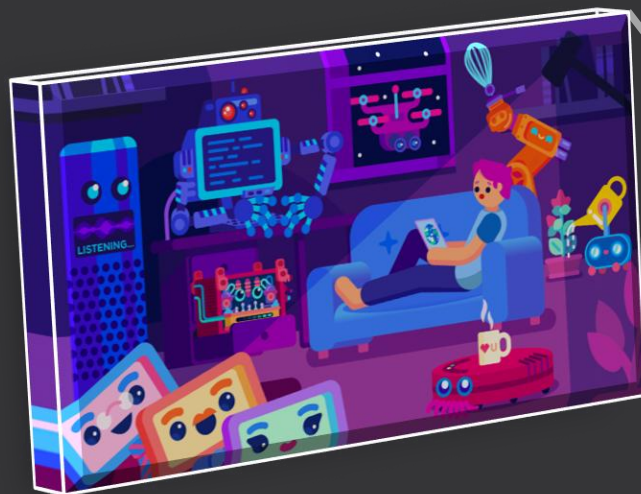
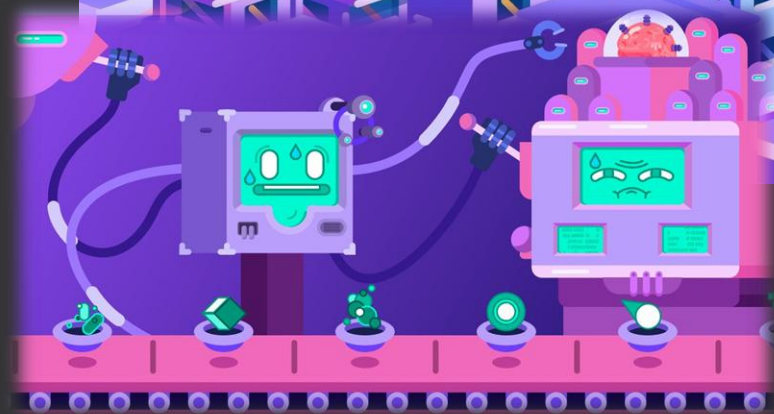




Metody i narzędzia *Big Data*

Agregaty – odpowiednia konstrukcja
algorytmu



Średnia i wariancja

Średnia i wariancja na bieżąco – przyrostowo, dla całego sygnału

Algorytmy do liczenia średniej i wariancji „w locie” mają ogólną postać

$$m_N = \Psi(m_{N-1}, x_N)$$

$$v_N = \Phi(v_{N-1}, m_N, x_N)$$

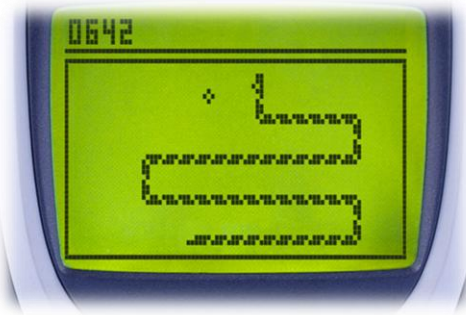
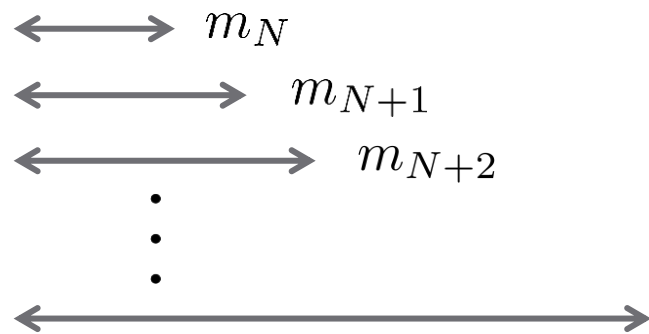
A konkretnie

$$m_N = \frac{N-1}{N}m_{N-1} + \frac{1}{N}x_N$$

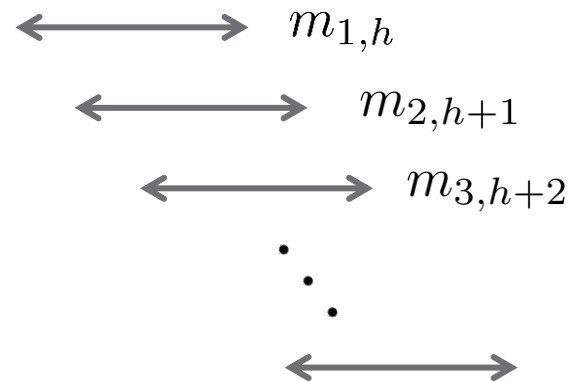
$$v_N = \frac{N-1}{N}v_{N-1} + \frac{1}{N-1}(m_N - x_N)^2$$

Przyrostowo vs okno przesuwne

Przyrostowo



Okno przesuwne



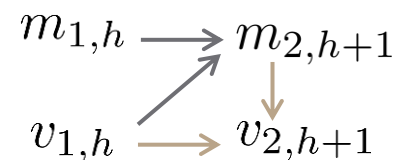
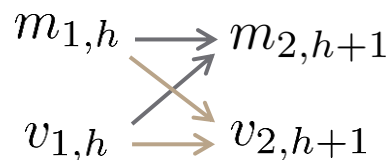
Średnia i wariancja na bieżąco – na oknie przesuwnym



Algorytmy do liczenia średniej i wariancji „w locie” mają ogólną postać

$$m_{2,h+1} = \Psi(m_{1,h}, x_{h+1}, x_1)$$

$$v_{2,h+1} = \Phi(v_{1,h}, m_{1,h}, x_{h+1}, x_1) \quad \text{lub} \quad v_{2,h+1} = \Phi(v_{1,h}, m_{2,h+1}, x_{h+1}, x_1)$$



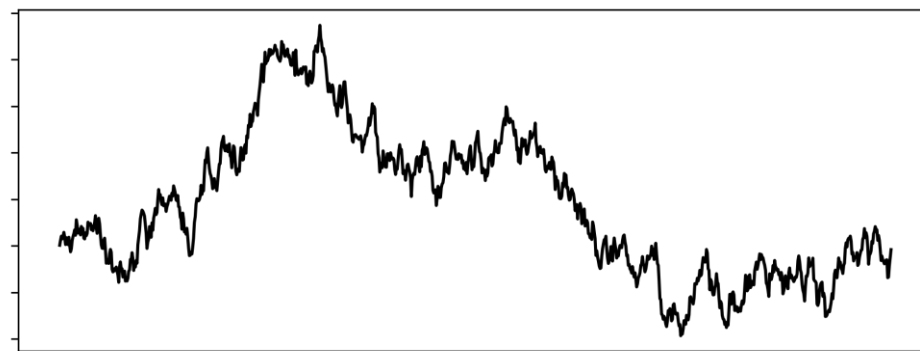
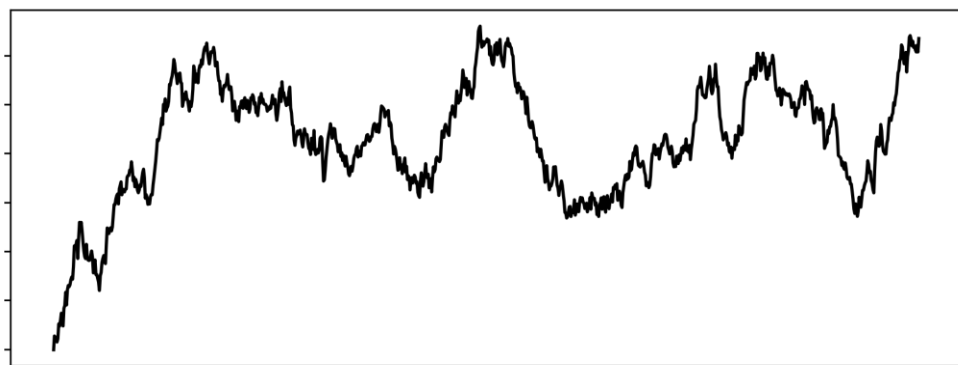
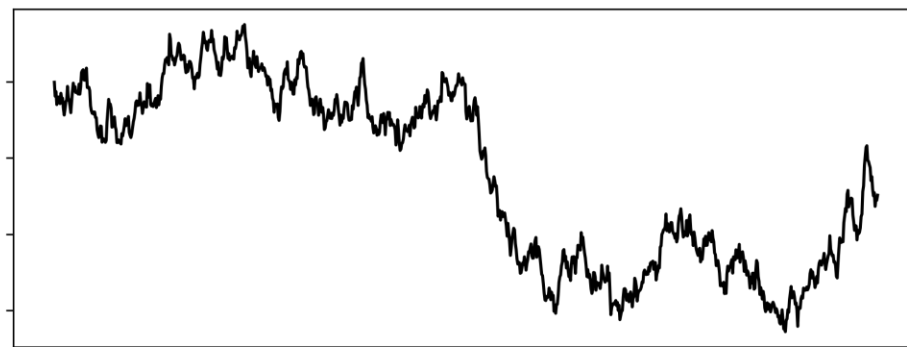
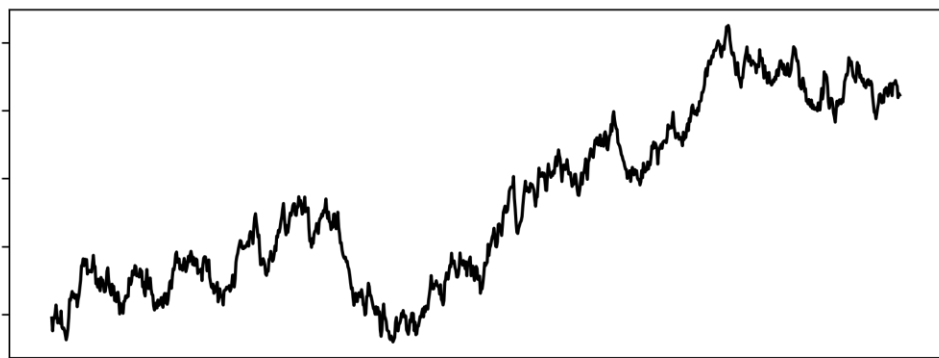
A konkretnie

$$m_{2,h+1} = m_{1,h} + \frac{1}{h}(x_{h+1} - x_1)$$

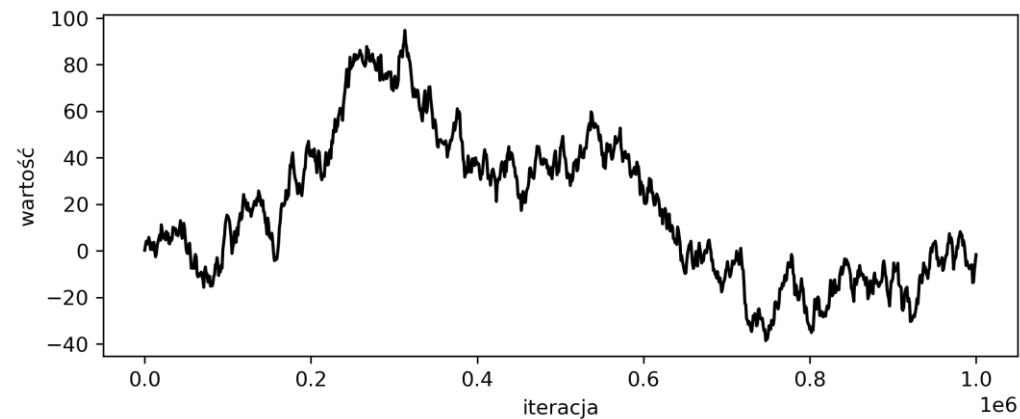
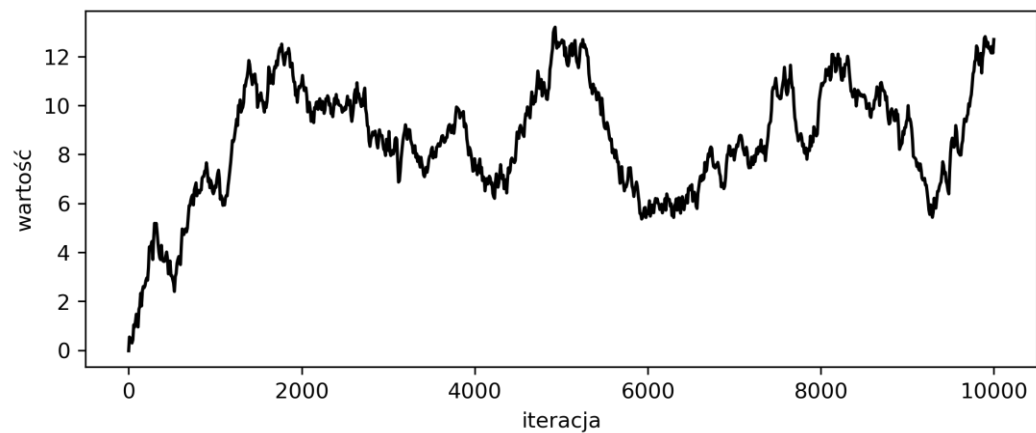
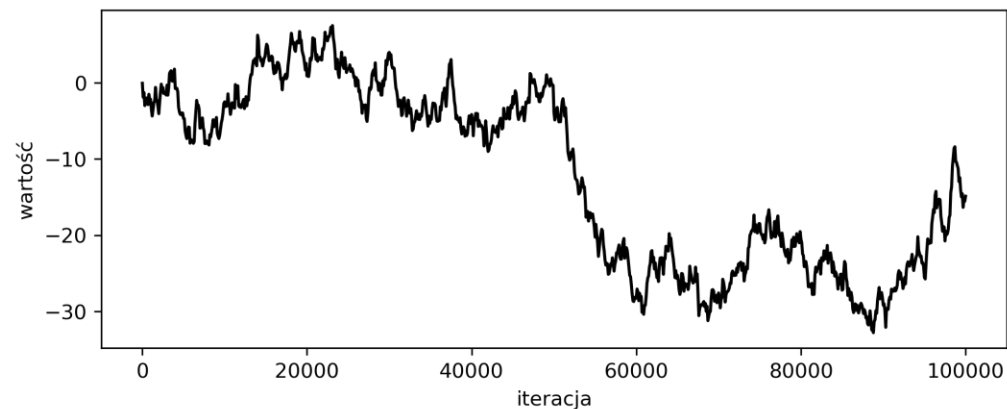
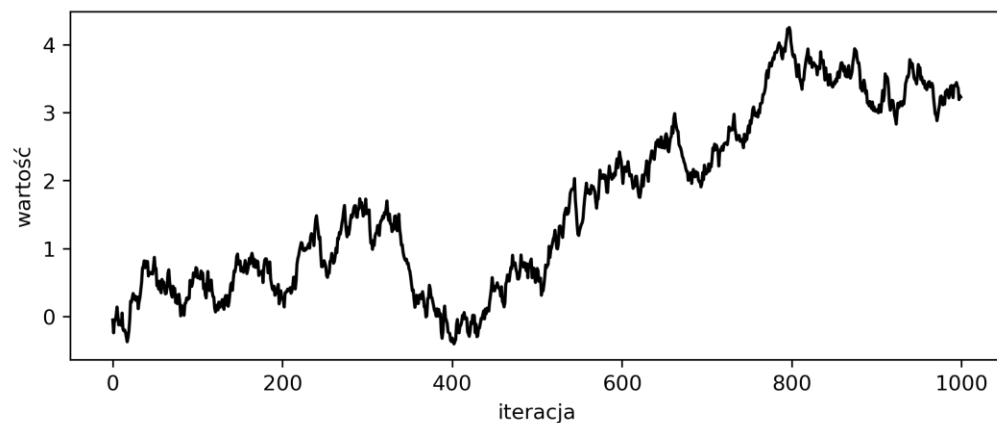
$$v_{2,h+1} = v_{1,h} - \frac{1}{h-1}(m_{1,h} - x_1)^2 + \frac{1}{h-1} \left(m_{1,h} - \frac{1}{h}x_1 + \frac{1-h}{h}x_{h+1} \right)^2$$

Agregaty szeregu czasowego

Wykładnik Hursta



Skala procesu



Wykładnik Hursta

$$\{x_i\}_{i=1}^N$$



$$X_N = \sum_{i=1}^N x_i$$



$$R/S(N) = \frac{\max_{1 \leq i \leq N} \left\{ X_i - \frac{i}{N} X_N \right\} - \min_{1 \leq i \leq N} \left\{ X_i - \frac{i}{N} X_N \right\}}{\sqrt{\frac{1}{N} \sum_{i=1}^N \left(x_i - \frac{1}{N} X_N \right)^2}}$$



$$R/S(N) \propto N^H$$

<https://pypi.org/project/nolds/>

<https://github.com/Mottl/hurst>

Stochastic Models That Separate Fractal Dimension and Hurst Effect, Tilmann Gneiting and Martin Schlather (2001)

Qian, B., & Rasheed, K. (2004). *Hurst exponent and financial market predictability*. Proceedings of the IASTED International Conference Cambridge, MA.

Graves, T., Gramacy, R., Watkins, N., & Franzke, C. (2017). *A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA*. Entropy, 19(9).

Wykładnik Hursta

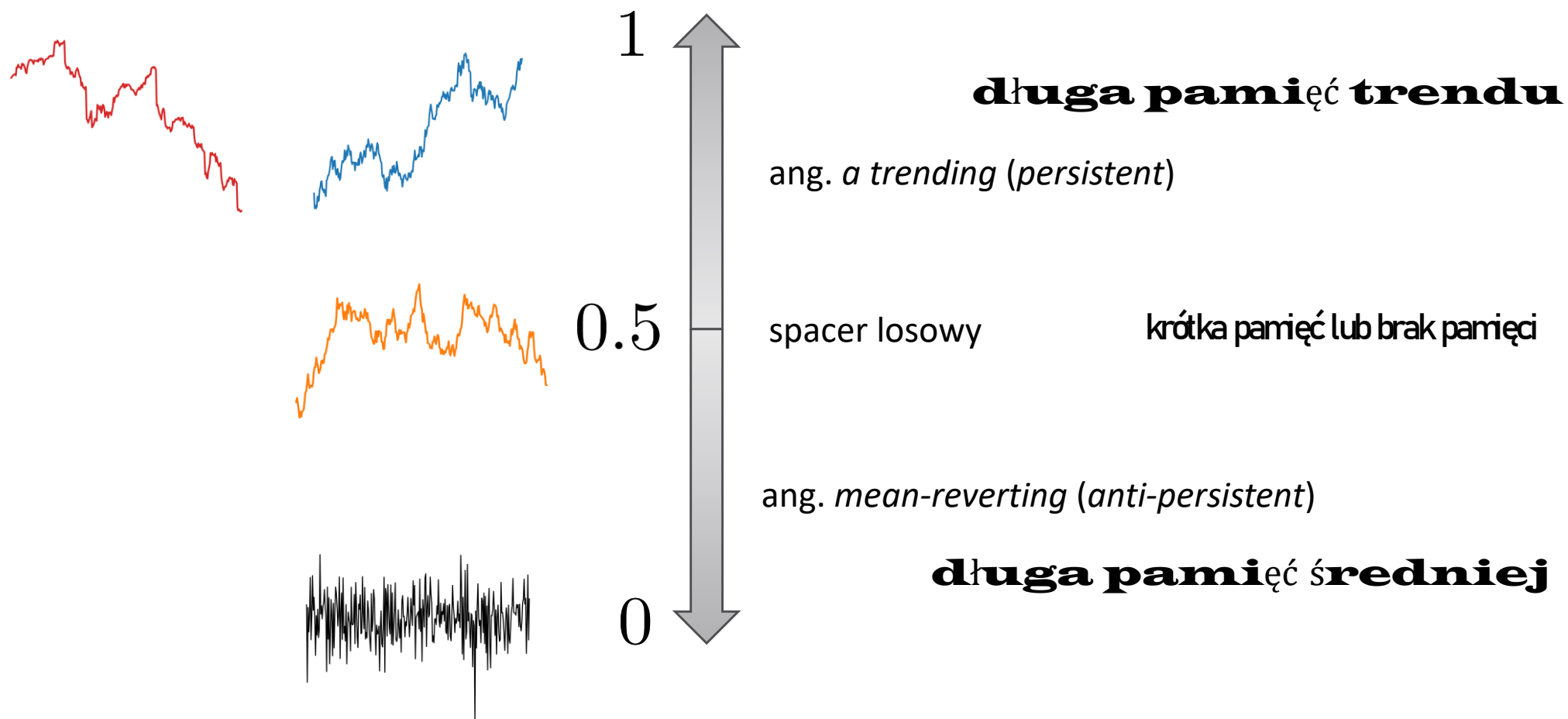
Inny sposób

$$y_i = \log x_i$$

$$V(\tau) \equiv \text{Var} \left(\left\{ y_{i+\tau} - y_i \right\}_{i=1}^{N-\tau} \right)$$

$$V(\tau) \propto \tau^{2H}$$

Jak interpretować H ?

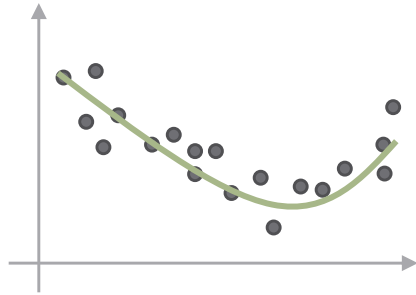


Agregaty szeregu czasowego

Interpolacja

Aproksymacja

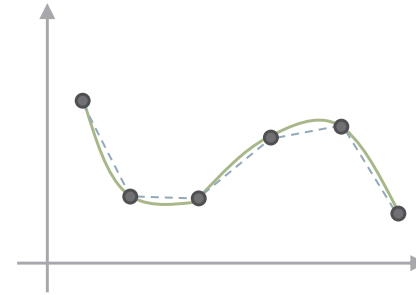
x_1	x_2	$\cdots$	x_N
y_1	y_2	$\cdots$	y_N



INTERPOLACJA

x_1	x_2	$\cdots$	x_N
y_1	y_2	$\cdots$	y_N

$$x_0 < x_1 < \cdots < x_N$$

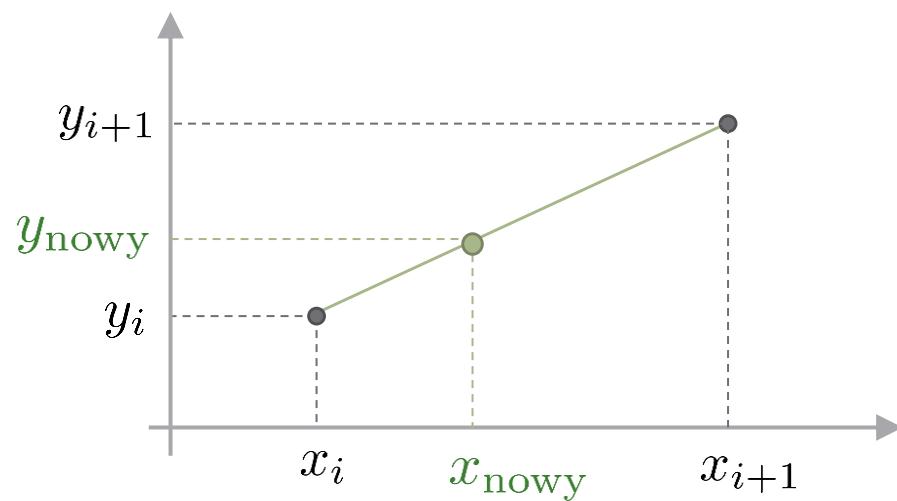


Wymóg: model dokładnie odtwarza tabelę

$$y_i = \text{model}(x_i)$$

- – węzeł (ang. *knot*)

Po co nam model interpolacyjny?



Model nie musi być jawny

$$\frac{y_i - y_{i+1}}{x_i - x_{i+1}} = \frac{y_i - y_{\text{nowy}}}{x_i - x_{\text{nowy}}}$$

$$y = ax + b$$

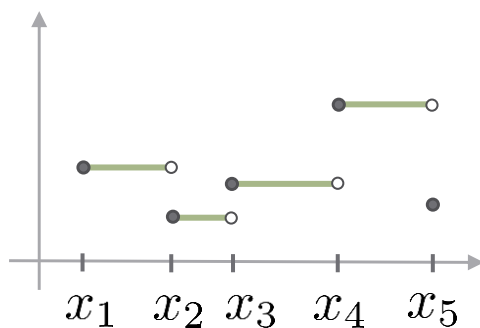
$$a = ? \quad b = ?$$

Modele interpolacyjne

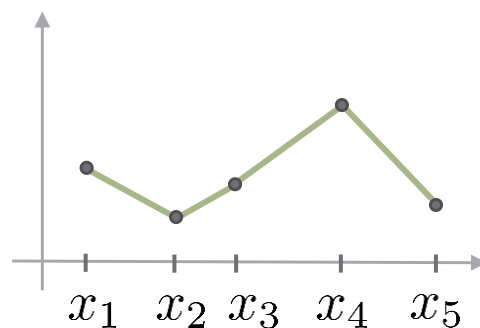
Model łączy tylko sąsiednie węzły

Splajny (funkcje sklejące, ang. *splines*)

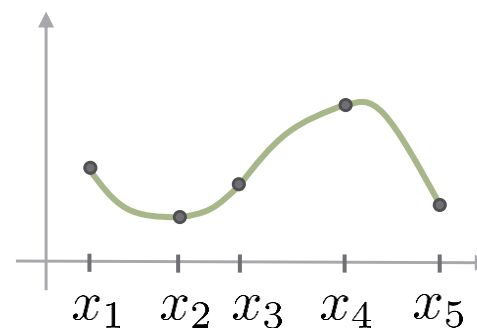
splajn 0 stopnia



splajn 1 stopnia



splajn 3 stopnia



Modele interpolacyjne

Model łączy tylko sąsiednie węzły

Splajn k -tego stopnia:

- w przedziałach $\langle x_i, x_{i+1} \rangle$ jest wielomianem k -tego stopnia
- ma ciągłą $(k-1)$ -wszą pochodną na przedziale $\langle x_0, x_N \rangle$

$$\bullet \quad S(x) = \begin{cases} S_0(x) & \text{dla } x \in \langle x_0, x_1 \rangle \\ S_1(x) & \text{dla } x \in \langle x_1, x_2 \rangle \\ \vdots & \\ S_{N-1}(x) & \text{dla } x \in \langle x_{N-1}, x_N \rangle \end{cases}$$

- przyjmuje się zwykle, że $S(x) = S_0(x)$ dla $x < x_0$
 $S(x) = S_{N-1}(x)$ dla $x > x_N$

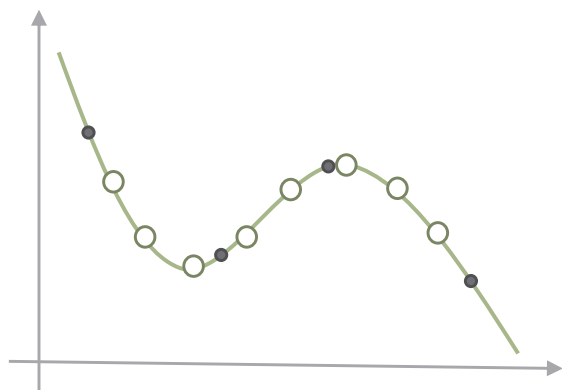
splajn kubiczny

$$S''(x_0) = S''(x_N) = 0$$

Modele interpolacyjne

Pojedynczy model dla wszystkich węzłów

np. wielomian N -tego stopnia



Reprezentacja Newtona

$$g(x) = \sum_{i=0}^N c_i \prod_{j=0}^{i-1} (x - x_j)$$

$c_i = ?$

Ilorazy różnicowe

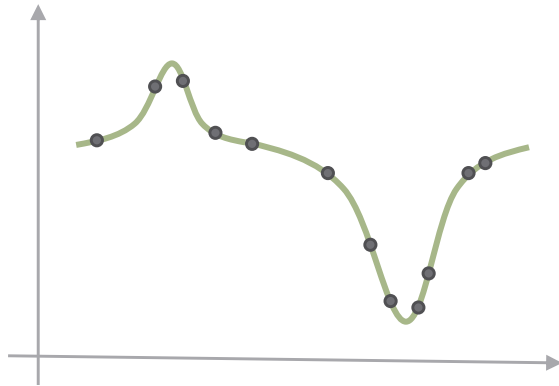
Reprezentacja Lagrange'a

$$g(x) = \sum_{i=0}^N y_i \prod_{\substack{j=0 \\ i \neq j}}^N \frac{x - x_j}{x_i - x_j}$$

Modele interpolacyjne

Pojedynczy model dla wszystkich węzłów

Uogólnione funkcje liniowe



$$g(\mathbf{x}) = \sum_{i=0}^N c_i \phi_i(\mathbf{x})$$

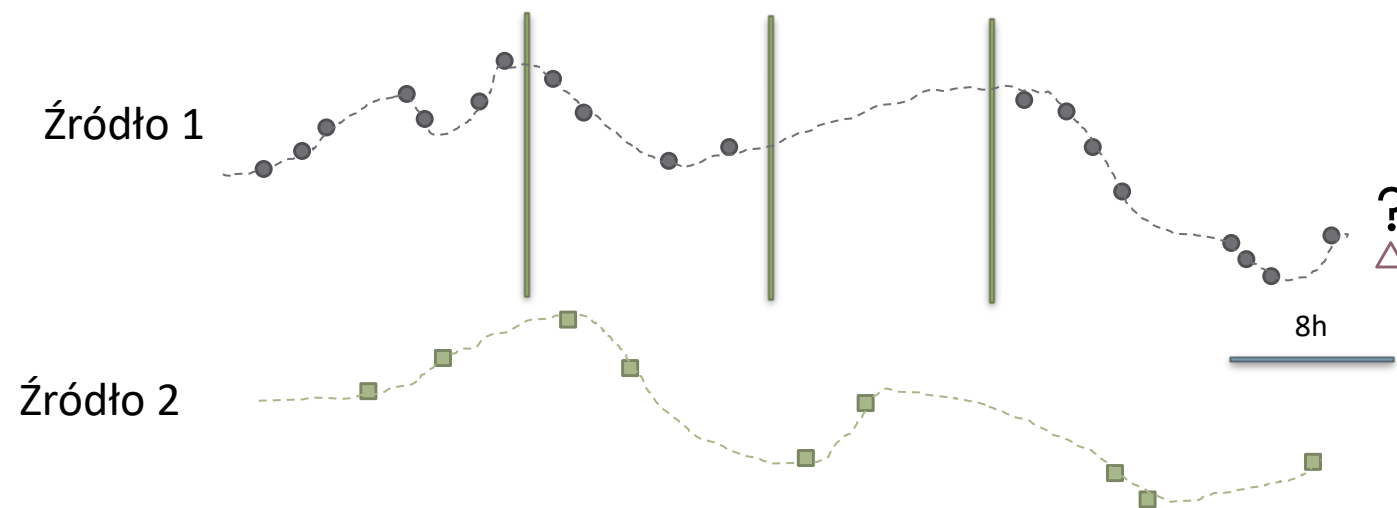
ϕ_i – funkcje bazowe

$c_i = ?$

Przykłady zestawów funkcji bazowych

- jednomiany
- wielomiany Czebyszewa
- szereg Fouriera

Łączenie szeregów czasowych z różnych źródeł



Co nam daje interpolacja?