SAVITZKY - GOLAY SMOOTING

Il modo più semplice per fare un algoritmo di smoothing è quello di rimpiazzare ogni punto con una combinazione lineare fatta dai punti del vicinato del punto target oltre che il punto stesso:

$$S_j = \sum_{n=-n_L}^{n_R} C_n Y_{j+n}$$

Dove nR e nL sono il numero di punti alla destra e alla sinistra del punto target.
Per capire meglio il filtro Savitzky-Golay consideriamo la più semplice procedura, usando S_j con i coefficienti C_n che vengono definiti dalla formula:

$$C_n = \frac{1}{n_L + n_R + 1}$$

Questo fa si che tutti i punti hanno egual peso nel calcolo della media-mobile.

Esempio:

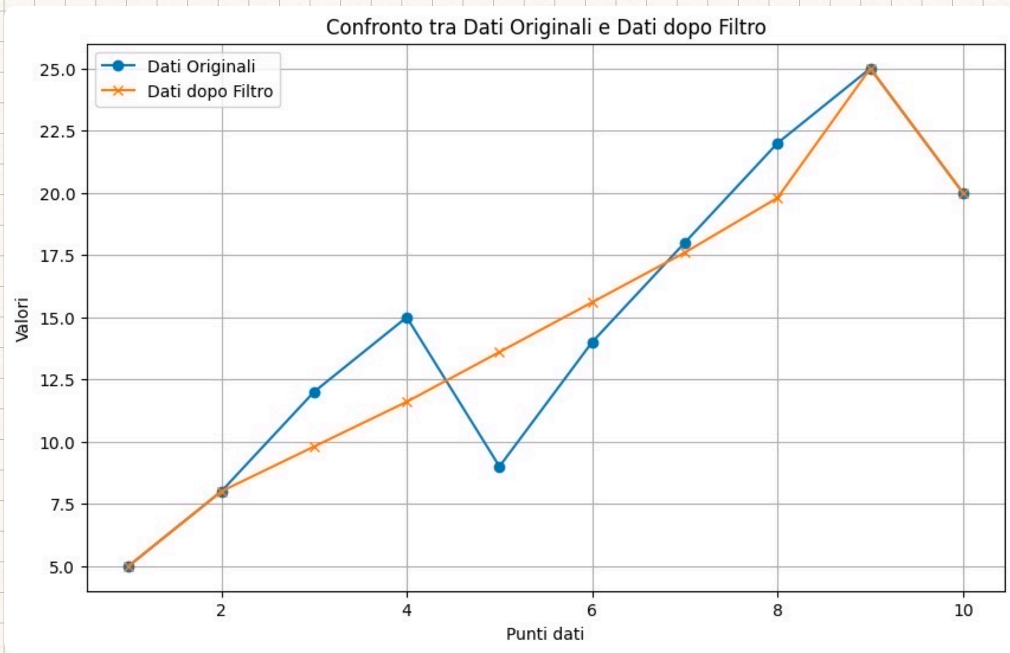
$$Y = [5, 8, 12, \overset{S_4}{15}, 9, 14, 18, 22, 25, 20]$$

$$C_n = \frac{1}{2+2+1} = \frac{1}{5} = C_{-2} = C_{-1} = C_0 = C_1 = C_2$$ } Perché usiamo sempre una finestra fatta da 5 p.

$$S_4 = \underbrace{\frac{1}{5} \cdot 8}_{C_{-2}} + \underbrace{\frac{1}{5} \cdot 12}_{C_{-1}} + \underbrace{\frac{1}{5} \cdot 15}_{C_0} + \underbrace{\frac{1}{5} \cdot 9}_{C_1} + \underbrace{\frac{1}{5} \cdot 14}_{C_2} = 11{,}6$$

Facendo tutti i calcoli esce:

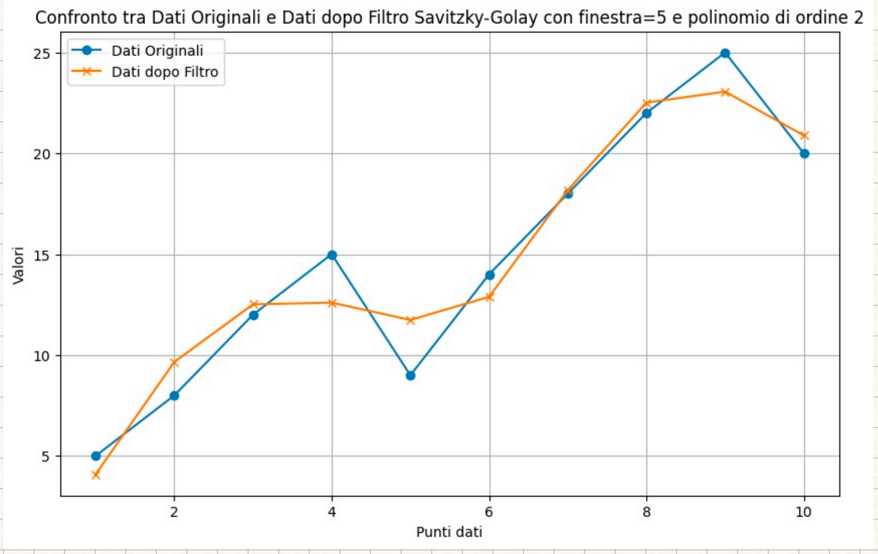
$$Y_{new} = [5 | 8 | 9{,}8 | 11{,}6 | 13{,}6 | 15{,}6 | 17{,}6 | 19{,}8 | 25 | 20]$$



Se la funzione sottostante è costante o cambia linearmente non viene introdotto nessun bias nella curva risultante.

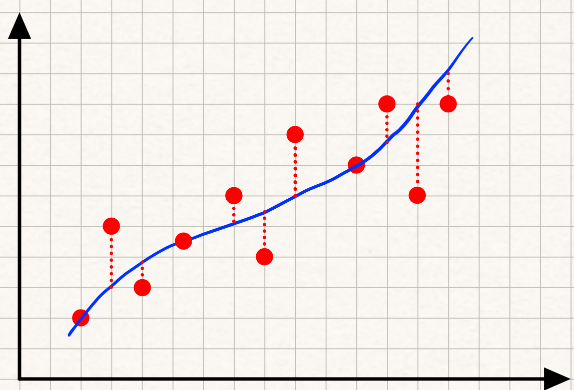
Sicuramente viene introdotto un bias se la funzione sottostante ha una derivata seconda in quel punto NON-ZERO. Per esempio un massimo locale nella curva risultante utilizzando la media dentro una finestra mobile, ridurrà il valore del massimo (Meno altezza ma più larghezza)

Questa scelta di coefficienti non è la migliore. L'idea di Savitzky-Golay è quella di inserire coefficienti che preservino meglio il trend. Per questo si usano polinomi di alto ordine, tipicamente di secondo o quarto ordine.
Quindi per ogni punto Y_j viene proposto un adattamento ai minimi quadrati trovati nella finestra. Il punto di smoothing viene sostituito col valore del polinomio in quella posizione. Questo polinomio viene usato solo per questo punto, scorrendo la finestra andrà ricalcolato un nuovo polinomio
Per calcolare tutti questi adattamenti ai minimi quadrati basta invertire una matrice lineare.
I coefficienti NON DIPENDONO dai dati, ma solo dalla posizione dei punti nella finestra. Possiamo quindi calcolare tutti i coefficienti in anticipo con dati fittizzi consistenti in tutti $\emptyset$ tranne per un solo 1.
Per fare il fit sui dati reali basta semplicemente calcolare una combinazione lineare dei coefficienti trovati.



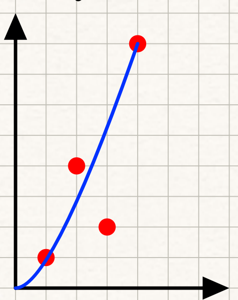
MINIMI QUADRATI (LEAST SQUARES)

Metodo di interpolazione che cerca di fare l'interpolazione di punti nello spazio in modo tale da minimizzare la somma dei quadrati delle differenze fra i nostri punti e il polinomio interpolante che approssima i dati:



Esempio

$$(1,1),\ (2,4),\ (3,2),\ (4,8)$$



Cerchiamo i minimi quadrati di secondo grado.

$$A_{ij} = i^j \Rightarrow A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$y = a_2 i^2 + a_1 i + a_0$$ dove la nostra variabile $x = i$.

Il vettore $f_j = [1, 4, 2, 8]$ perciò ricaviamo i nostri coefficienti

$$a = (A^T A)^{-1} (A^T \cdot f_j) = \begin{array}{c} \begin{array}{ccc} a_0 & a_1 & a_2 \end{array} \\ [2.5,\ -2.25,\ 1.25] \end{array}$$

TABELLA

Abbiamo detto questo filtro cerca di approssimare i punti dentro la finestra a un polinomio (2° o 4°) Il polinomio viene calcolato con i minimi quadrati. Non abbiamo bisogno di calcolare i coefficienti del polinomio per ogni finestra. Questi coefficienti possono essere calcolati offline. In altre parole quando fissiamo: 1) IL GRADO DEL POLINOMIO 2) n_R 3) n_L abbiamo i coefficienti dentro una tabella:

M	n_L	n_R	Sample Savitzky-Golay Coefficients										
2	2	2				−0.086	0.343	0.486	0.343	−0.086			
2	3	1			−0.143	0.171	0.343	0.371	0.257				
2	4	0		0.086	−0.143	−0.086	0.257	0.886					
2	5	5	−0.084	0.021	0.103	0.161	0.196	0.207	0.196	0.161	0.103	0.021	−0.084
4	4	4		0.035	−0.128	0.070	0.315	0.417	0.315	0.070	−0.128	0.035	
4	5	5	0.042	−0.105	−0.023	0.140	0.280	0.333	0.280	0.140	−0.023	−0.105	0.042