人工智慧理論與實務

基因演算法與LPPL

Linear Regression

- Signal generator
  - $F1(t) = 0.063\ t^3 - 5.284\ t^2 + 4.887\ t + 412 + noise$
- Problem settings
  - Input: a series of F1 signal with t in [0.0 100.0]
  - Prior knowledge: F1 is a linear equation of t
  - Goal: Reverse the original equation of F1(t)

Non-linear cases

- Signal generator
  - $F2(t) = 0.6\ t^{1.2} + 100\cos(0.4t) + noise;$
- Assume
  - Given: $F2(t) = A*t^{B} + C*\cos(D*t) + noise;$
  - Find the best parameters A,B,C, and D
- Fitness function
  - $Energy(A,B,C,D) = |\ F2(t) - (A*t^{B} + C*\cos(D*t))\ |$
- Exhaustive search
  - A = -5.11 : 0.01 : 5.12
  - B = -5.11 : 0.01 : 5.12
  - C = -511 : 512
  - D = -5.11 : 0.01 : 5.12

Exhaustive Search

- Experiment 1
  - Fix A,B,C to ground truth and estimate the fitness under different D settings
  - Plot the curve where Y axis is the Energy and X axis is the D value
- Experiment 2
  - Fix B,D and estimate the fitness under different combination of A and C settings
  - Plot the surface

Problem

- Exhaustive Search
  - It requires 2^{40} function calls
  - If the computational time of experiment 2 in previous slides is around 30 seconds, to examine 4 variables requires 364 days
- Solution
  - Model the candidate solution and apply evolutionary algorithm, such as genetic algorithm, to find the optimal solution.

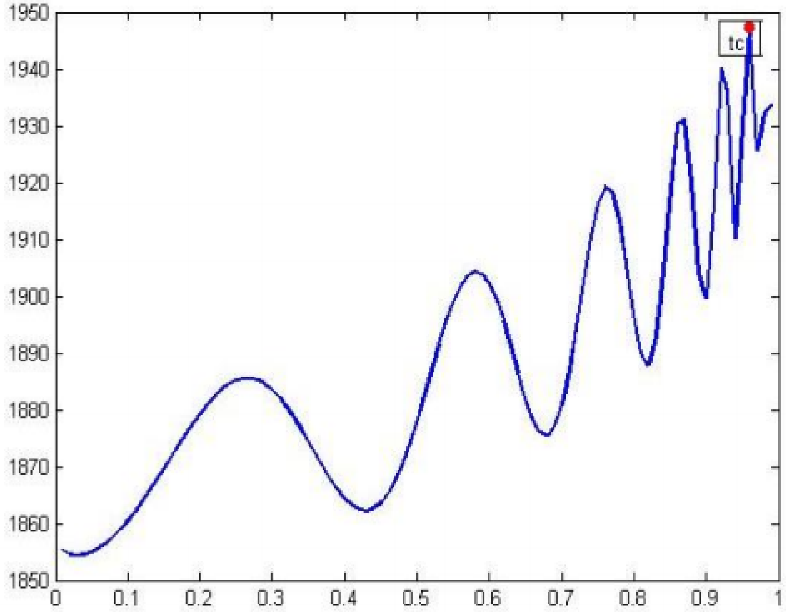
Genetic algorithm

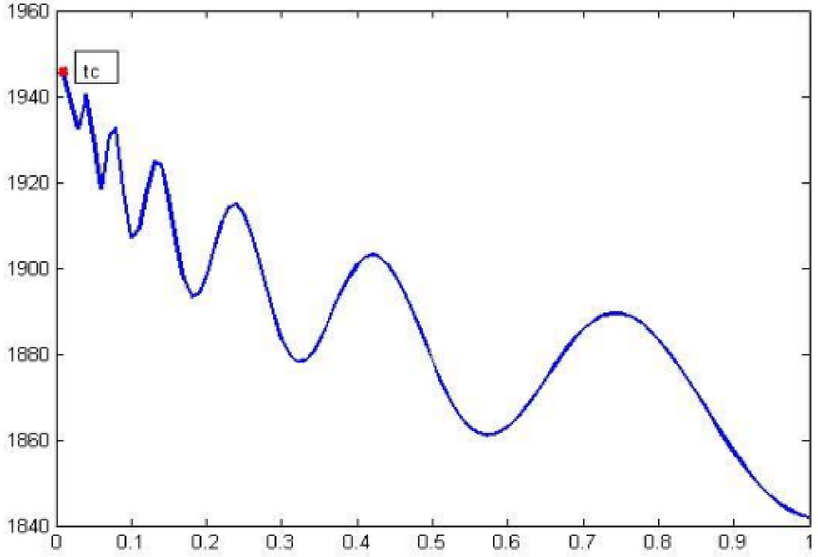
- 定義基因
  - For example, 在我們的問題中使用一組40 bits code來代表4個變數
- 初代
  - 利用亂數或expert knowledge產生一群初始的族群
- 複製 (reproduction)
  - 計算fitness
  - 利用fitness決定適者生存
  - 輪盤式選擇 (roulette wheel selection)
    - 依照fitness分割輪盤大小，面積比例越大越容易被選中
  - 競爭式選擇 (tournament selection)
    - 只留fitness最高的一小群人survive，淘汰適應不佳的

Genetic algorithm

- 交配 (crossover)
  - 單點交配
    - 此點以後的基因互換
  - 雙點交配
    - 兩點間的基因互換
  - 遮罩交配
    - 產生一個0/1 mask或filter，mask為1的bit互換
- 突變
  - 少數bit 0->1或1->0

Exercise



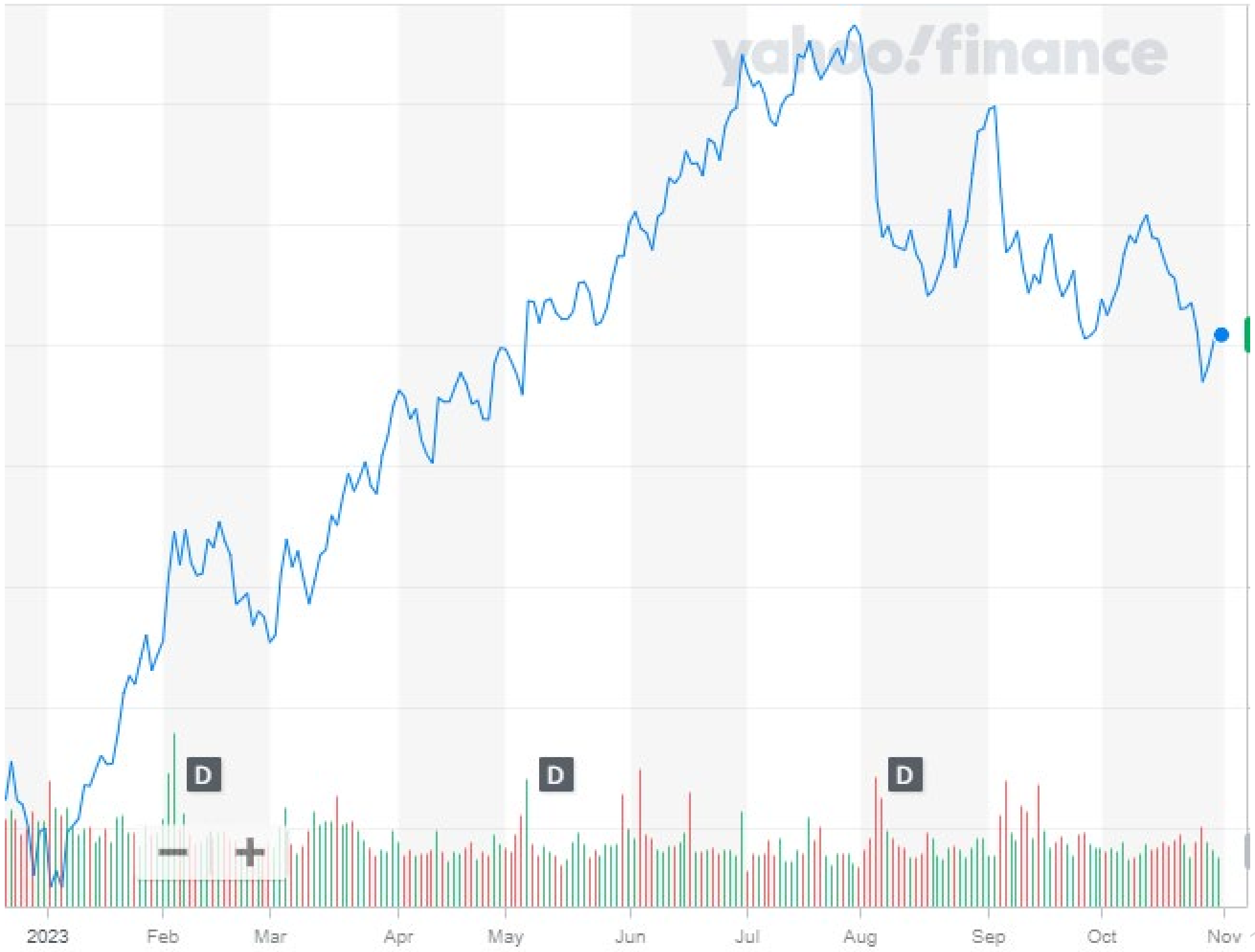


圖一 泡沫

圖二 反泡沫

資料來源：國泰君安證券研究所

2023 apple股價



log-periodic power laws (LPPL) for bubble modeling

$$\ln[p(t)] \approx A + B(t_c - t)^{\beta}\{1 + C\cos[\omega\ln(t_c - t) + \phi]\}, \tag{12}$$

where $A > 0$ is the value of $[\ln p(t_c)]$ at the critical time, $B < 0$ is the increase in $[\ln p(t)]$ over the time unit before the crash if C were to be close to zero, $C \neq 0$ is the proportional magnitude of the oscillations around the exponential growth, $0 < \beta < 1$ should be positive to ensure a finite price at the critical time t_c of the bubble and quantifies the power law acceleration of prices, and ω is the frequency of the oscillations during the bubble, while $0 < \phi < 2\pi$ is a phase parameter. Expression (12), which is known as the LPPL, is the fundamental equation that describes the temporal growth of prices before a crash and it has been proposed in different forms in various papers (e.g. Sornette 2003a, Lin, Ren, and Sornette 2009 and references therein). We remark that A, B, C and ϕ are just units distributions of betas and omegas, as described in Sornette and Johansen (2001) and Johansen (2003), and do not carry any structural information.

作業題目

- 請先到yahoo finance下載2023年APPL的historical data
- 試以2023年每日的調整過後股價為真實資料
- 假設我們已知t_c發生在7/24~8/4間(含)
- 試求出一組LPPL模型參數，可讓模型產生的資料與真實資料間，擁有最小的mean squared error

- Two step algorithm
  - Each gene includes 4 non-linear variables t_c, β, ω, Φ
  - Use linear regression to estimate best A, B, C
- For each parameter setting, we can measure the fitness between synthetic signals and real financial time-series data.
- Apply genetic algorithm to approximate the optimal solution by minimizing the average fitness error between time 0 and t_c.
- Homework:
  - LPPL
    - Find the optimal LPPL parameters, suppose t_c in given range
    - Plot the synthetic signals and real time-series data with different colors in a figure
    - Calculate the mean squared error between your model and real data