

Programación Genética y Métodos Numéricos en Regresión Simbólica

Rocío Luján Cecchini ¹,
Gustavo Esteban Vazquez ¹, Nélidea Beatriz Brignole ^{1,2}

¹ Laboratorio de Investigación y Desarrollo en Computación Científica (LIDeCC),
Departamento de Ciencias e Ingeniería de la Computación,
Universidad Nacional del Sur, Bahía Blanca

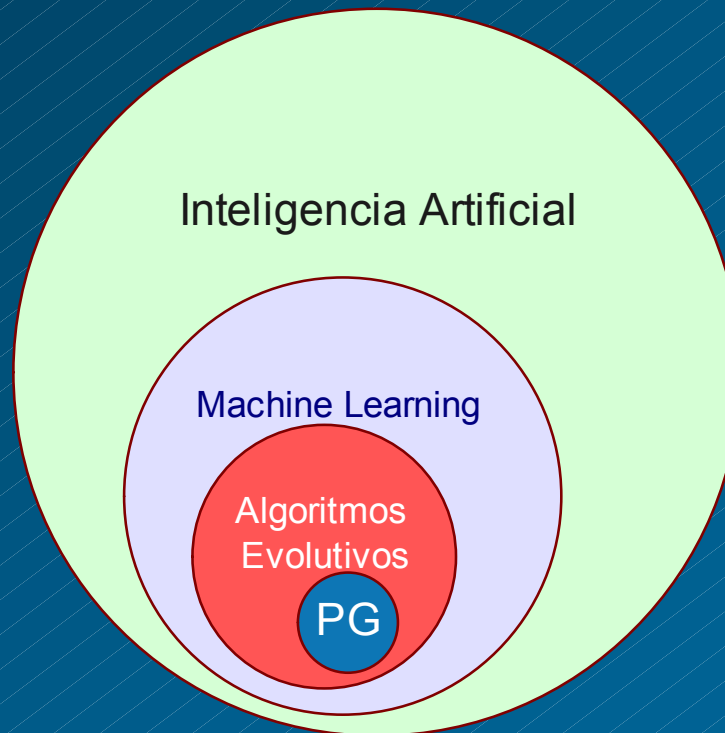
² Planta Piloto de Ingeniería Química – UNS – CONICET, Bahía Blanca



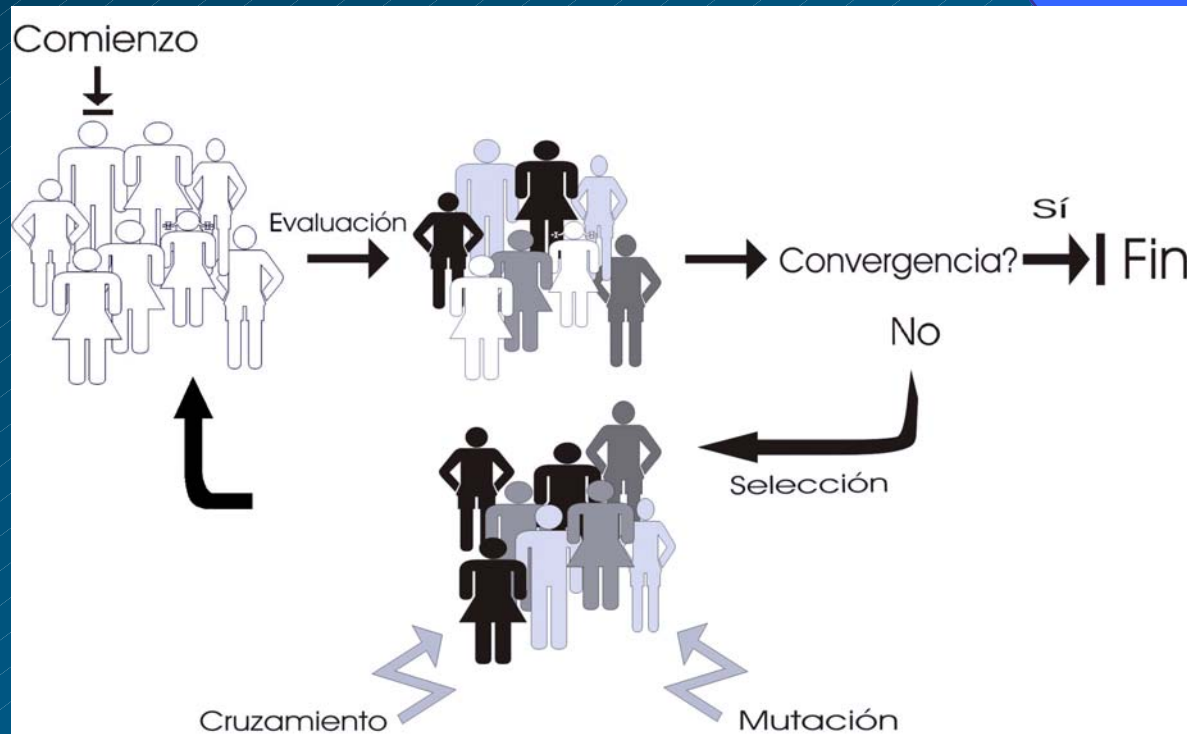
- **Programación Genética (PG) - Revisión**
 - Contexto
 - Descripción
- **Regresión Simbólica**
 - Pasos básicos
 - Pasos previos a la implementación
 - Ejemplo clásico de individuos de PG
- **Combinación de PG y Métodos Numéricos**
 - Diferencia con las técnicas conocidas
- **Algoritmo Básico**
- **Experimentos**
 - Caso de estudio
 - Pasos Previos
 - Resultados
 - Problema de las expresiones oscilatorias
- **Algoritmo Multi-Objetivo**
 - Pasos Previos
 - Resultados
 - Otro Caso de Estudio
- **Conclusiones y Trabajo Futuro**

- Contexto

- La Programación Genética (PG) es un caso particular de algoritmos evolutivos. Los cuales se basan en los principios de evolución definidos por Darwin.



- Descripción
 - Pasos Básicos
 - Estos principios incluyen mecanismos como reproducción, mutación (operadores genéticos), selección natural y supervivencia del más apto.





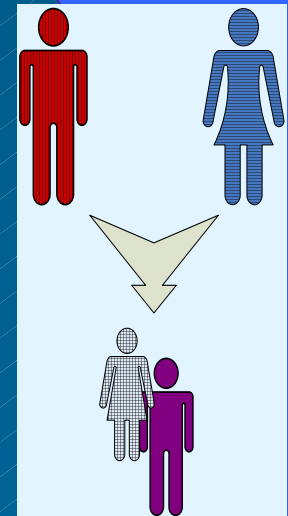
Programación Genética – Conceptos Básicos

- En resumen cualquier algoritmo evolutivo consta de cuatro pasos principales:

- generación de la población inicial, (ramped half and half)
- evaluación de los individuos, (fitness)
- selección de padres para producir una nueva población, (lexictour)
- generación de la nueva población mediante la aplicación de operadores genéticos adecuados,



Programas !!





Qué es Regresión Simbólica?

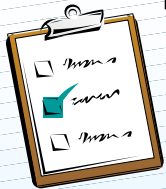


Es el proceso de descubrir tanto la forma funcional de una nube de datos especificada, como todos sus coeficientes involucrados (o al menos una aproximación de ellos)

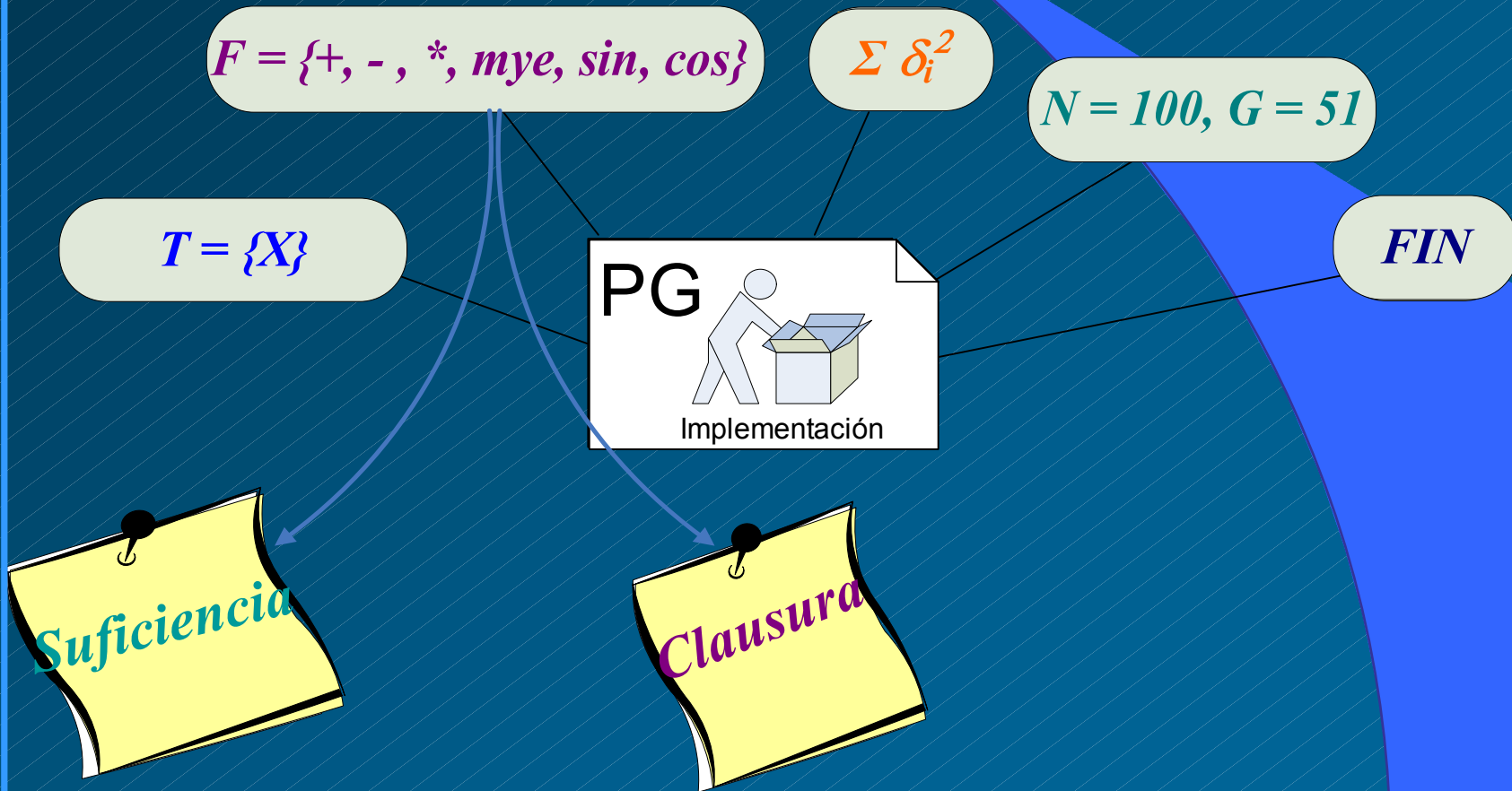


Programación Genética – Conceptos Básicos

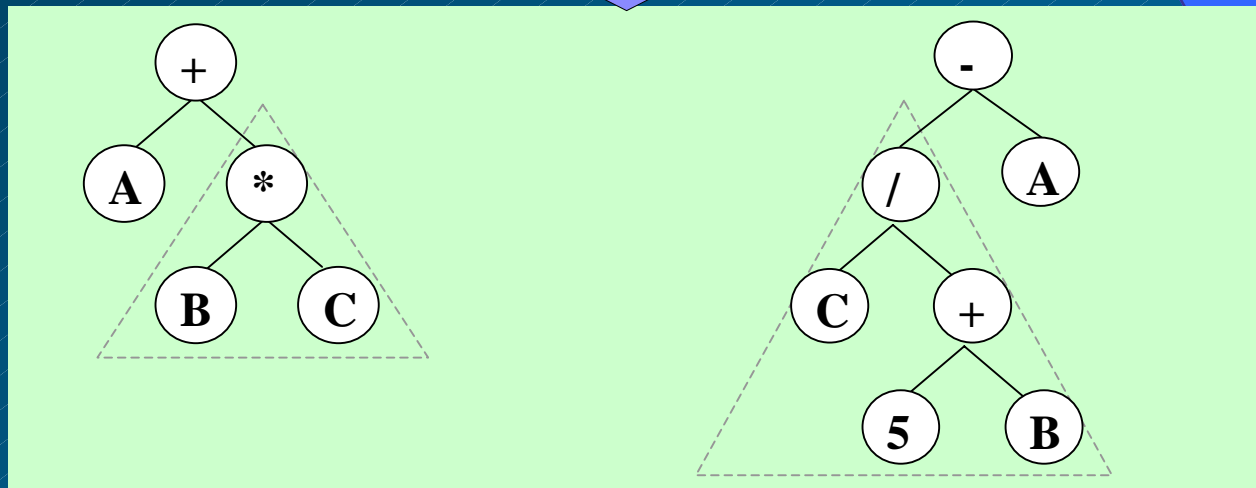
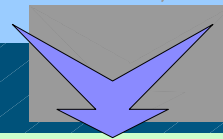
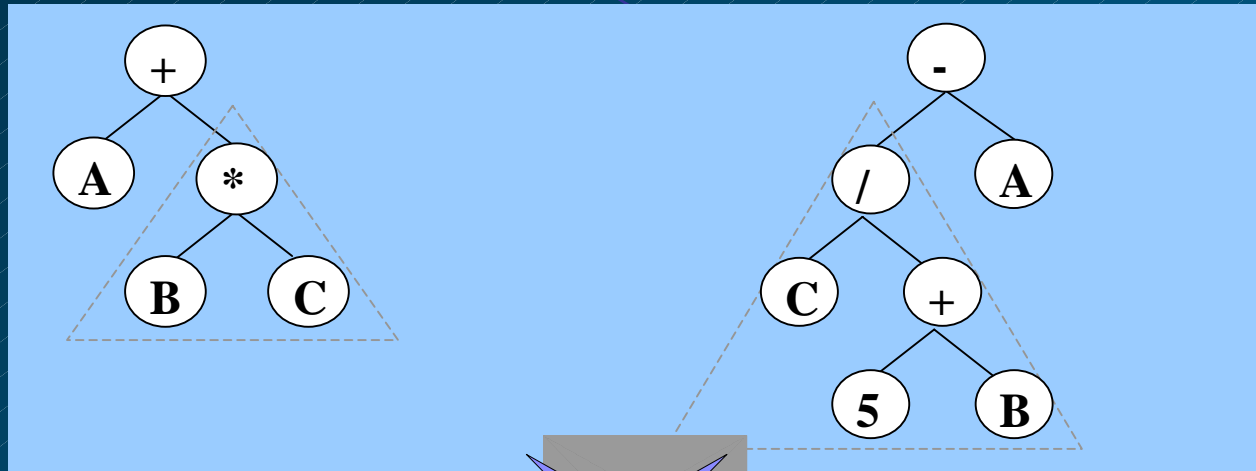
– Pasos Previos



- Cuando se lleva a cabo cualquier implementación de programación genética para resolver un problema particular se requieren pasos preparatorios.



– Ejemplo: Crossover





- Diferencias con las técnicas conocidas:

Implementaciones clásicas:



No contemplan coeficientes de regresión:

$$F = \{ +, -, *, / \}$$
$$T = \{ x_1, x_2 \}$$

$x_1 * (x_1 + x_2)$

$\beta_1 x_1 * () + \beta_4$

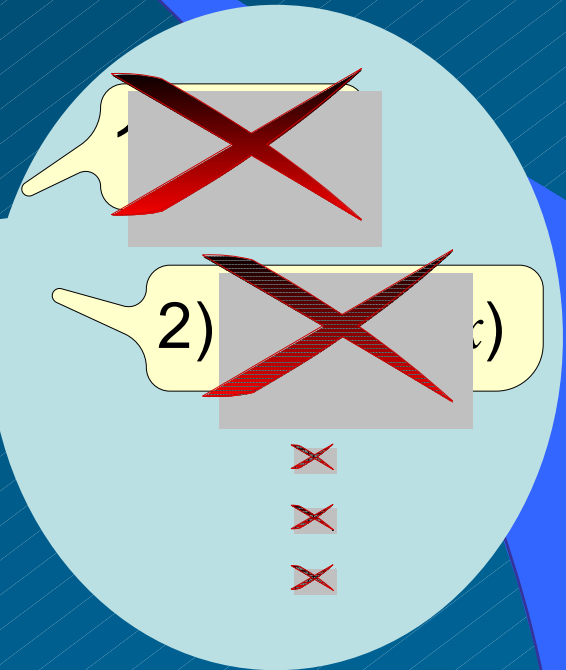
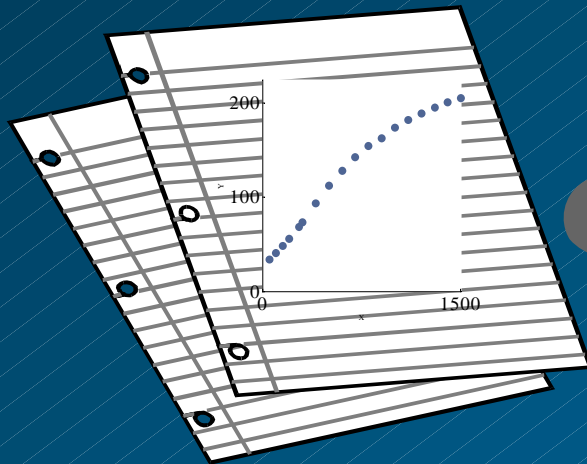
Coeficiente de regresión

Contemplan coeficientes pero siguiendo un esquema de PG puro



Combinación de Programación Genética y Métodos Numéricos

Por otro lado usar sólo métodos matemáticos constituye una limitación pues debemos conocer de antemano la forma que siguen los datos.





Algoritmo GP_NLMS_básico (g, n, muestra, parámetros, best)
crear una nueva población de hijos

(full, grow, *ramped half and half*)

mientras no se alcance la condición de terminación

1 - para cada hijo i en la nueva población:

ajustar coeficientes con CMNL:

exp \leftarrow obtener expresión matemática del individuo

expSimple \leftarrow eliminar redundancias matemáticas(exp)

familiai \leftarrow obtener familia(expSimple)

si esta familia de funciones no fue evaluada

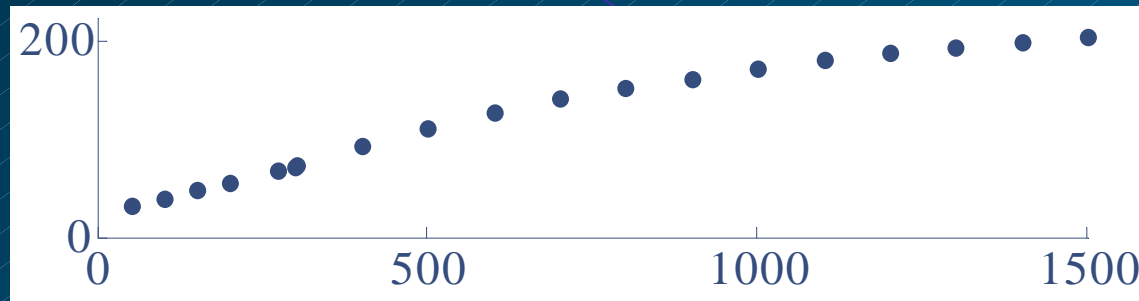
coeficientesi \leftarrow ajustar por cuadrados mínimos no
lineales(familiai, muestra)

f \leftarrow evaluar error(familiai, coeficientesi, muestra)

2 - construir la nueva generación:

- seleccionar padres (roulette, tournament, *lexictour*)
- generar nuevos individuos, aplicar operadores (*crossover, mutation*)
- crear la nueva población seleccionando los sobrevivientes de entre la población actual y los nuevos individuos generados (Replace, *Keepbest*, Halfelitism, Totalelitism).

- Caso de estudio



Muestra de datos correspondiente a los experimentos realizados sobre la Capacidad Calorífica del Gas Propano a Presión Constante

– Pasos previos:

- $F = \{ +, -, *, \text{mye}, \sin, \cos \}$

- $T = \{ x \}$

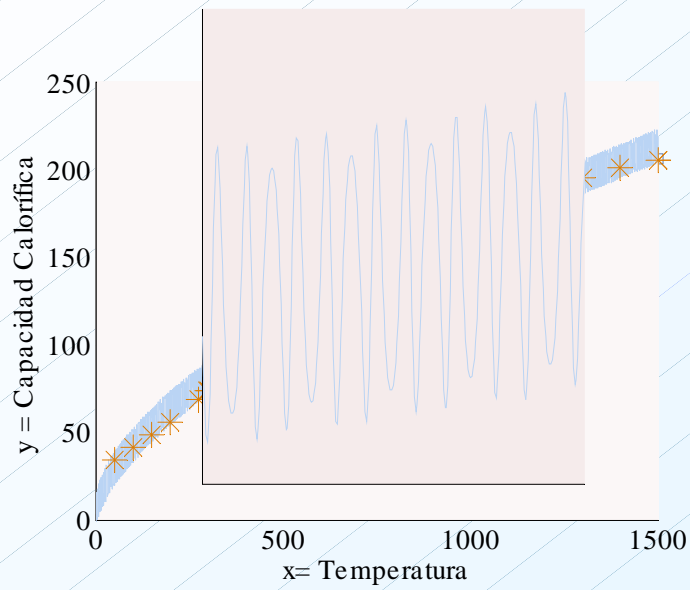
- *Fitness*: $\sqrt{\sum \delta^2}$

- *Parámetros*: $g = 150; n = 20.$

Resultados

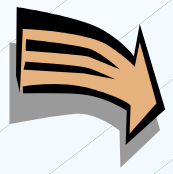


**Individuo óptimo
hallado con sus
coeficientes de
regresión**



- Resultado de evaluar la expresión en los puntos de la muestra
- ✱ Valores de los puntos de la muestra
- Gráfica completa del individuo

Fitness
8.3898

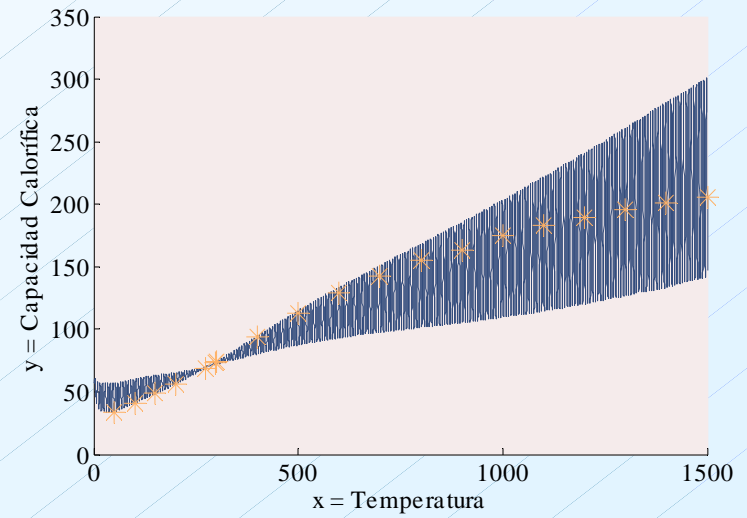
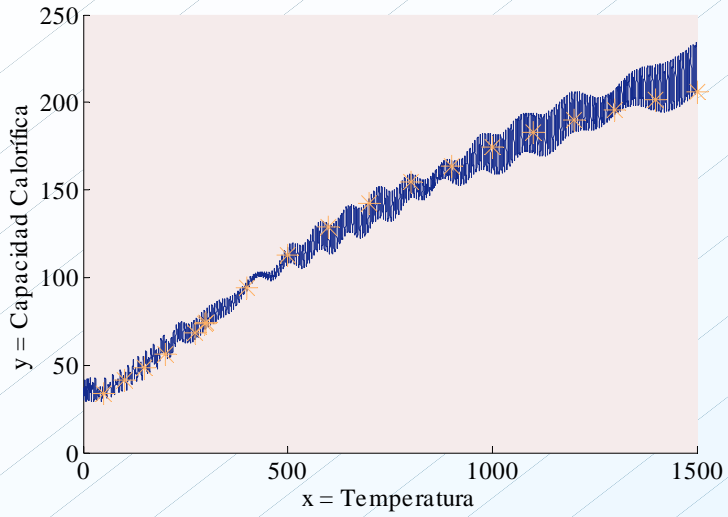


$$f_1(x) = \beta_1 x^{\beta_2} - \beta_3 \cos(\beta_4 x) - \beta_5 e^{(\beta_6 \sin(\beta_7 \cos(\beta_8 x^{\beta_9})))} + \beta_{10}$$



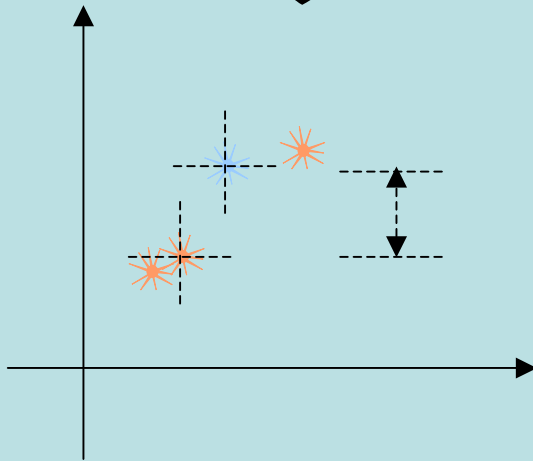
- Otros resultados oscilatorios

$$f(x) = my \log(x)^2 + 2 \sin(x) + \sin(\sin(x))x + \cos(\sin(my \log(x)))$$

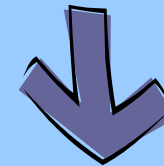
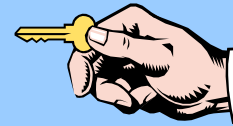


$$f(x) = (\cos(\sin(x)) + my \log(x) - \cos(x)) * (my \log(2x) - x) - 2 * \cos(my \log(x)) * x$$

Cual fue el problema?



Qué solución podemos aplicar?



Búsqueda Multi-Objetivo

- ▶ Ajuste en los puntos
- ▶ Poco comportamiento oscilatorio

Algoritmo GP_NLMS_Multi-Objetivo(g, n, muestra, parámetros, best)

crear una nueva población de hijos

(full, grow, *ramped half and half*)

mientras no se alcance la condición de terminación

1 - para cada hijo i en la nueva población:

ajustar coeficientes con CMNL:

exp ← obtener expresión matemática del individuo

expSimple ← eliminar redundancias matemáticas(exp)

familiai ← obtener familia(expSimple)

si esta familia de funciones no fue evaluada

coeficientesi ← ajustar por cuadrados mínimos no
lineales(familiai, muestra)

*%calcular fitness (f) teniendo en cuenta dos objetivos (error y
%oscilaciones)*

fe ← evaluar error(familiai, coeficientesi, muestra)

fo ← evaluar oscilaciones(familiai, coeficientesi, muestra)

*f ← pf*fo + fe*

2 - construir la nueva generación:

- seleccionar padres (roulette, tournament, *lexictour*)
- generar nuevos individuos, aplicar operadores (*crossover, mutation*)
- crear la nueva población seleccionando los sobrevivientes de entre la población actual y los nuevos individuos generados (Replace, *Keepbest*, Halfelitism, Totalelitism).

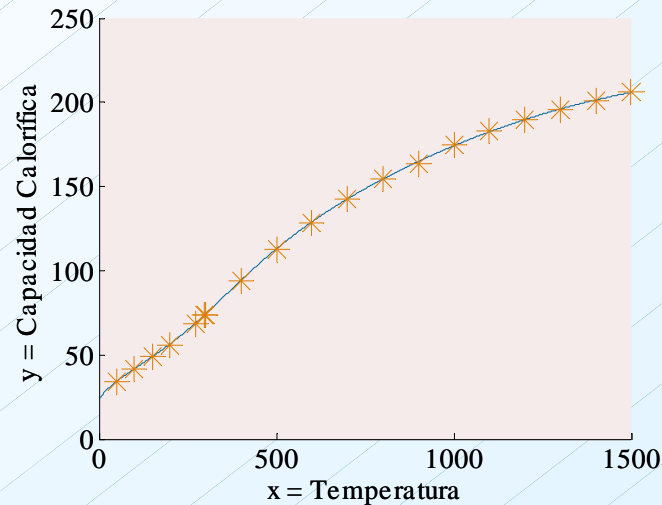
• Resultados



Individuo óptimo hallado con sus coeficientes de regresión



$$f_2(x) = \beta_1 e^{(\beta_2 e^{(\beta_3 x^{\beta_4} - \beta_5 x)})} x^{\beta_6} (\beta_7 + \beta_8 x) + \beta_9$$

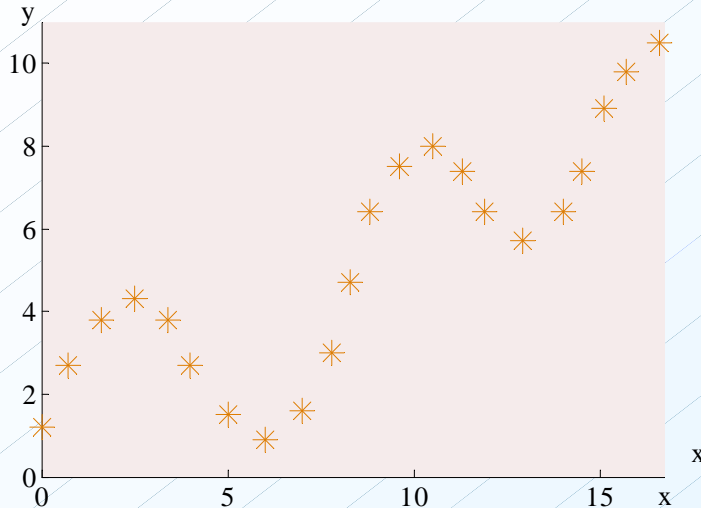


$$\sqrt{\sum \delta^2} + \text{M}$$

Fitness

1.6792

- Otro caso de estudio



Muestra de datos generada de forma artificial

– Pasos previos:

- $F = \{ +, -, *, \text{mye}, \text{sin}, \text{cos} \}$

- $T = \{ x \}$

- *Fitness*: $\sqrt{\sum \delta^2} + \text{MNV}$

- Parámetros: $g = 80; n = 100.$

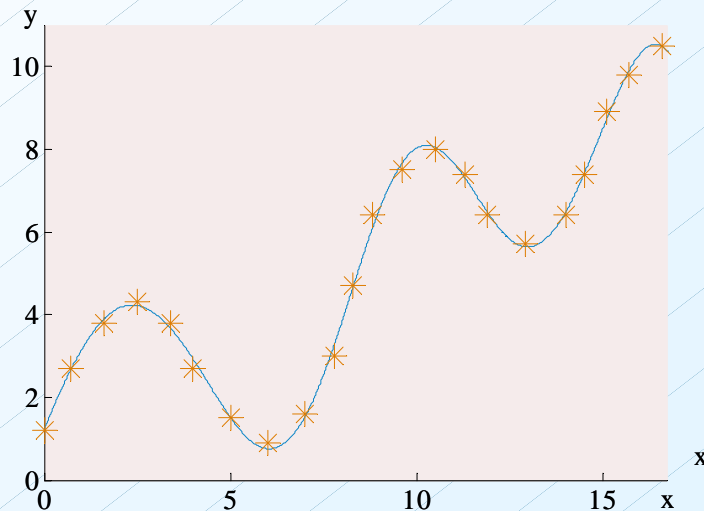
• Resultados



Individuo óptimo hallado con sus coeficientes de regresión



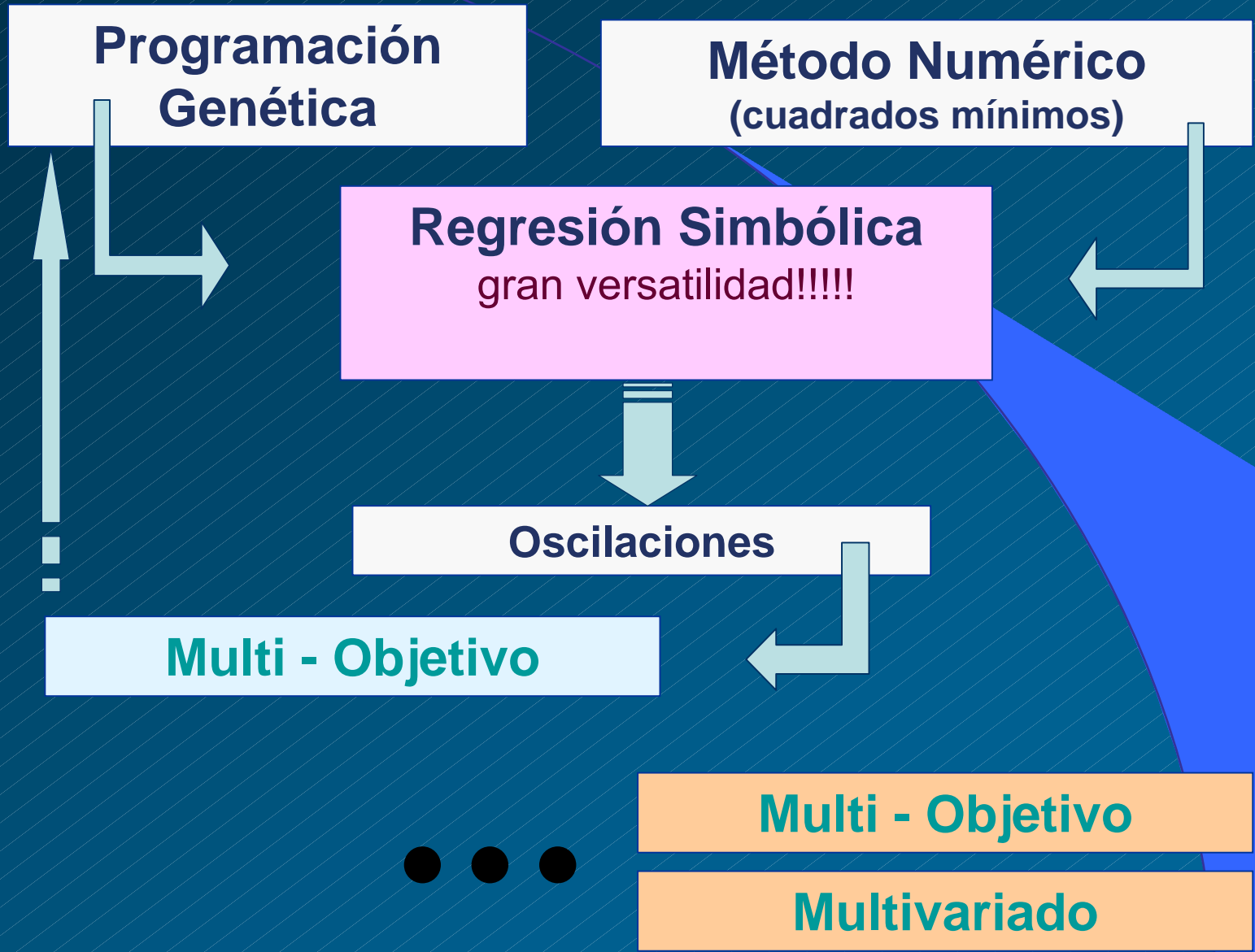
$$f_3(x) = -\beta_1 x^{\beta_2} + \beta_3 x^{\beta_4} - \beta_5 x^{\beta_6} + \beta_7 e^{(\beta_8 x - \beta_9 x^{\beta_{10}})} + \beta_{11} e^{(\beta_{12} x - \beta_{13} x^{\beta_{14}})} + \beta_{15}$$



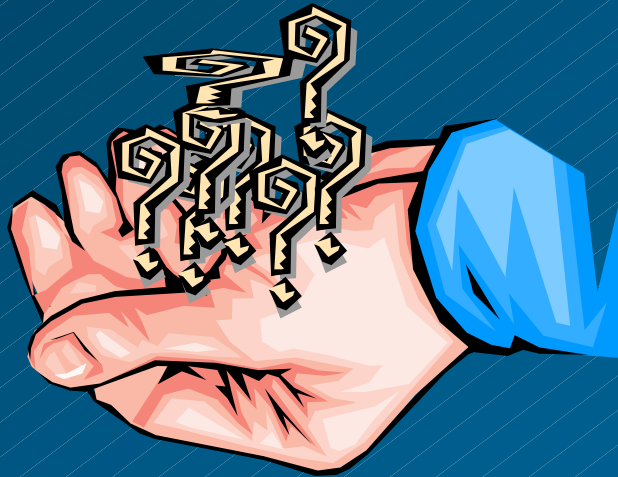
Fitness
1.6792



Conclusiones y Trabajo Futuro



Muchas gracias





REFERENCIAS - REGRESIÓN SIMBÓLICA CON PG

- Koza J.R. *“Genetic Programming: On the Programming of Computers by Means of Natural Selection”*. The MIT Press. Chapter. 10. 1992.
- Koza J.R. *“Genetic Programming: A Paradigm for Genetically Breeding Populations of Computer Programs to Solve Problems”*. Sec. 4.3.2, 1990.
- Robinson A. *“Genetic Programming: Theory, Implementation and the Evolution of Unconstrained Solutions”*. Hampshire College. Thesis. Sec. 4.4. 2001.
- Sugimotoa M., Kikuchia S., Tomitaa M *“Reverse Engineering of Biochemical Equations from Time-Course Data by Means of Genetic Programming”*. BioSystems, 80, 155-164. Elsevier. 2005.
- Cai W., Pacheco-Vega A., Sen M., Yang K.T. *“Heat Transfer Correlations by Symbolic Regression”*. International Journal of Heat and Mass Transfer. En prensa. Elsevier. 2006.
- Grosman B., Lewin D.R. *“Adaptive Genetic Programming for Steady-State Process Modeling”*. Computers & Chemical Engineering, 28 (2004) 2779–2790. Elsevier. 2004.
- Greeff D.J., Aldrich C. *“Empirical Modelling of Chemical Process Systems with Evolutionary Programming”*. Pergamon. 1997.

- Se hizo en MATLAB modificando un Toolbox existente, el GPLab. El cual cuenta con las operaciones necesarias para la ejecución de la metodología.
- Se consideró una versión válida para muestras de una sola variable independiente.

