

Introduction to Artificial Intelligence



VICTORIA UNIVERSITY OF
WELLINGTON
TE HERENGA WAKA

COMP307/AIML420

**Evolutionary Computation 3:
GP for Regression and Classification**

Dr Fangfang Zhang

fangfang.zhang@ecs.vuw.ac.nz

COMP307/AIML420 Week 6

1. Announcements

- Assignment 2 (**12%**)
- Part 1
- Helpdesk available during the teaching break
- https://www.researchgate.net/publication/216301261_A_Field_Guide_to_Genetic_Programming (a useful book, part 1)
- Andrew's teaching evaluation (*until 12th April*)

2. Genetic Programming for **Regression**

Part 2

3. Genetic Programming for **Classification**

Book

I Basics	7
2 Representation, Initialisation and Operators in Tree-based GP	9
2.1 Representation	9
2.2 Initialising the Population	11
2.3 Selection	14
2.4 Recombination and Mutation	15
3 Getting Ready to Run Genetic Programming	19
3.1 Step 1: Terminal Set	19
3.2 Step 2: Function Set	20
3.2.1 Closure	21
3.2.2 Sufficiency	23
3.2.3 Evolving Structures other than Programs	23
3.3 Step 3: Fitness Function	24
3.4 Step 4: GP Parameters	26
3.5 Step 5: Termination and solution designation	27
4 Example Genetic Programming Run	29
4.1 Preparatory Steps	29
4.2 Step-by-Step Sample Run	31
4.2.1 Initialisation	31
4.2.2 Fitness Evaluation	32
4.2.3 Selection, Crossover and Mutation	32
4.2.4 Termination and Solution Designation	35

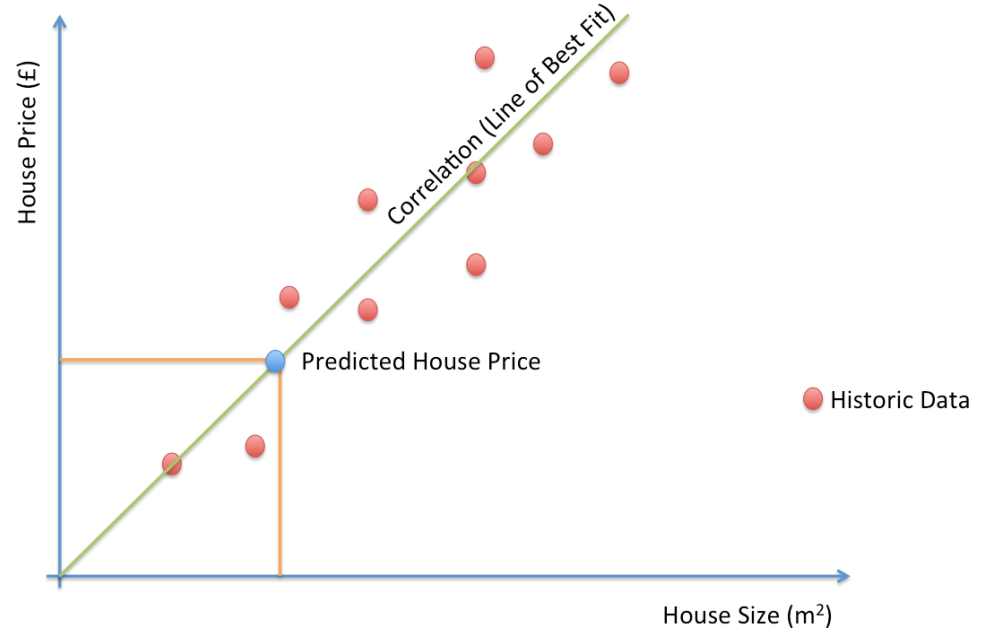
Outline

- Statistical parameter regression
- Symbolic regression
- GP for symbolic regression
- GP for binary classification

Learn the pattern of data 😊

House Price Prediction

- How **much** should your tender be?



Price	Floor space	Rooms	Lot size
250000	71	4	92
209500	98	5	123
349500	128	6	114
250000	86	4	98
419000	173	6	99
225000	83	4	67
549500	165	6	110
240000	71	4	78
340000	116	6	115

(Statistical) Regression Analysis

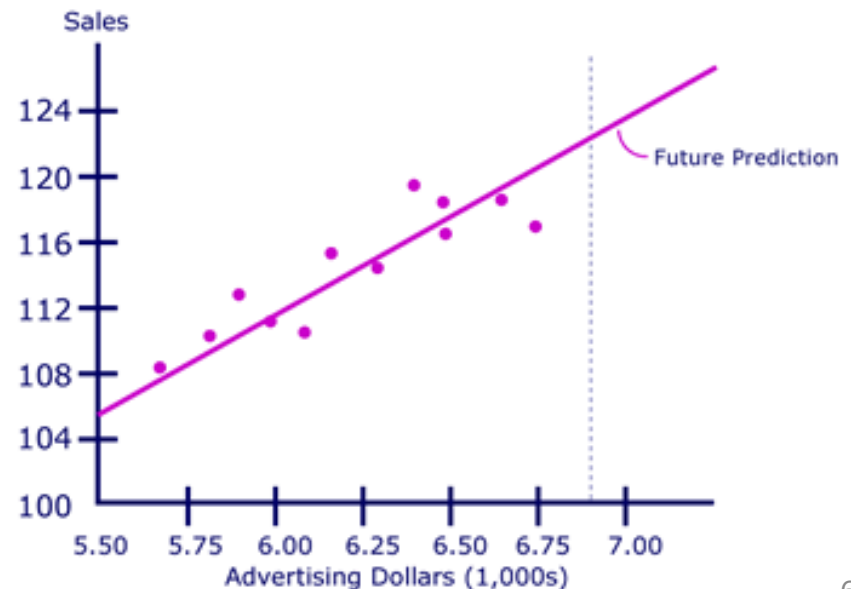
- In statistics, regression analysis examines the relation of a **dependent variable** (*response variable*) to specified **independent variables** (*explanatory variables*)

$$y = x^2$$

- The mathematical model of their relationship is the **regression equation** (e.g. $f(x, y) = 0$), e.g., $y - x^2 = 0$
- estimates of one or more hypothesized regression **parameters** (“**constants**”)

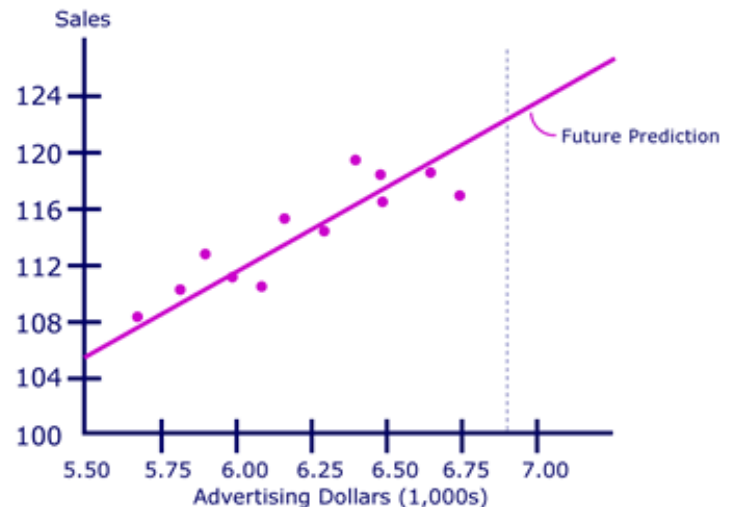
- **Examples**

- Financial prediction
- Saving prediction
- Ad cost vs sales
- Natural law discovery



Regression

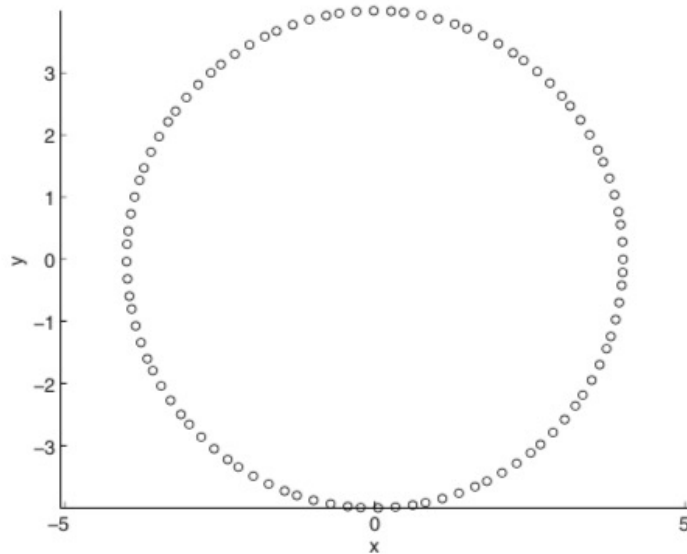
- Two main tasks:
 - **Regression equation:** relationship between a **dependent variable** (response variable) and **independent variables** (explanatory variables)
 - **Parameters/Coefficients**
- Example: linear regression
 - Regression equation is a **linear model**: $y = \alpha \cdot x + \beta + \epsilon$
 - **Coefficients**: α and β
 - α is the slope, β is the intercept
 - ϵ is the *error* term (assume normally distributed)
 - Estimate α and β to minimise ϵ
$$\epsilon = y - \alpha \cdot x - \beta$$



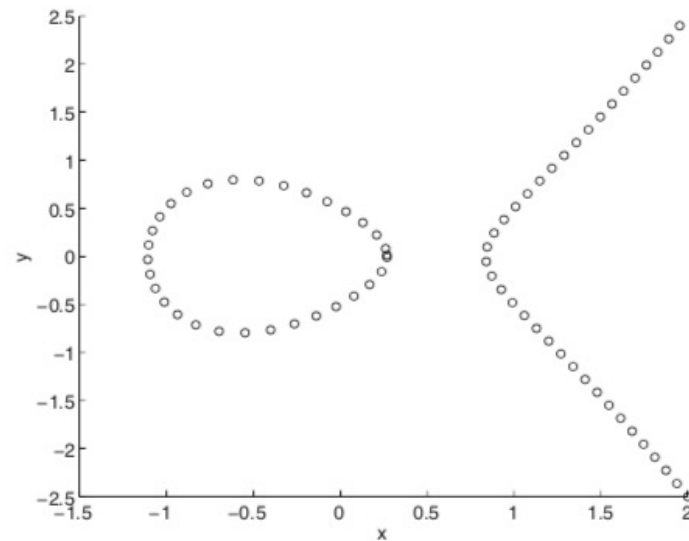
Assume model structure, estimate model parameters

Symbolic regression

- However, linear model is too **simple** for many real-world data
- Hard to find the proper regression equation



(a) Circle: $x^2 + y^2 - 4^2$



(b) Elliptic Curve: $x^3 + x - y^2 - 1.5$

Symbolic regression: A physics example

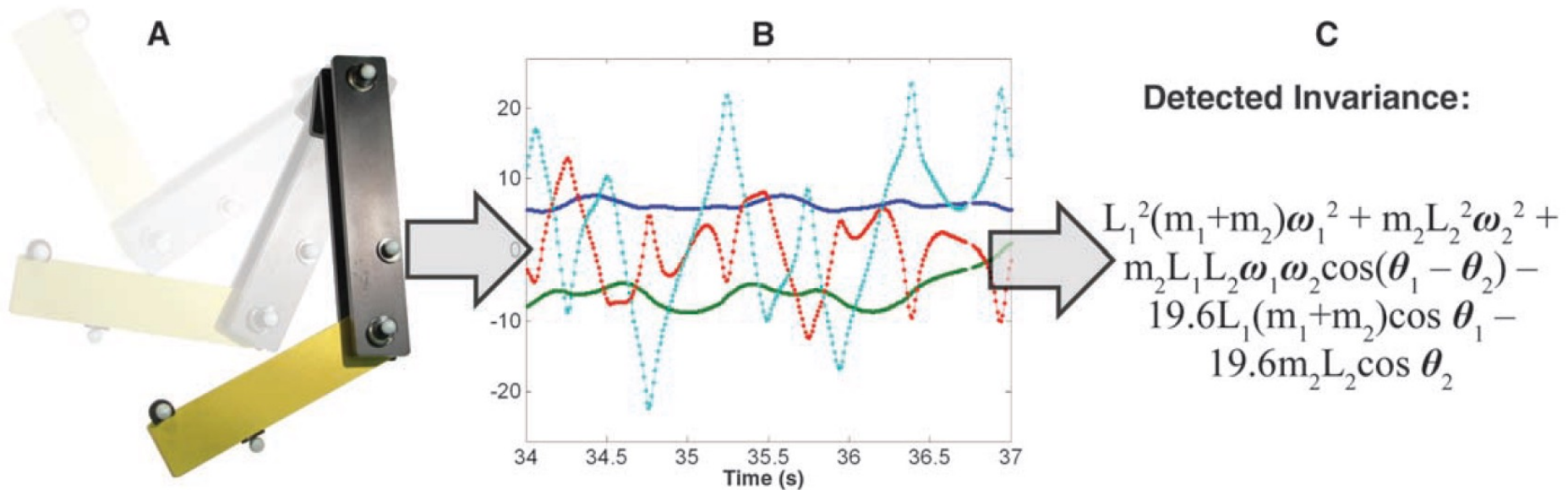


Fig. 1. ... We captured the angles and angular velocities of a chaotic double-pendulum (A) over time using motion tracking (B), then we automatically searched for equations... without any prior knowledge about physics or geometry, the algorithm found the conservation law (C)

Schmidt, Michael, and Hod Lipson. "Distilling free-form natural laws from experimental data." *Science* 324, no. 5923 (2009): 81-85. **1600 citations!**

Symbolic Regression Applications

- Symbolic regression has many real-world applications:
 - Economics prediction: [stock market](#) prediction, [GDP](#) prediction
 - Industrial prediction: prediction of [container handling capacity](#) at a sea port; short/medium/long-term prediction of [power load](#) at a region
 - Experimental formula modelling in engineering:
formulating the [amount of gas](#) emitted from a coal surface
 - Time series projection: CPI projection for a country or a region (or less capitalistic: [Gross National Happiness \(GNH\) index](#)?)
 - **Others?**

Symbolic regression

- Symbolic regression: finds a **symbolic description** of a model, not just a set of coefficients/parameters in a pre-specified model.
- Finds both:
 - the **model structure**, and
 - the corresponding **coefficients/parameters**

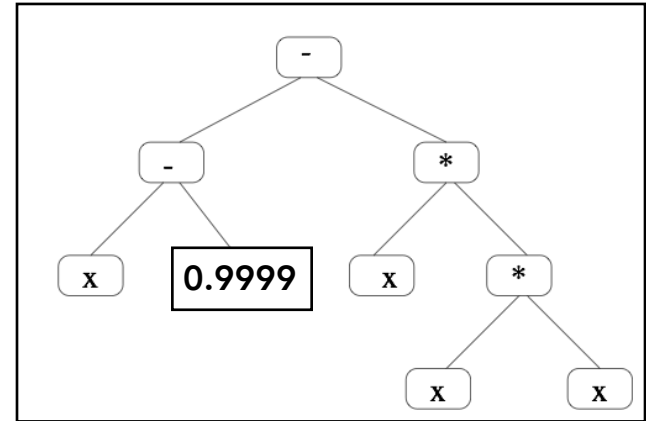
GP for Symbolic Regression (Example)

- **Objective:** Find a **program/model** that produces the correct value of the **dependent variable y** when given the value of an **independent variable x**
- **Terminal Set:** x , random constant
- **Function Set:** $\{+, -, \times, \div\}$
- **Fitness Cases:** 50 cases of x and the corresponding y values
(e.g. *50 instances/patterns/cases*)
- **Fitness Measure:** Sum of the **absolute** errors for the 50 cases
- **Parameters:** Population = 100, Generations = 50, MaxDepth = 17
reproduction rate: 5%, crossover rate: 90%, mutation rate: 5%
- **Success:** The fitness value is smaller than a pre-defined value, e.g. 0.01
- **Termination criteria:** satisfactory solutions found, or at generation 50

$$y = x^2 = x * x$$

GP for Symbolic Regression (Example)

- One GP run gave: $y = (x - 0.9999) - x^3$
- Successful? If the true model is
$$y = (x - 1) - x^3$$



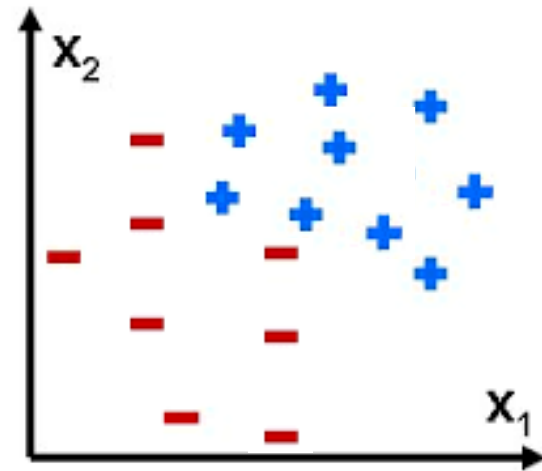
- Sometimes....:
`(% (% (* (* X 0.571) (* (- (* (+ (% 0.634094 0.68469) (+ (+ X X) - 0.5992))(* (* (+ (% 0.634094 0.68469) (+ X -0.5992)) (* (% 0.354904 - 0.7549) (* X 0.571)))) (- X 0.395493))) - 0.4665)another 15 lines) ☹`
- This example: one input variable (x), training set only
- Real-world applications: usually **multiple variables**, can have a **separate test set**, but use the same principles

GP for Symbolic Regression

- Compared with statistical parameter regression methods, GP method has the following properties:
 - Does **NOT** need to assume any **distribution** of data set,
 - Does **NOT** assume the **independence** of the input variables
 - Does **NOT** require any **statistical background knowledge** to assume any model
- Can **automatically** learn/evolve **both** the **model structure** and the **model parameters** at the same time
- System input: just the **data** with unknown model/parameters
- System output: a “**white-box**” model **structure** with appropriate **parameters** and **coefficients**

Binary Classification

- Binary classification is the task of classifying the instances of a given set into **two categories** based on whether they have some property or not
- Two target classes, e.g.
 - Disease vs non-disease
 - normal vs abnormal
 - approve loan or reject
 - fault vs normal state
 - object vs non-object

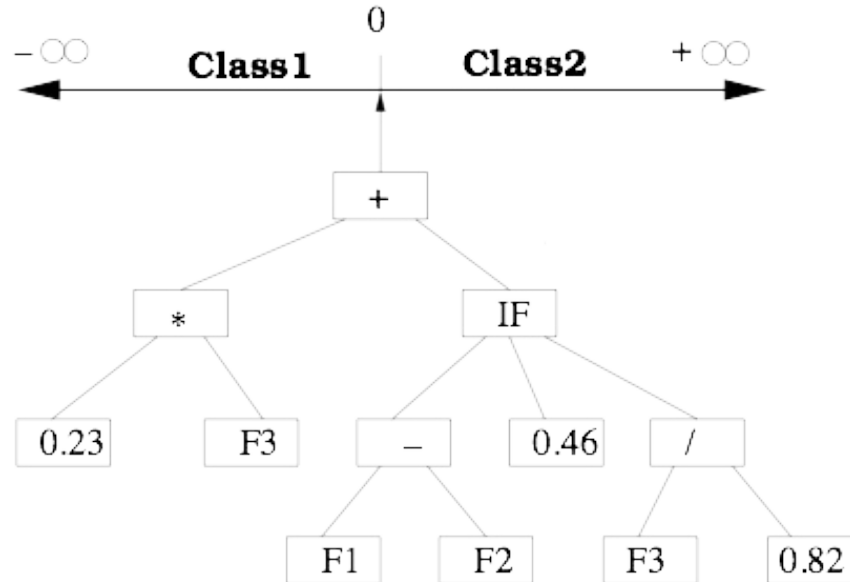


GP for Binary Classification

- Compared with GP for symbolic regression problems, the terminal set and function set can be the same or **very similar**, but the **fitness function is normally very different**
- For the **fitness function**, we can simply use **classification accuracy** or **error rate**.
- How do we determine **which class an example belongs to?**
 - **Classification Strategy or Program Class Translation Rule**
- For binary classification problems, this is quite easy: we can use the value “**zero**” in the real number space for the **program output** to separate the two classes

Program Class Translation Rule

- Is zero the best threshold?
- What if all features are positive?



Genetic Program: (+ (* 0.23 F3)
 (IF (- F1 F2) 0.46 (/ F3 0.82))
)

```
if ProgOut < 0 then Class1 else Class2;
```

Where have we seen this before?

GP for Classification Example

- **Task:** (Image) Object classification: *objects vs non-objects*
- **Objective:** Find a program which can successfully *split* the instances into *two* classes
- **Terminal Set:** Object *attributes*: pixels, pixel statistics, or specific features, and *random numbers*.
- **Function Set:** $\{+, -, \times, \div, |x|, e^x, \log x, \sin x, \cos x\}$
- **Fitness Cases:** Build a training set of *patterns (feature vectors)*, some are objects, some not.
- **Fitness Measure:** classification accuracy or error rate
- **Classification strategy:** ProgOut > 0 for objects, else non-objects

GP for Classification Example

F1	F2	F3	F4	Label
23	11	33	21	A
20	12	10	31	B

1. Terminal set: F1, F2, F3, F4 (or with constant)
2. Function Set: +, -, *, / (protected)
3. Fitness function: Accuracy
4. GP parameters: Population size, crossover, etc.
5. Termination: Generations, a good accepted accuracy

Summary

- GP for symbolic regression
- Properties of GP for symbolic regression
- GP for binary classification
- How do you use GP for multi-class classification? Can we get better translation rules?
- That's it for this half!



Summary

This Friday

- Go through part 2 of assignment 2
- Introduce possible ways to work on it easier, i.e., libraries
- How to write a good report