國立臺灣科技大學
NATIONAL TAIWAN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Class 4 - 時域訊號特徵抽取演算法

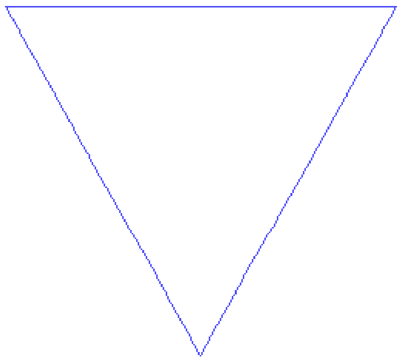
課程設計：蕭玉聰
指導教授：劉益宏 老師

日期：2021/8/27

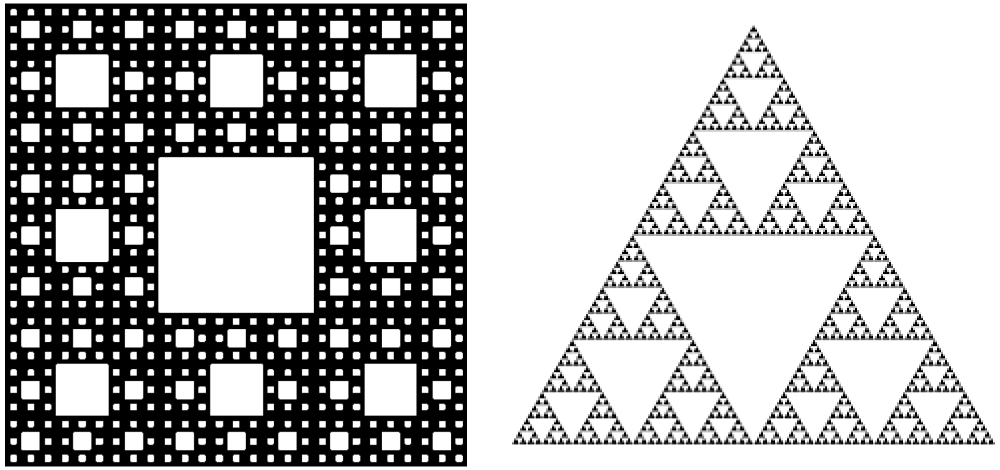
神經工程及智慧系統實驗室
Neural Engineering & Smart System Lab , NESS Lab

碎形維度 (Fractal Dimension)

- 碎形，又稱分形、殘形，通常被定義為「一個粗糙或零碎的幾何形狀，可以分成數個部分，且每一部分都（至少近似地）是整體縮小後的形狀」，即具有自相似的性質。
- 碎形維度是一種量測複雜度的方法，通過在某個時間間隔內估計自我相似度來測量該時間序列的複雜度。



科赫雪花(Koch snowflake)



謝爾賓斯基碎形(Sierpinski FD)

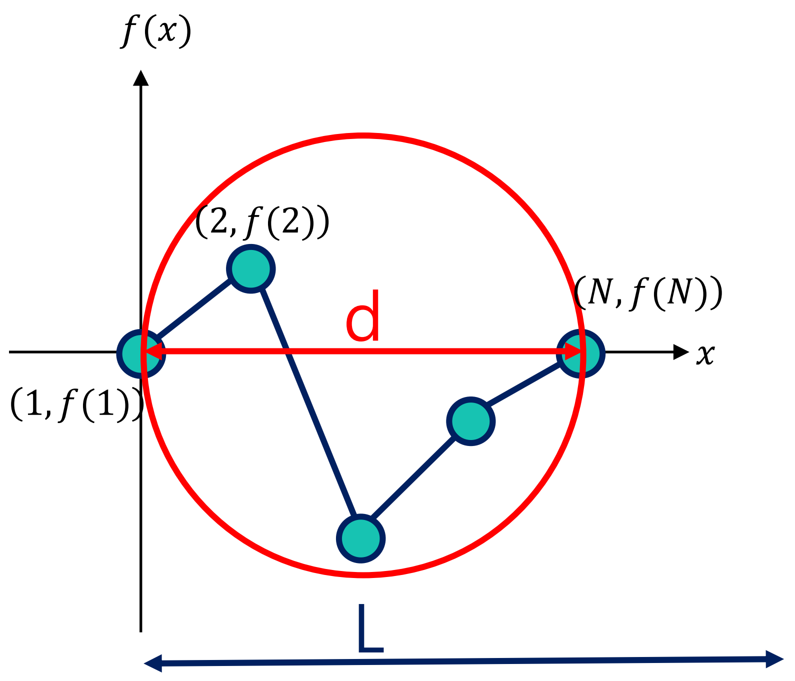
Katz Fractal Dimension

- 計算時間序列中，各資料點之歐式距離，以距離總長及最大直徑評估資料複雜度。

$$D = \frac{\log_{10}(L)}{\log_{10}(d)}$$

$L = sum\big(dist(i, i+1)\big)$，$i = 1,2 \ldots N-1$

$d = \max(dist(1, i))$



L ：時間序列的總長度
d ：時間序列的最大直徑
N ：時間序列的資料筆數
註:log以10為底

Katz Fractal Dimension
Normalization

- 將總長與直徑除以平均長度，進行正規化。

$$D = \frac{\log(L/a)}{\log(d/a)} \longrightarrow D = \frac{\log(n)}{\log(d/L) + \log(n)}$$

n：$\frac{L}{a}$，時間序列的間距數量
a：相鄰資料點間平均距離

Higuchi Fractal Dimension

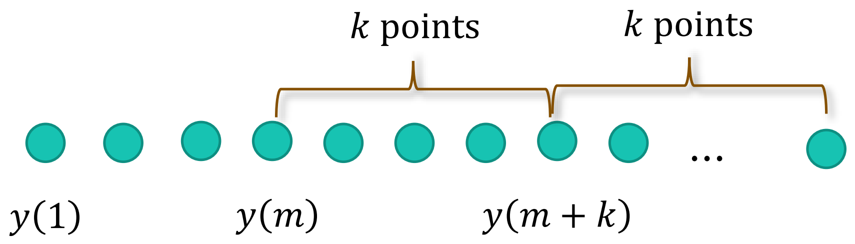
- 計算時間序列碎形維度的一種方式，能為較短的時間序列訊號提供合理的FD估計。

- 對於長度為N的訊號：$\mathbf{x} = [\mathrm{x}(1), \mathrm{x}(2), \dots, \mathrm{x}(N)]$

$$\boldsymbol{y}_m^k = \left[y(m), y(m+k), y(m+2k), \dots, y\left(m + int\left(\frac{N-m}{k}\right)k\right)\right], m = 1,2,\dots,k$$

m:每次重新採樣的初始時間點
k :時間間格

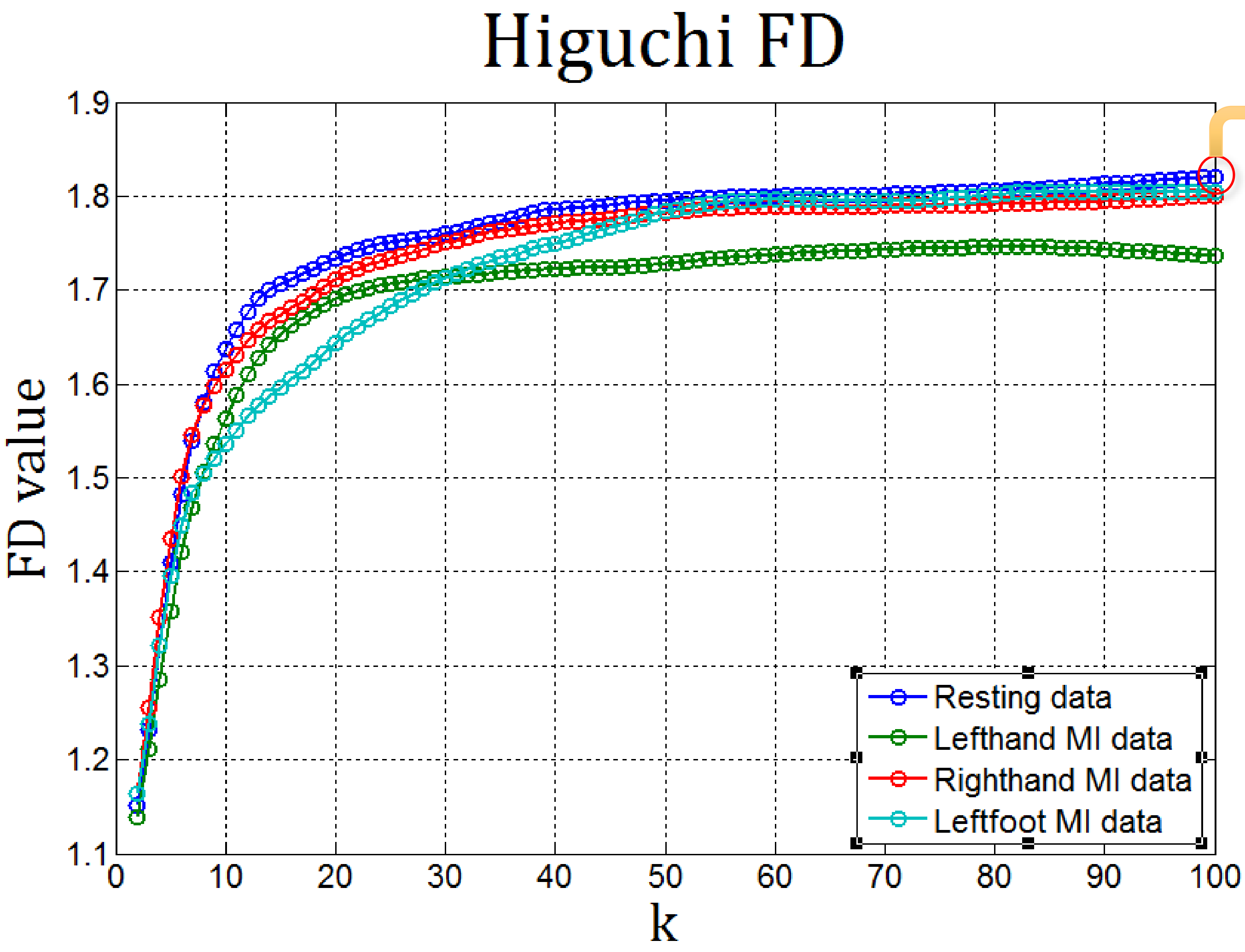
$$L_m(k) = \frac{1}{k}\left[\frac{N-1}{int\left(\frac{N-m}{k}\right)k}\left(\sum_{i=1}^{int\left(\frac{N-m}{k}\right)} |y(m+ik) - y(m+(i-1)k)|\right)\right]$$

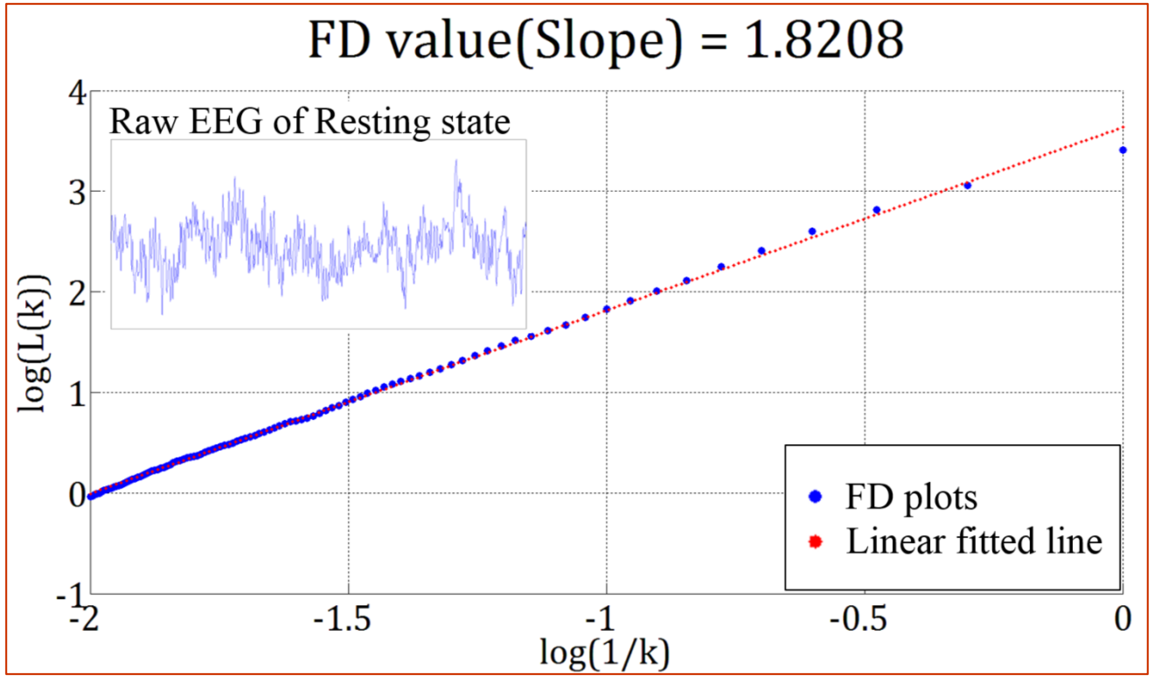


$$L(k) = \frac{1}{k}\sum_{m=1}^{k} L_m(k) \text{，} k = 1,2 \dots k_{max}$$

$\frac{N-1}{int\left(\frac{N-m}{k}\right)k}$:Normalize factor

HFD
Example





K_{max} : 100

熵(Entropy)

- 熵是用來描述混亂的程度(亂度)的一種方法，自然界必定趨向於於最大亂度與最低能量，因此是持續散亂的過程；若越來越亂則稱之為「熵增」，若越來越有條理則是「熵減」。
- 熵是一種機率問題，高熵狀態的可能數量遠高於低熵狀態(混亂機率遠高於有條理機率)。
- 熵越高代表亂度越高，反之熵越低代表亂度越低

樣本熵(Sample Entropy)

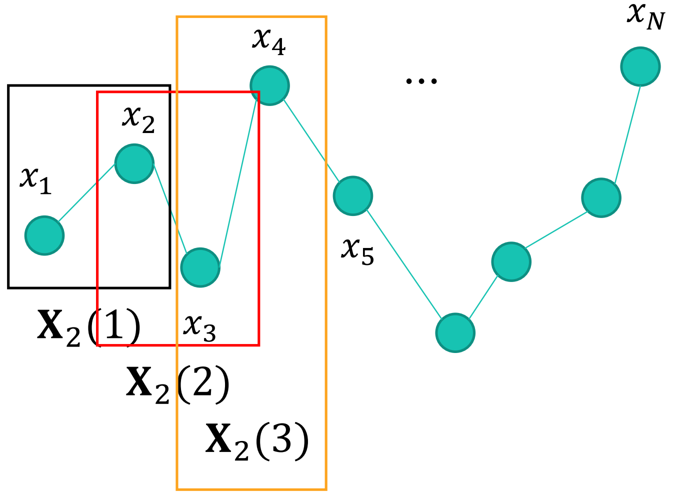
- 是一種利用時間序列計算亂度的方法，計算結果值越高代表亂度越高，反之則越低。
- 假設一長度為N的時序訊號為 $\mathbf{X} = \{x_1, x_2, \ldots, x_N\}$ ，將連續m個點取樣成一個新的時序訊號為 $\mathbf{X}_m$
- 設定參數 : m嵌入維度(dimensional reconstruction vector):通常取2或3。

$\mathbf{X}_m(i) = \{x_i, x_{i+1}, \ldots, x_{i+m-1}\}$ ， $1 \leqq i \leqq N - m$，代表第i個新時序訊號。

時序訊號之取樣熵計算如下:

ex:設$\mathrm{m} = 2$時，第一個新時序取樣
$\mathbf{X}_2(1) = \{x_1, x_2\}$

m=2時的新取樣



樣本熵(Sample Entropy)

切比雪夫距離:
假設以(x_1, y_1)和(x_2, y_2)二點為例，其切比雪夫距離為$\max(|x_2 - x_1|, |y_2 - y_1|)$

(1) 以$\mathbf{B}_i = \mathrm{d}[\mathbf{X}_m(i), \mathbf{X}_m(j)] = \boxed{\max(|\mathbf{X}_m(i) - \mathbf{X}_m(j)|)}$，$i \neq j$，計算時序訊號與新時序訊號之間的距離，取計算距離最大者，並計算滿足不等式 $d[X_m(i), X_m(j)] \leqq r$之數量，

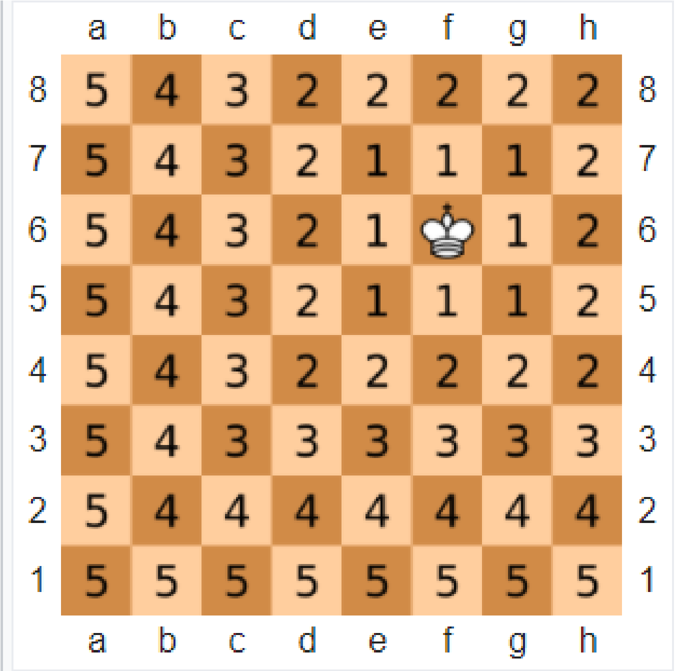
r(兩個樣本距離閥值)=k*std($\mathbf{X}$)，std =標準差， k 介於0~1之間。

(2) $\mathbf{P}_i = \dfrac{\mathbf{B}_i}{N-m-1}$， $\mathbf{P}_i = \mathbf{B}_i$之平均值

(3) $\mathbf{B}_m = \dfrac{1}{N-m} \sum_{i=1}^{N-m} \mathbf{P}_\mathrm{i}$

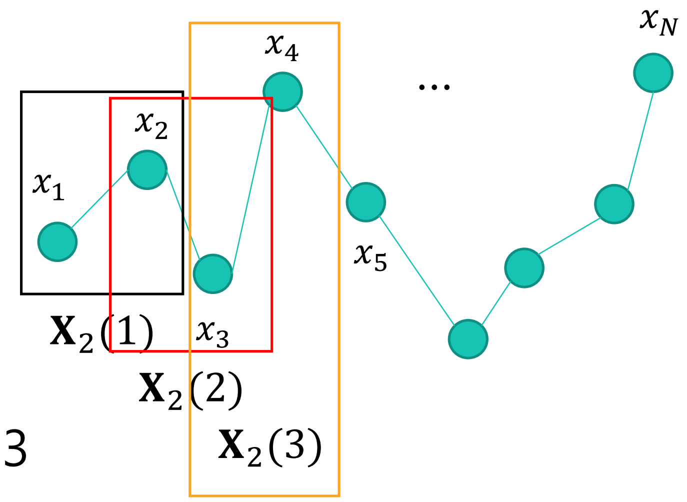
(4)重複(1)~(3)步驟，其中把m設定為 m+1，獲得$\mathbf{B}_{m+1}$

(5) $SampEn = -ln \dfrac{\mathbf{B}_{m+1}}{\mathbf{B}_m}$



若將西洋棋棋盤放在二維直角座標系中，格子的邊長定義為1，座標的x軸及y軸和棋盤方格平行，王從一個位置走到其他位置需要的步數恰為二個位置的切比雪夫距離，因此切比雪夫距離也稱為**棋盤距離**

樣本熵舉例



設$\mathbf{X} = \{x_1, x_2, \ldots, x_{10}\}$，$\mathrm{m} = 2$，$r = \mathrm{k} * std(\mathbf{X})$

m=2

$\mathbf{X}_2(1)$={x_1, x_2}

$\mathbf{X}_2(2)$={x_2, x_3}

⋮

$\mathbf{X}_2(9)$={x_9, x_{10}}

$$\mathrm{d}[\mathbf{X}_2(1), \mathbf{X}_2(2)] = \max(|\mathbf{X}_2(2) - \mathbf{X}_2(1)|) = \max(|x_2 - x_1|, |x_3 - x_2|)$$

共有$C_2^9 = 36$個d值,假設其中有2個$\leqq r$，分別是$\mathbf{B}_2 = d(2)$、$\mathbf{B}_3 = d(3)$。

m=m+1 =3

$\mathbf{X}_3(1)$={x_1, x_2, x_3}

$\mathbf{X}_3(2)$={x_2, x_3, x_4}

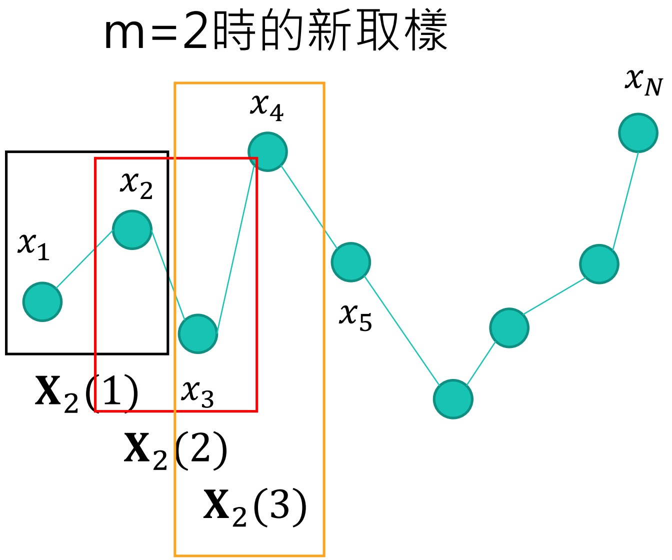
⋮

$\mathbf{X}_3(8)$={x_8, x_9, x_{10}}

$$\mathrm{d}[\mathbf{X}_3(1), \mathbf{X}_3(2)] = \max(|x_2 - x_1|, |x_3 - x_2|, |x_4 - x_3|)$$

共有$C_2^8 = 28$個d值,假設其中有2個$\leqq r$，分別是$\mathbf{B}_4 = d(4)$、$\mathbf{B}_5 = d(5)$。

樣本熵舉例



$\mathbf{P_2} = \frac{\mathbf{B_2}}{N-m-1}$, $\mathbf{P_3} = \frac{\mathbf{B_3}}{N-m-1}$

$\mathbf{B}_{m=2} = \frac{1}{N-m}(\mathbf{P_2} + \mathbf{P_3})$

$\mathbf{P_4} = \frac{\mathbf{B_4}}{N-m-1}$, $\mathbf{P_5} = \frac{\mathbf{B_5}}{N-m-1}$

$\mathbf{B}_{m=3} = \frac{1}{N-m}(\mathbf{P_4} + \mathbf{P_5})$

$$\mathrm{SampEn}(\mathbf{x}, \mathrm{m}, \mathrm{r}) = -ln\frac{\mathbf{B_3}}{\mathbf{B_2}}$$

$$\mathbf{P_i}(m,r) = \frac{\mathbf{B_i}}{N-m-1} \quad \dots$$

$$\mathbf{B_m}(r) = \frac{1}{N-m}\sum_{i=1}^{N-m}\mathbf{P_i}$$

Q&A