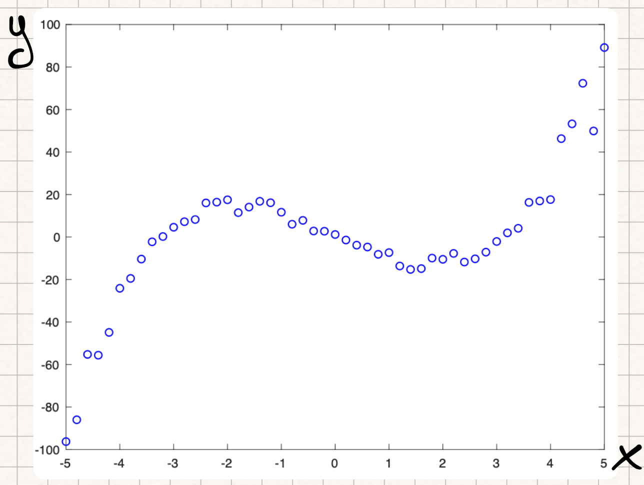
REGRESSIONE

POLINOMIO DI REGRESSIONE

Dati ℓ dati sperimentali $y_1 \dots y_\ell \in \mathbb{R}$ che corrispondono ai dati $x_1 \dots x_n \in \mathbb{R}$ osservati:



Con la regressione vogliamo trovare la migliore approssimazione dei punti osservati, tramite un polinomio p di grado $(n-1)$ con $n \leq \ell$.

Il polinomio p è nella forma:

$$p(x) = z_1 + z_2 x + z_3 x^2 + \dots + z_n x^{n-1}$$

RESIDUAL

Chiamiamo residual la differenza fra il valore del polinomio di approssimazione in x_i e il valore del punto osservato in x_i:

$$r_i = p(x_i) - y_i$$

Obiettivo: trovare $z = (z_1, \dots, z_n)$: $\|r\|$ sia minima :

$$\begin{cases} \min \|Az - y\| \\ z \in \mathbb{R}^n \end{cases}$$

Dove:

$$A = \begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_\ell & \cdots & x_\ell^{n-1} \end{bmatrix} \in \mathbb{R}^{\ell \times n}, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_\ell \end{bmatrix}$$

Funzione obiettivo:

$$f(z) = \|Az - y\| \quad \text{Convessa}$$

POLINOMIO DI REGRESSIONE CON $\|\cdot\|_2$

Data A matrice di Vandermonde, sopra descritta, abbiamo il seguente problema di programmazione quadratica non vincolato:

$$\begin{cases} \min \frac{1}{2} \|Az - y\|_2^2 \\ z \in \mathbb{R}^n \end{cases}$$

Dove, sviluppando la norma euclidea

$$\frac{1}{2}\|Az - y\|_2^2 = \frac{1}{2}(Az-y)^T(Az-y) = \frac{1}{2}\left(z^TA^TAz - z^TA^Ty - Azy + y^Ty\right)$$

$$= \frac{1}{2} z^TA^TAz - z^TA^Ty + \frac{1}{2}y^Ty$$

Possiamo provare che $\mathrm{rank}(A) = n$ e che $A^TA > 0$.

Dato che l'unica soluzione ottima del sistema con $\|\cdot\|_2$ è il punto stazionario della funzione obiettivo, la soluzione del sistema è:

$$f'(z) = 0$$

$$A^TAz - A^Ty = 0$$

$$A^TAz = A^Ty$$

$$z = (A^TA)^{-1}A^Ty$$

POLINOMIO DI REGRESSIONE CON $\|\cdot\|_1$

Ora abbiamo un problema di programmazione lineare:

$$\begin{cases} \min \frac{1}{2}\|Az-y\|_1 \\ z \in \mathbb{R}^n \end{cases}$$

Dove, sviluppando la $\|\cdot\|_1$:

$$\frac{1}{2}\|Az-y\|_1 = \sum_{i=1}^{e} |A_i z - y_i|$$

Il sistema precedente è equivalente del risolvere:

$$\begin{cases} \min \sum_{i=1}^{e} u_i \\ u_i = \|A_i z - y_i\|_1 = \max(A_i z - y_i,\ y_i - A_i z) \end{cases}$$

$$\begin{cases} \min \sum_{i=1}^{\ell} u_i \\ u_i \geq \max(A_i z - y_i,\ y_i - A_i z) \end{cases}$$

$$\boxed{\begin{cases} \min \sum_{i=1}^{\ell} u_i \\ u_i \geq A_i z - y_i & \forall i = 1 \dots \ell \\ u_i \geq y_i - A_i z \end{cases}}$$

In forma vettoriale, quest'ultimo sistema si trasforma

$$\begin{cases} \min\ e_\ell^T u \\ Az - u \leq y \\ -Az - u \leq -y \end{cases}$$

Con $e^T = (1, \dots, 1) \in \mathbb{R}^\ell$.

Settando:

$$D = \begin{pmatrix} A & -I_\ell \\ -A & -I_\ell \end{pmatrix}, \quad d = \begin{pmatrix} y \\ -y \end{pmatrix}$$

Abbiamo il sistema finale:

$$\boxed{\begin{cases} \min\limits_{z,\, u}\ \left(\emptyset_n^T,\ e_n^T\right) \begin{pmatrix} z \\ u \end{pmatrix} \\ D \begin{pmatrix} z \\ u \end{pmatrix} \leq d \end{cases}}$$

Polinomio di regressione con $\|\cdot\|_\infty$

Abbiamo un problema di programmazione lineare:

$$\begin{cases} \min \|Az-y\|_\infty \\ z \in \mathbb{R}^n \end{cases}$$

$$\|Az-y\|_\infty = \max |A_i z - y_i|$$

Che è uguale a

$$\begin{cases} \min\limits_{z,u} u \\ u = \max\limits_{i=1\dots l} |A_i z - y_i| \end{cases} \longrightarrow \begin{cases} \min\limits_{z,u} u \\ u \geq A_i z - y_i \\ u \geq y_i - A_i z \end{cases} \quad \forall i = 1 \dots l$$

Settando

$$D = \begin{pmatrix} A & -e_l \\ -A & -e_l \end{pmatrix}, \quad d = \begin{pmatrix} y \\ -y \end{pmatrix}$$

Abbiamo

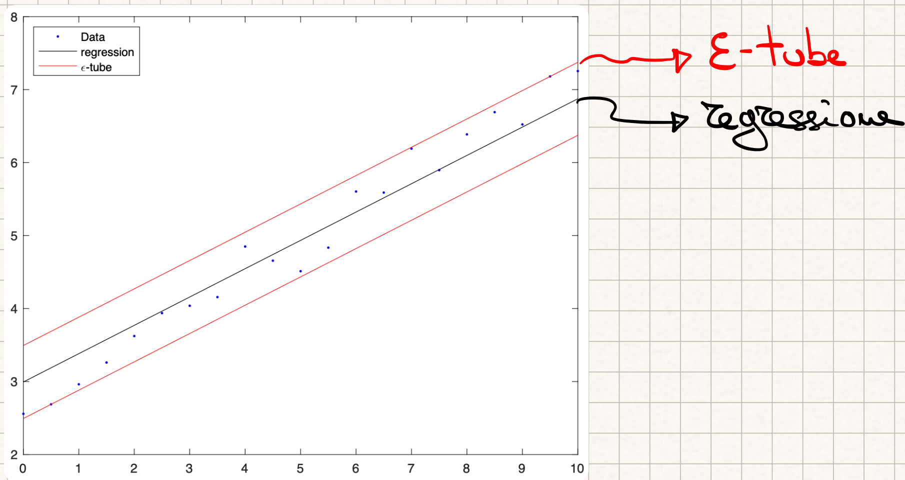
$$\begin{cases} \min\limits_{z,u} (0, \dots, 0, 1) \begin{pmatrix} z \\ u \end{pmatrix} \\ D \begin{pmatrix} z \\ u \end{pmatrix} \leq d \end{cases}$$

Con u scalare

ε-SV REGRESSION

Dato un set di punti di training $\{(x_1, y_1) \dots (x_\ell, y_\ell)\}$ dove $x_i, y_i \in \mathbb{R}^n$, e dato $\varepsilon > 0$, nella regressione di tipo ε-SV, l'obiettivo è trovare la funzione f:

$$|f(x_i) - y_i| \leq \varepsilon \quad \forall i = 1 \dots \ell$$



ε-SV LINEARE

Nel caso lineare consideriamo la funzione **affine** f:

$$f(x) = w^T x + b \quad \text{e} \quad \varepsilon > 0$$

Vogliamo che $f(x)$ sia "piatta", o per meglio dire, che abbia pendenza $w \to 0$. Però vogliamo rispettare anche che i punti siano compresi fra $\pm\varepsilon$ rispetto al nostro iperpiano $(-\varepsilon \leq w^T x_i + b - y_i \leq \varepsilon)$, da cui il sistema:

$$\begin{cases} \min \frac{1}{2}\|w\|^2 \\ y_i \leq w^T x_i + b + \varepsilon \\ y_i \geq w^T x_i + b - \varepsilon \end{cases} \quad \forall i = 1 \dots \ell$$

In forma vettoriale definiamo

$$Q = \begin{pmatrix} I_\ell & \emptyset \\ \emptyset_\ell^T & \emptyset \end{pmatrix}, \quad D = \begin{pmatrix} -x & -e_\ell \\ x & e_\ell \end{pmatrix}, \quad d = \begin{pmatrix} \varepsilon e_\ell - y \\ \varepsilon e_\ell + y \end{pmatrix}$$

Da cui il sistema:

$$\begin{cases} \min\limits_{w,b} \frac{1}{2} (w^T, b) \, Q \begin{pmatrix} w \\ b \end{pmatrix} \\ D \begin{pmatrix} w \\ b \end{pmatrix} \leq d \end{cases}$$

LINEAR ε-SV con variabili di rilassamento

Potrebbe succedere che ε sia troppo piccolo, e per tanto il modello non sia abbastanza flessibile. Estendiamo il modello aggiungendo le variabili "slack" o di rilassamento $\xi_i^{\pm}$:

$$\begin{cases} \min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{\ell}(\xi_i^+ + \xi_i^-) \\ y_i \leq w^T x_i + b + \varepsilon + \xi_i^+ \\ y_i \geq w^T x_i + b - \varepsilon - \xi_i^- \end{cases} \quad \forall i = 1 \dots \ell$$

Dove il parametro C fa da trade-off fra la piattezza di $f(x)$ e la tolleranza su quanto deviare da ε.

$$\begin{cases} \min\limits_{w,b} \ \frac{1}{2}(w^T, b, (\xi^+)^T, (\xi^-)^T) Q1 \begin{pmatrix} w \\ b \\ \xi^+ \\ \xi^- \end{pmatrix} + c^T \begin{pmatrix} w \\ b \\ \xi^+ \\ \xi^- \end{pmatrix} \\ D1 \begin{pmatrix} w \\ b \\ \xi^+ \\ \xi^- \end{pmatrix} \leq d1 \\ \xi^+ \geq 0, \ \xi^- \geq 0 \end{cases}$$

$$Q1 = \begin{pmatrix} I_n & 0_n & O_{n\times 2\ell} \\ 0_n^T & 0 & 0_{2\ell}^T \\ O_{2\ell\times n} & 0_{2\ell} & O_{2\ell\times 2\ell} \end{pmatrix} \qquad c^T = (O_n^T, 0, C * e_\ell^T, C * e_\ell^T)$$

$$D1 = \begin{pmatrix} -x & -e_\ell & -I_\ell & O_{\ell\times\ell} \\ x & e_\ell & O_{\ell\times\ell} & -I_\ell \end{pmatrix} \qquad d = \begin{pmatrix} \varepsilon e_\ell - y \\ \varepsilon e_\ell + y \end{pmatrix}$$

DUALE

Vediamo adesso il duale ε-SV lineare con variabili di rilassamento.

Iniziamo definendo la Lagrangiana

$$L(w, b, \xi^+, \xi^-, \lambda^+, \lambda^-, \eta^+, \eta^-) =$$

$$= \frac{1}{2}\|w\|^2 - w^T\left[\sum_{i=1}^{\ell}(\lambda_i^+ - \lambda_i^-)\, x_i\right] - b\sum_{i=1}^{\ell}(\lambda_i^+ - \lambda_i^-)$$

$$+ \sum_{i=1}^{\ell}\xi_i^+\left(C - \lambda_i^+ - \eta_i^+\right) + \sum_{i=1}^{\ell}\xi_i^-\left(C - \lambda_i^- - \eta_i^-\right)$$

$$-\varepsilon\sum_{i=1}^{\ell}(\lambda_i^+ + \lambda_i^-) + \sum_{i=1}^{\ell}y_i(\lambda_i^+ - \lambda_i^-)$$

SE $\sum_{i=1}^{\ell}(\lambda_i^+ - \lambda_i^-) \neq 0$

oppure $\left(C - \lambda_i^+ - \eta_i^+\right) \neq 0$

oppure $\left(C - \lambda_i^- - \eta_i^-\right) \neq 0$ per qualche i

ALLORA:

$$\min_{w, b, \xi^+, \xi^-} L = -\infty$$

ALTRIMENTI:

$$1)\ \nabla_w L = w - \sum_{i=1}^{\ell}(\lambda_i^+ - \lambda_i^-)\, x_i = 0$$

Da cui ricaviamo, sotto queste condizioni:

$$w = \sum_{i=1}^{\ell}(\lambda_i^+ - \lambda_i^-)\, x_i$$

Allo stesso modo possiamo calcolare:

$$2)\ \nabla_b L = -\sum_{i=1}^{\ell} (\lambda_i^+ - \lambda_i^-) = 0$$

$$3)\ \nabla_{\xi_i^+} L = C - \lambda_i^+ - \eta_i^+ = 0 \implies \eta_i^+ = C - \lambda_i^+$$

$$4)\ \nabla_{\xi_i^-} L = C - \lambda_i^- - \eta_i^- = 0 \implies \eta_i^- = C - \lambda_i^-$$

Nota: Questi ultimi 3 gradienti sono i vincoli del duale.

Sostituendo $w = \sum_{i=1}^{\ell} (\lambda_i^+ - \lambda_i^-) x_i$ alla Lagrangiana abbiamo che il sistema duale è:

$$\begin{cases} \max\limits_{\lambda^+,\lambda^-,\eta^+,\eta^- \geq 0} -\frac{1}{2}\sum\limits_{i=1}^{\ell}\sum\limits_{j=1}^{\ell}(\lambda_i^+ - \lambda_i^-)(\lambda_j^+ - \lambda_j^-)(x_i)^T x_j \\ \qquad -\varepsilon \sum\limits_{i=1}^{\ell}(\lambda_i^+ + \lambda_i^-) + \sum\limits_{i=1}^{\ell} y_i(\lambda_i^+ - \lambda_i^-) \\ \sum\limits_{i=1}^{\ell}(\lambda_i^+ - \lambda_i^-) = 0 \\ C - \lambda_i^+ - \eta_i^+ = 0,\ \ i = 1,..,\ell \\ C - \lambda_i^- - \eta_i^- = 0,\ \ i = 1,..,\ell \end{cases}$$

Notando che $C = \lambda_i^+ + \eta_i^+$ e $C = \lambda_i^- + \eta_i^-$ possiamo riscriver che $\lambda_i^\pm \in [0, C]$:

$$\begin{cases} \max\limits_{\lambda^+,\lambda^-} -\frac{1}{2}\sum\limits_{i=1}^{\ell}\sum\limits_{j=1}^{\ell}(\lambda_i^+ - \lambda_i^-)(\lambda_j^+ - \lambda_j^-)(x_i)^T x_j \\ \qquad -\varepsilon \sum\limits_{i=1}^{\ell}(\lambda_i^+ + \lambda_i^-) + \sum\limits_{i=1}^{\ell} y_i(\lambda_i^+ - \lambda_i^-) \\ \sum\limits_{i=1}^{\ell}(\lambda_i^+ - \lambda_i^-) = 0 \\ \lambda_i^+ \in [0, C],\ \ i = 1,..,\ell \\ \lambda_i^- \in [0, C],\ \ i = 1,..,\ell \end{cases}$$

In forma vettoriale:

$$\begin{cases} \max_{\lambda^+,\lambda^-} \ -\frac{1}{2}((\lambda^+)^T,(\lambda^-)^T)Q\begin{pmatrix}\lambda^+\\ \lambda^-\end{pmatrix} + [-\epsilon(e_\ell^T,e_\ell^T)+(y^T,-y^T)]\begin{pmatrix}\lambda^+\\ \lambda^-\end{pmatrix} \\ (e_\ell^T,-e_\ell^T)\begin{pmatrix}\lambda^+\\ \lambda^-\end{pmatrix}=0 \\ \lambda_i^+\in[0,C],\ i=1,..,\ell \\ \lambda_i^-\in[0,C],\ i=1,..,\ell \end{cases}$$

Con $Q=\begin{pmatrix} X & -X \\ -X & X\end{pmatrix}$ e $X=[(x_i)^T x_j]$ $i,j=1\dots\ell$

Nota:

1) Il duale è un problema convesso

2) SE $\lambda_i^+>0$ o $\lambda_i^->0$ ALLORA x_i = support vector

3) SE (λ^+,λ^-) è ottimo del duale ALLORA

$$w=\sum_{i=1}^{\ell}(\lambda_i^+-\lambda_i^-)\,x_i$$

4) b si ottiene usando

$$\begin{cases} \lambda_i^+\left(\varepsilon+\xi_i^+-y_i+w^Tx_i+b\right)=0 \\ \lambda_i^-\left(\varepsilon+\xi_i^-+y_i-w^Tx_i-b\right)=0 \\ \xi_i^+(C-\lambda_i^+)=0 \\ \xi_i^-(C-\lambda_i^-)=0 \end{cases}$$

Da cui fissando $0<\lambda_i^\pm<C$:

$$b=y_i-w^Tx_i\mp\varepsilon$$

KERNEL

Uguale identico, sostituendo x_i con $\Phi(x_i)$