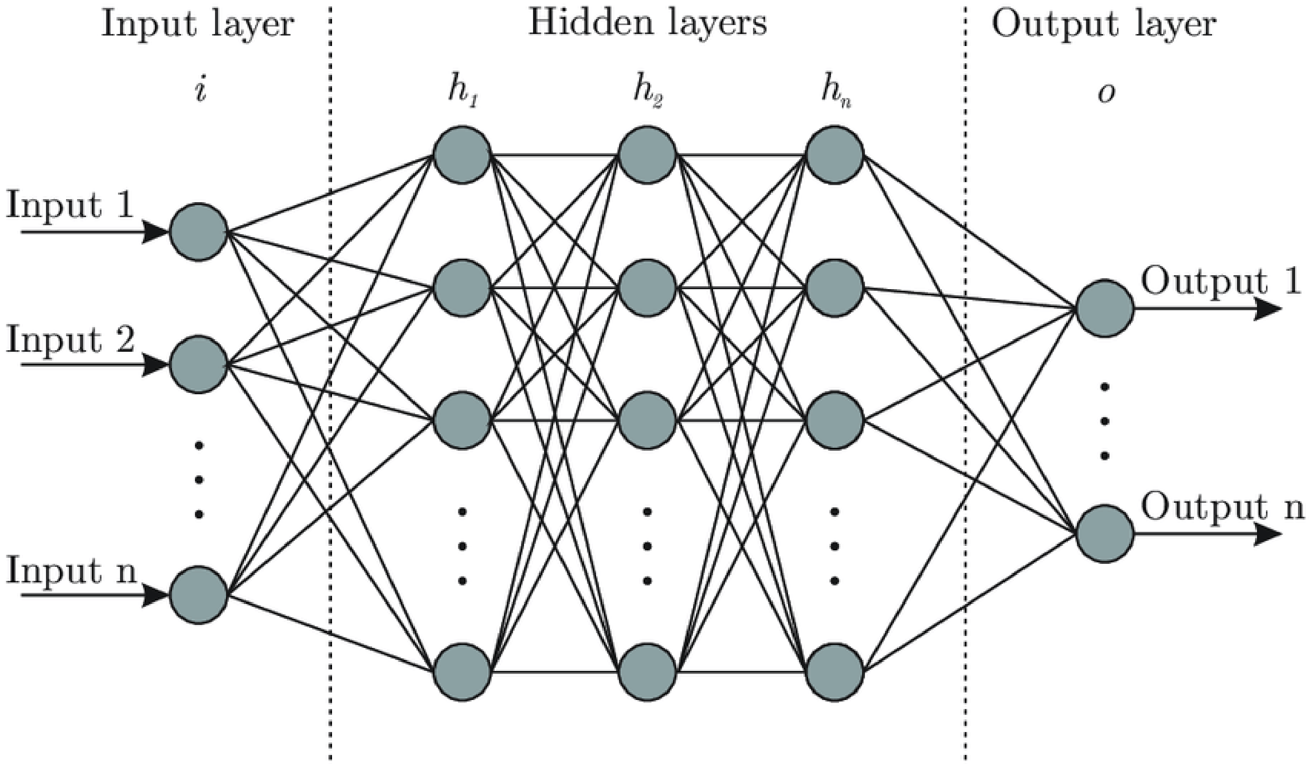
A search operator $M : S \times S \to S$ is a ϵ-*Geometric Mutation* w.r.t. the metric d if for any choice of parent p, any of its offspring $o = M(p)$ is in the metric ball of radius ϵ centered in the parent.

Geometric Semantic Mutation

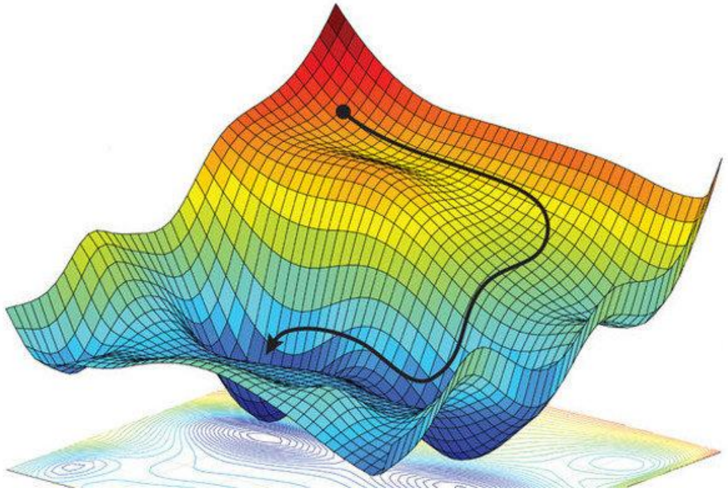
Given a parent function $T : \mathrm{R}^n \to \mathrm{R}$, the *Geometric Semantic Mutation* returns the real function $T_M = T + ms \times (T_{R1} - T_{R2})$ where T_{R1} and T_{R2} are random real function.

GEOMETRIC SEMANTIC GENETIC PROGRAMMING HYBRIDIZED WITH GRADIENT DESCENT

NEURAL NETWORK



Input features	x_i
Weights	w_i
Output	y
Loss function	$\mathrm{Min}\mathcal{L}(y, f(x, w))$



ADAM

The **Adam** algorithm is an extension to stochastic gradient descent.

Adam is an Optimization Algorithm to update network weights iterati ve based in training data.

Specifically:

- Calculates an exponential moving average of the gradient and the squared gradient
- The parameters β1 and β2 control the decay rates of these moving averages.

ADAM

- $\alpha = 0.001, \beta_1 = 0.9, \beta_2 = 0.999, \eta = 10^{-8}$ (Defaults)

$m_0 \leftarrow 0$ (Initialize 1st moment vector)
$v_0 \leftarrow 0$ (Initialize 2nd moment vector)
$i \leftarrow 0$ (Initialize step)
while Θ_i not converged **do**
  $i \leftarrow i + 1$
  $g_i \leftarrow \nabla_{\Theta} f_i(\Theta_{i-1})$ (Get gradients at step *i*)
  $m_i \leftarrow \beta_1 \cdot m_{i-1} + (1 - \beta_1) \cdot g_i$ (Update biased first moment estimate)
  $v_i \leftarrow \beta_2 \cdot v_{i-1} + (1 - \beta_2) \cdot g_i^2$ (Update biased second raw moment estimate)
  $\hat{m}_i \leftarrow m_i / (1 - \beta_1^i)$ (Compute bias-corrected first moment estimate)
  $\hat{v}_i \leftarrow v_i / (1 - \beta_2^i)$ (Compute bias-corrected second raw moment estimate)
  $\Theta_i \leftarrow \Theta_{i-1} - \alpha \cdot \hat{m}_i / (\sqrt{\hat{v}_t} + \eta)$ (Update parameters)
end while
return Θ_i (resulting parameters)

GSGP HYBRIDIZED WITH ADAM

A combination of these techniques should guarantee a jump in promising areas of the solution space, thanks to the evolutionary search of GSGP, and subsequent refinement of the solution obtained with the Adam algorithm.

GSGP		ADAM
• Providing **structural changes in the shape** of individuals • **New areas** of the solution space can be explored	STRENGHT	• Optimizes a series of parameters of the individuals • Perform **small shifts** in the **local area** of the solution space
Doesn't optimizes a series of parameters of the individuals	WEAKNESSES	Get stuck in local optima

BUT.. HOW CAN WE COMBINE THEM???

Q : What should Adam optimize?

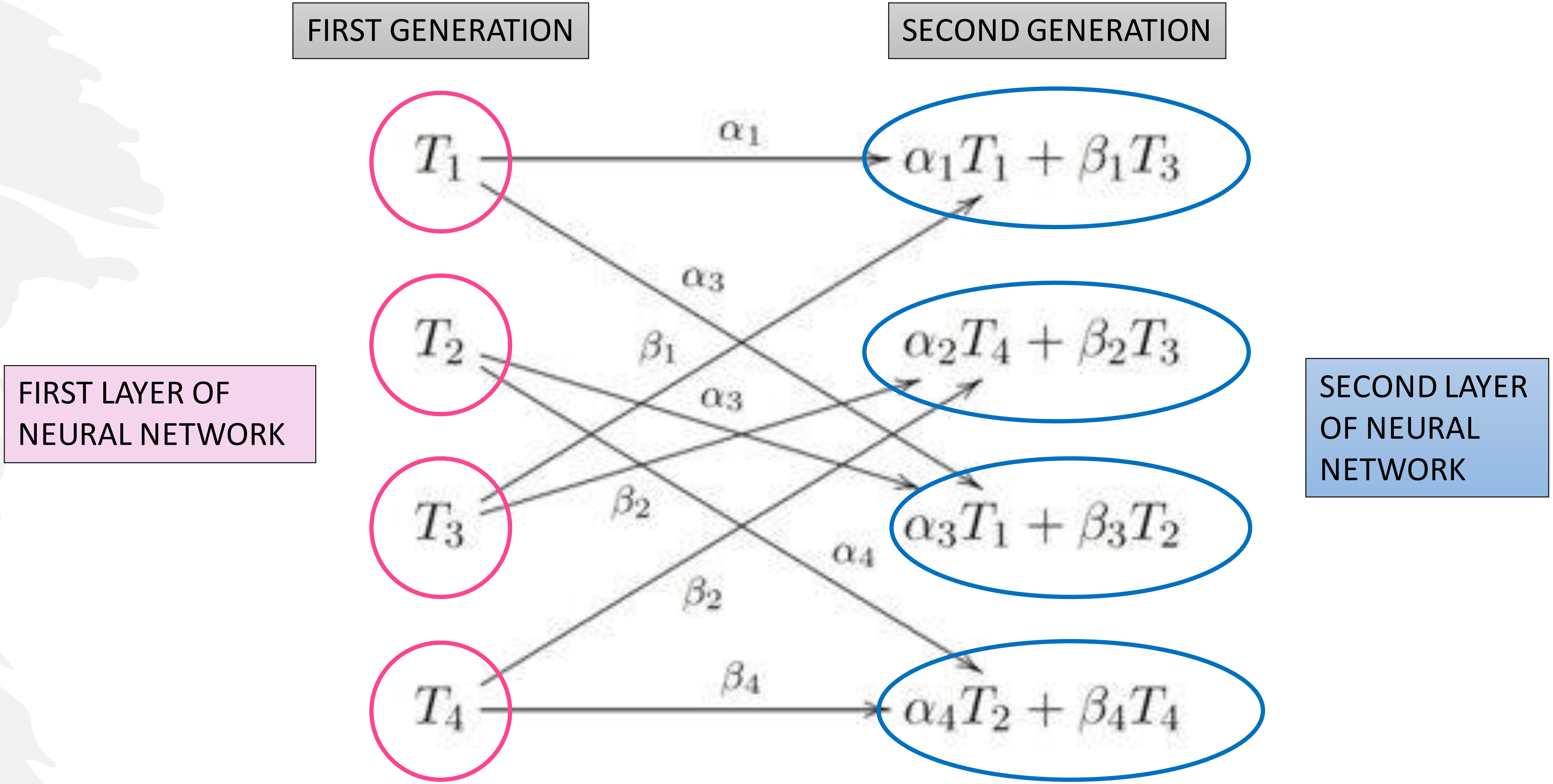
A : The vector $T = (T_1, \ldots, T_N)$, that is the vector of the new generations obtained, is composed of **derivable functions**.

Q : Respect to which parameters should Adame optimize?

$$Geometric\ Semantic\ Mutation: \ T_M = T + ms \times (T_{R1} - T_{R2})$$

$$Geometric\ Semantic\ Crossover: \ T_{X0} = (T_1 xa) + ((1 - a) \times T_2)$$

A : The values of a , m and $b = 1 - a$ are randomly initialized...and we can easily derive!


FIRST GENERATION
SECOND GENERATION
FIRST LAYER OF NEURAL NETWORK
SECOND LAYER OF NEURAL NETWORK
T_1
T_2
T_3
T_4
α_1
α_3
β_1
α_3
β_2
α_4
β_2
β_4
$\alpha_1 T_1 + \beta_1 T_3$
$\alpha_2 T_4 + \beta_2 T_3$
$\alpha_3 T_1 + \beta_3 T_2$
$\alpha_4 T_2 + \beta_4 T_4$

THE TWO METHOD PROPOSED

HYB-GSGP

Hybrid Geomettric Semantic Genetic Programming

One step of GSGP is alternated to one step of the Adam optimizer

HeH-GSGP

Half et Half Geometric Semantic Genetic Programming

Initially, all the GSGP genetic steps are performed, followed by an equal number of Adam optimizer steps

RESULTS

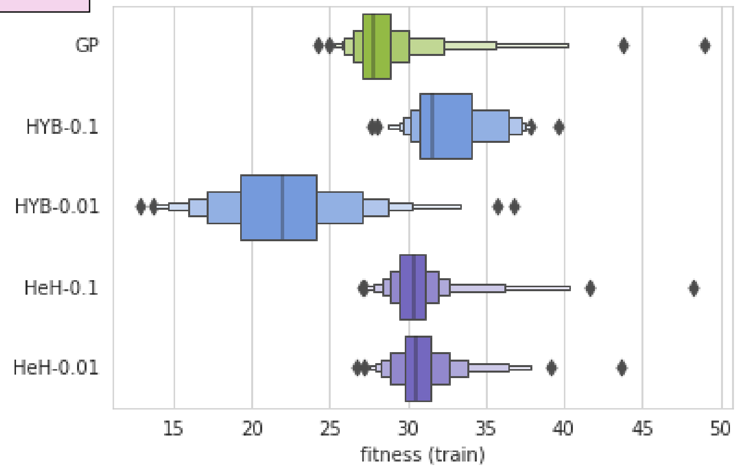
Principal characteristics of the considered datasets

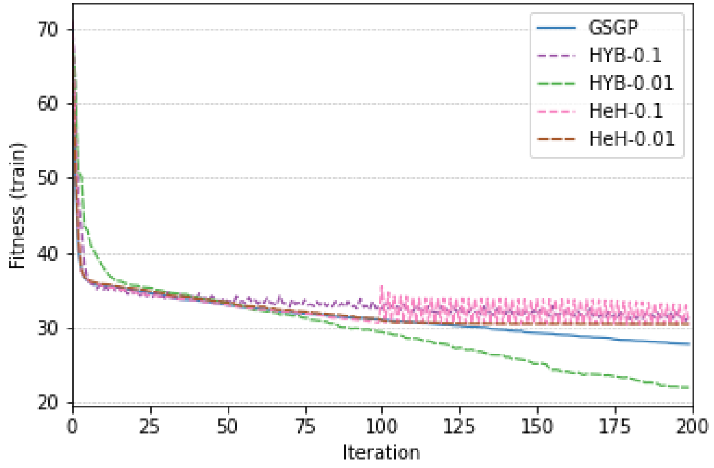
Dataset	Variables	Instances	Area	Task
%F	242	359	Pharmacokinetic	Regression
LD50	627	234	Pharmacokinetic	Regression
%PPB	627	131	Pharmacokinetic	Regression
yac	7	308	Physics	Regression
slump	10	102	Physics	Regression
conc	9	1030	Physics	Regression
air	6	1503	Physics	Regression

Training and testing fitness (RMSE)

		GSGP	HYB-0.1	HYB-0.01	HeH-0.1	HeH-0.01
%F	Train	38.08	37.74	**36.80**	39.61	40.60
	Test	40.15	40.48	**39.61**	40.85	41.23
LD50	Train	2118.00	**2086.56**	2128.22	2144.27	2161.00
	Test	2214.78	**2203.25**	2229.87	2221.72	2215.09
%PPB	Train	30.15	27.00	**24.32**	34.79	33.26
	Test	328.1	401.43	263.81	**213.86**	235.53
yac	Train	**11.83**	11.92	12.48	12.28	12.31
	Test	11.92	**11.83**	12.52	12.38	12.48
slump	Train	4.56	3.47	**2.92**	5.19	4.41
	Test	5.08	3.63	**3.32**	5.77	4.76
conc	Train	9.62	8.86	**8.50**	10.59	10.05
	Test	9.65	8.88	**8.69**	10.47	10.07
air	Train	27.76	31.54	**21.98**	30.37	30.46
	Test	27.94	31.71	**21.97**	30.15	30.53

TRAIN SET



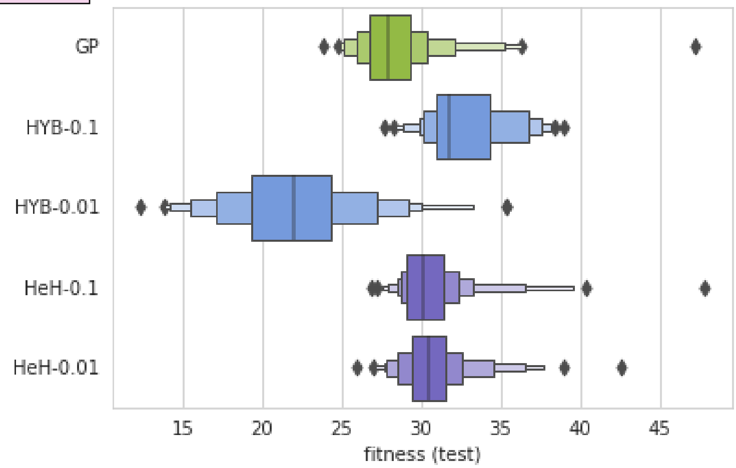


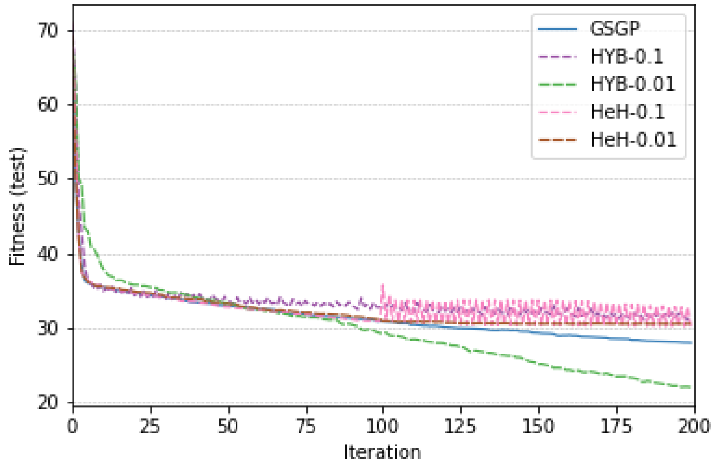
Air

Airfoil self-noise

NASA dataset obtained from a series of aereodynamic and acoustic test of airfoil blade section

TEST SET



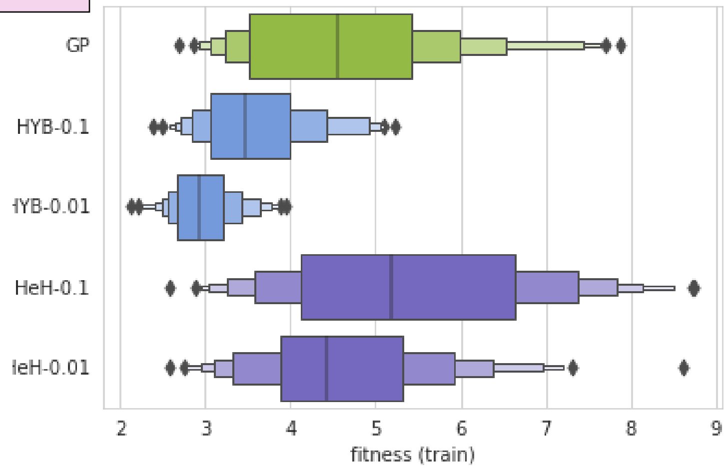


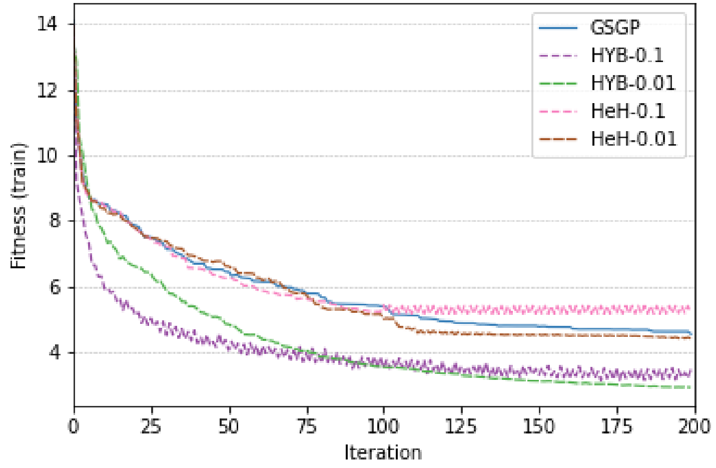
Slump

Concrete Slump

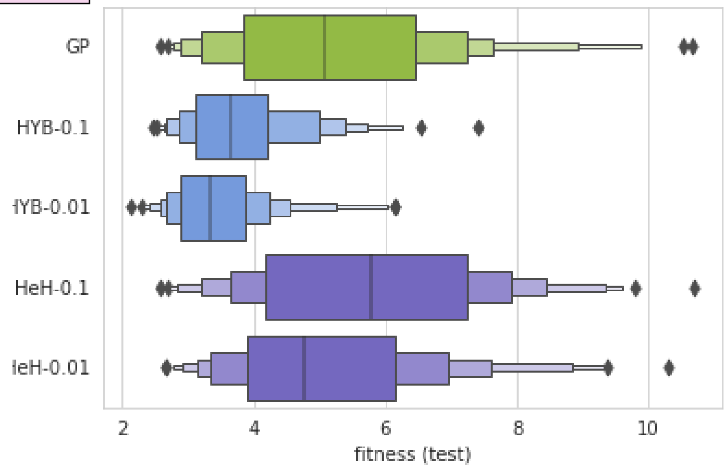
Measures the value about the slump flow of the concrete.

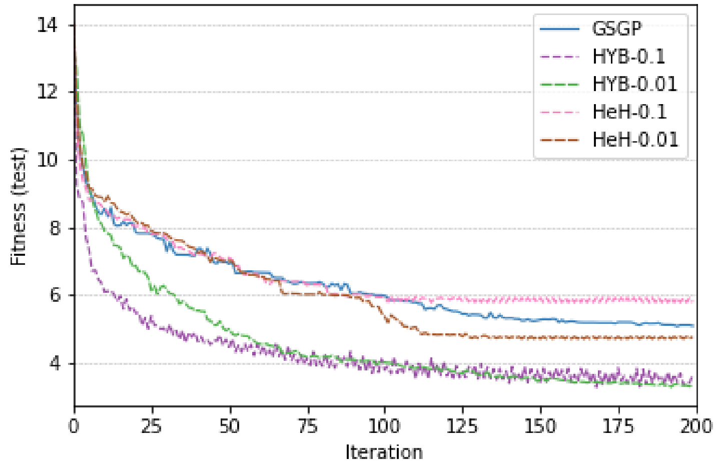
TRAIN SET





TEST SET



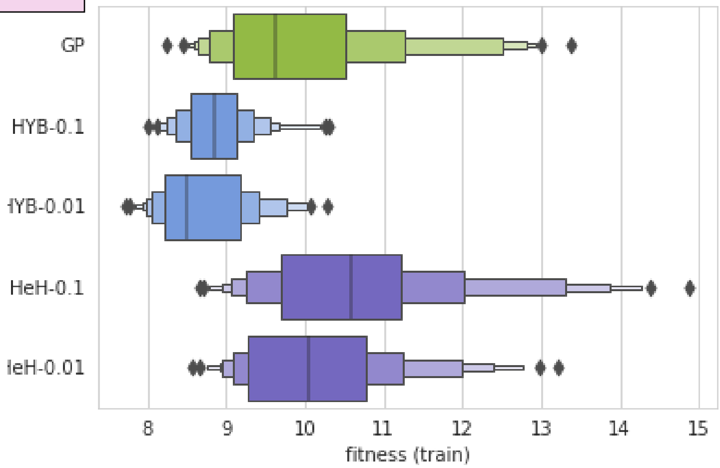


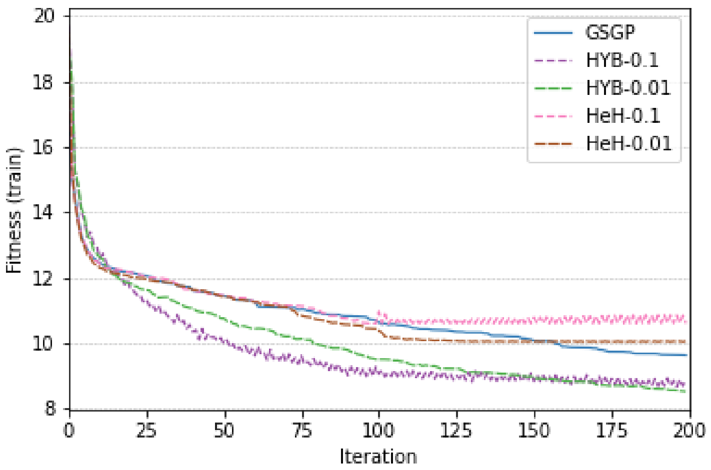
Conc

Concrete Complessive Strength

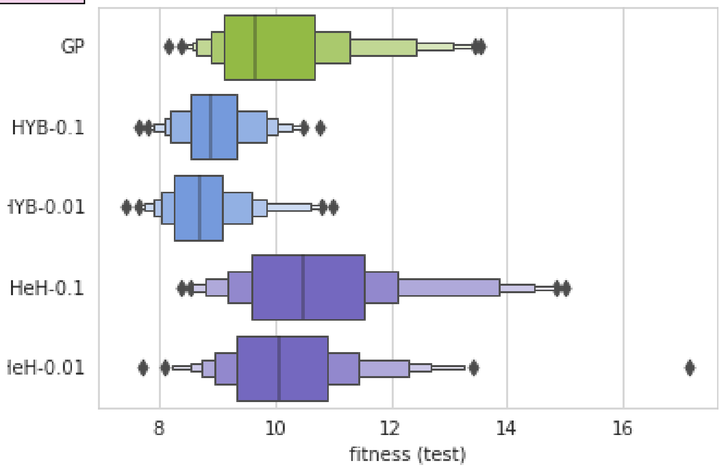
Measures the value about the compressive strength of the concrete.

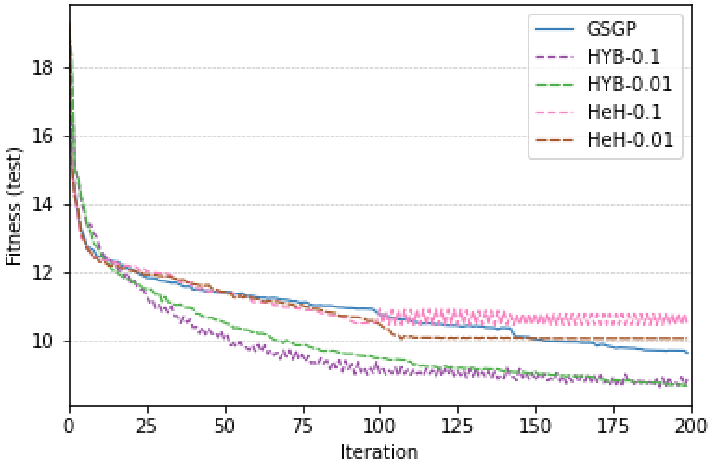
TRAIN SET





TEST SET





CONCLUSION

The combination of GSGP with the Adam optimizer can improve the performance of GSGP

HYB-GSGP outperforms classic GSGP in both training and test sets with a statistically significant difference on the test set

HYB-GSGP converges to good-quality solutions faster than classical GSGP as it requires fewer epochs to converge

HeH-GSGP does not outperform **GSGP** even if it generally ensures good quality results on the test set

GRAZIE PER L'ATTENZIONE