overdetermined vuol dire che è un sistema in cui ci sono più equazioni che incognite.

overdetermined systems of linear equations

Questa lezione spiega come sovradeterminare i sitemi di equazioni lineari.
La formula qui sotto è la definizione formale, abbiamo un sistema lineare dove le equazioni sono in numero maggiore rispetto alle incognite, A è una matrice rettangolare e x è un vettore colonna. L'output è anch'esso un vettore colonna.
Noi già sappiamo risolverlo, l'idea è trovare i coefficienti del polinomio che mi permettono di creare quella linea grafica che passa più vicino ai punti a disposizione.
In pratica questo problema va sotto il nome di DATA FITTING.
In ogni caso noi abbiamo A e b, dobbiamo trovare x che contiene i coefficienti del polinomio che fitta i dati.
A è la matriche ce contiene tutte le coordinate X e b è il vettore che contiene le Y, l'obiettivo è trovare i coefficienti X

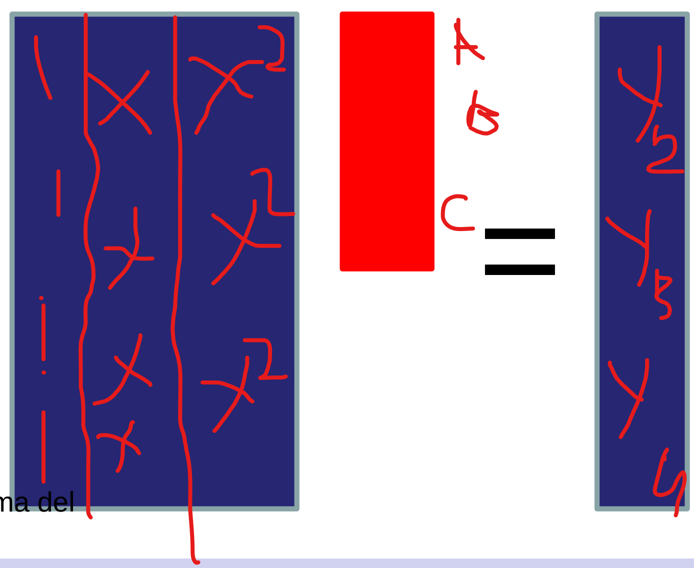
$$A \in \Re^{m \times n},\ m > n$$

$$Ax = b$$

$$x \in \Re^{n}, b \in \Re^{m}$$

(x con i, y con i)

a + bx + cx^2
Questo dovrebbe essere la forma del polinomio.



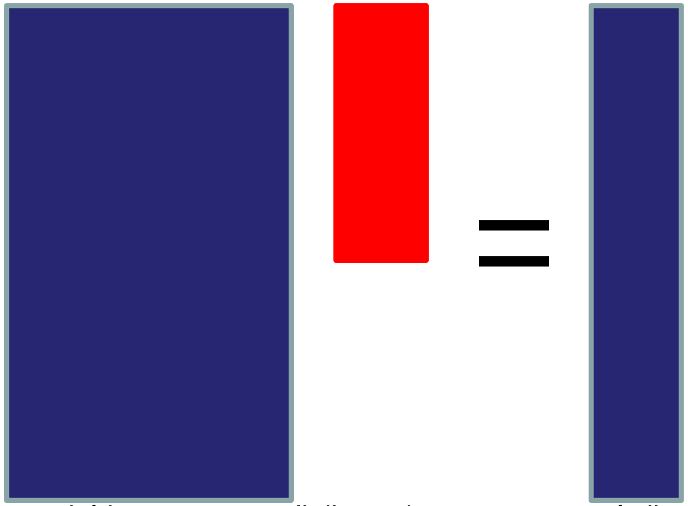
Questo problema in statica prende il nome di regressione. Dove abbiamo che A contiene le features, Y i target e X i coefficienti.

it admits a *true* solution if and only if $b \in R(A)$

Questo problema ammette una soluzione vera se e solo se b appartiene al range di A, altrimenti non sono ammesse reali soluzioni. Per soluzioni reali si intende che Ax = esattamente a B, speigo dopo perché non è possibile.

overdetermined systems of linear equations

$$Ax = b$$



$$A \in \Re^{m\times n},\ m > n$$

$$x \in \Re^{n}, b \in \Re^{m}$$

non è comune avere una true solution perché b e un vettore di dimensione M mentre x è di N ed essendo m > n, allora per esempio ho uno spazio di dimensione 100 (dimensione di B) e supponiamo che ci sarà un piano per esempio di dimensione 3 (3 dimensione di x) allora dobbiamo essere fortunati che il nostro vettore appartenga a questo particolare iperpiano. Per questo in geenrale non abbiamo una soluzione piuttosto ci accontentiamo di determinare una pseudo soluzione,

since a ***true* solution** rarely exists, we are content to determine some kind of **pseudo-solution**

how can we characterize such a (pseudo) **solution** ?

Ma come la caratterizziamo questa pseudo soluzione. (Il punto è che determinata questa pseuodo solutione e calcolando quindi AX= b non otteniamo esattamente b piuttosto un qualcosa di MOLTO SIMILE A B)

Piuttosto siamo interessati a individare il vettore x che minimizzala norma del residuo.

overdetermined systems of linear equations

we look for a vector x that **minimizes** the norm of the **residual**

$$\boxed{r = b - Ax}$$

Cioè cerchiamo il vettore x che risolive il problema di minimizzare, rispetto ad x, la funzione del vettore x chiamata appunto residuo.

i.e, we seek a vector x **solving** :

questa è la definizione di residuo.

$$\min_x \|r\| = \min_x \|b - Ax\| = \min_x \|Ax - b\|$$

such a vector x is called the **solution, in the chosen norm sense, of the overdetermined system**

X è la soluzione, nel senso della norma scelta, del sistema.

overdetermined systems of linear equations

Se è scelta la norma 2 (che è cponsiderata un ottima scelta.), abbiamo il problema dei minimi quadrati.

if the 2-norm is chosen, then we have a **Least Squares problem** (**LS**)

$$x_{LS} = \min_{x} \|Ax - b\|_2$$

Xls è la soluzione del prooblema dei minimi quadrati del sistema.

the vector x_{LS} is the **solution in the Least Squares sense of the overdetermined system**

Cambiare la norma vuol dire cambiare la soluzione del problema.

Vedremo due metodi per risolvere il problema.

overdetermined systems of linear equations

Example

$$Ax = b \qquad A = (1,1,1)^T, b = (b_1, b_2, b_3)^T$$

Questo dice è un esempio (strano non so perché)

i.e.

$$\begin{cases} x = b_1 \\ x = b_2 \\ x = b_3 \end{cases}$$

questo è il sistema da risolvere.

se b1 b2 e b3 sono differenti non c'è una soluzione vera

se usiamo la norma 1 la soluzione è x ?= b2 cioè il mediano del dataset.

$$\|\cdot\|_1 \Rightarrow x = b_2$$

median

se usiamo la norma 2 la soluzione è la media del dataset

$$\|\cdot\|_2 \Rightarrow x = \frac{b_1 + b_2 + b_3}{3}$$

mean

se usiamo la norma infinito otteniamo il punto medio

$$\|\cdot\|_\infty \Rightarrow x = \frac{\min(b_i) + \max(b_i)}{2}$$

mid-point

overdetermined systems of linear equations

Example $Ax = b \qquad A = (1,1,1)^T, b = (b_1, b_2, b_3)^T$

i.e.

$$\begin{cases} x = b_1 \\ x = b_2 \\ x = b_3 \end{cases}$$

Ora analizza la scelta della norma 1, cioè il minimo della somma della differenza tra x e ogni valore del dataset

$$\min_x \left(|x - b_1| + |x - b_2| + |x - b_3| \right)$$

il risultato è il mediamo, e questo lo dimostra un teorema della statica.

$$\|\cdot\|_1 \Rightarrow x = b_2$$

median

il mediano minimizza la somma dei valori assoluti delle deviazione, cioè le differenze.

the median of a dataset minimizes the sum of the absolute values of the deviations

overdetermined systems of linear equations

Example $Ax = b \qquad A = (1,1,1)^T, b = (b_1, b_2, b_3)^T$

i.e.

$$\begin{cases} x = b_1 \\ x = b_2 \\ x = b_3 \end{cases}$$

La norma due invece è la somma dei quadrati delle deviazioneni, prendiamo la x che minimizza questo. Lui dice di essersi dimenticato la radice quadrata, l'ho aggiunta io.

$$\min_x \sqrt{\left(|x - b_1|^2 + |x - b_2|^2 + |x - b_3|^2 \right)}$$

Il risultato è la media e questo lo dimostra un teorema statistico.

$$\|\cdot\|_2 \Rightarrow x = \frac{b_1 + b_2 + b_3}{3}$$

mean

La media aritmetica minimizza la somma dei quadrati delle deviazioni.

the arithmetic mean of a dataset minimizes the sum of the squared deviations

overdetermined systems of linear equations

Example $Ax = b \qquad A = (1,1,1)^T, b = (b_1, b_2, b_3)^T$

i.e.

$$\begin{cases} x = b_1 \\ x = b_2 \\ x = b_3 \end{cases}$$

QUesta è la norma infinito è ad esempio avendo 3 punti , il punto medio che ci permette di minimizzare la distanza tra i tre punti.

$$\min_x \left(\max_x \left(|x - b_1|, |x - b_2|, |x - b_3| \right) \right)$$

$$\|\cdot\|_\infty \Rightarrow x = \frac{\min(b_i) + \max(b_i)}{2}$$

mid-point of the dataset

Vediamo alcune proprietà del problema dei minimi quadrati, non sappiamo quante soluzioni ha il sistema ne se le ha, questo dipende dalla metrica A.
Noi ~~chiamimao X come l'insieme che contiene le x che risolvono questo problema (minimi quadrati), si tratta di un insieme con~~vesso.

overdetermined systems of linear equations

Dato un insieme considerando due punti, se traccio la linea che collega due punti e la linea è nell'0insieme allora l'insieme è convesso altrimenti no

properties of the set X of **LS** solutions

$$\min_x \|Ax-b\|_2$$

$$X = \{ x \in \Re^n : \underbrace{\|Ax-b\|_2 = \text{minimum}} \}$$

Questa di sotto mostra come risolvere il problema in questione. Allora innazitutto Ax-b = minimum è la normal equation (equazione normale). Per risolvere prendiamo la trasposta di A la moltiplichiamo ambo i membri. Porta b a sx e lo mette in evidenza. In matlab basta scrivere x = A\b e risolve il problema, cioè l'equazione della norma.

X is a convex set

Ax = b è l'equazione norma, semplicemente ha portato b a DX

$$x \in X \Rightarrow A^T A x = A^T b \quad , \quad A^T (Ax-b) = 0$$

Queste sono proprietà delle soluzioni. Se esiste almeno soluzione al problema (le solu sono nell'insieme X convesso), allora possiamo prendere la più piccola, questa è la minimum norm solution.

$$\exists! x_{LS} \in X : \|x_{LS}\|_2 = \min \{ \|x\|_2, x \in X \}$$

minimum norm solution

Se succede che il massimo rank di A è n l'insieme convesso delle soluzioni è constituito da un solo vettore, quindi c'è una sola soluzione
Questa proprietà è la più importante.

$$X = \{x_{LS}\} \Leftrightarrow rank(A) = n$$

ro è la norma del residuo, cioè corrisponde alla soluzione del Least Square Problem, perché è il minimum sum of suqares.

$$\rho_{LS}^2 = \|Ax_{LS} - b\|_2^2 = \|r\|_2^2$$

minimum sum of squares

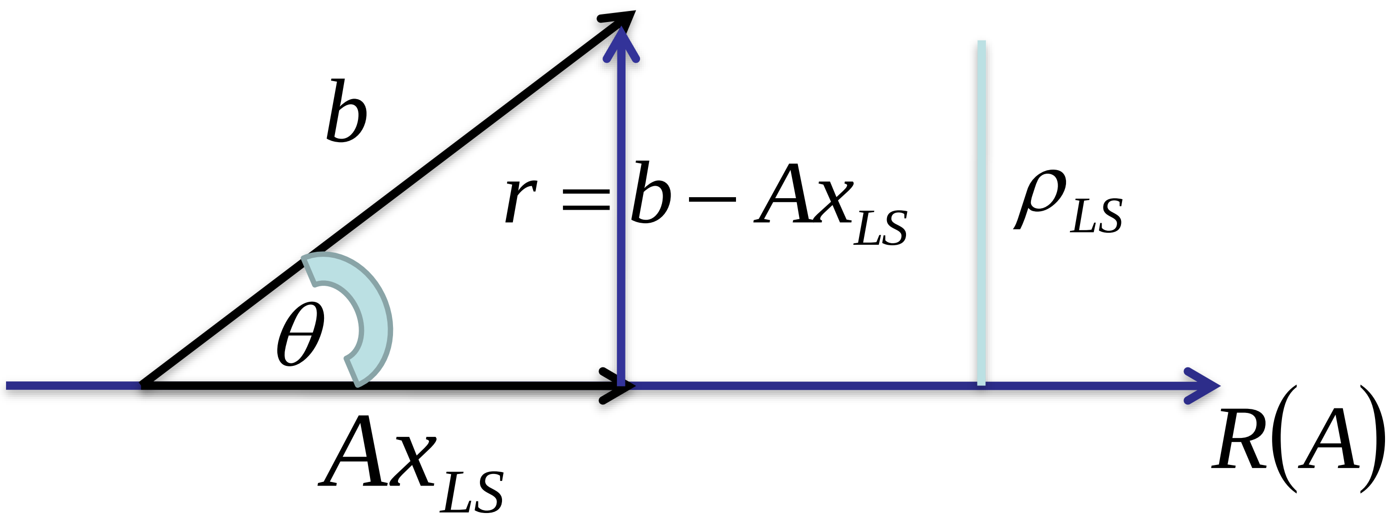
QUESTA SLIDE NON L'HO CAPITA MINIMIAMTNE. lezione 4.1 MINUTO 24

overdetermined systems of linear equations

PARLA MALISSIMO

Ax_{LS} is the orthogonal projection of b onto the range of A

Al massimo dirà questa definizione.



Questo è l'angolo tra range di A e b.

$$\sin(\theta) = \frac{\|r\|_2}{\|b\|_2} = \frac{\rho_{LS}}{\|b\|_2}$$

Qui dice che la soluzione del LS può essere denotata anche

overdetermined systems of linear equations

an **LS** solution

$$\min_x \|Ax-b\|_2$$

can be denoted as

A plus è la pseudo inversa della matrice A, quindi mette in relazione A plus con il vettore b. Ovviamente a partire poi dalla definizione classica di ro con la A plus diventa:

$$x_{LS} = A^+ b$$

$$\rho_{LS} = \left\| \left(AA^+ - I \right) b \right\|_2$$

ricordiamo che ro è la norma del residuo, quindi la soluzione al problema.

A^+ is the **pseudoinverse** matrix

a formal definition of pseudoinverse is the following:

$$A^+ = \min_{V \in \Re^{n \times m}} \|AV - I\|_F$$

Questa è la definizione fermale ed è la generalizzazione della matrice di identità nel caso in cui A sia non quadrata.

overdetermined systems of linear equations

an **LS** solution $\min_x \|Ax-b\|_2$

can be denoted as

$$\boxed{x_{LS} = A^{+}b}$$

$$\rho_{LS} = \left\|\left(AA^{+} - I\right)b\right\|_2$$

se il rank di A è n allora la pseudo inversa
è quella roba li

if $rank(A) = n$ then the pseudoinverse becomes

P.s ci dice che spesso
vedremo questo trucco
cioè per isolare x
moltiplichiamo ambro i
metri per l'inversa.

$$A^{+} = \left(A^{T}A\right)^{-1}A^{T}$$

$$A^{T}Ax = A^{T}b$$

W questa è il normal equations
system

Ovviamente se A è quadrata
la pseudo è = all'inversa.

and if A is square $\Rightarrow$ $A^{+} = A^{-1}$

overdetermined systems of linear equations

Questo è il punto centrale della lezione: la soluzione le LS può essere scritto come il punto 1. Se a quella formula moltiplico ad ambo i membri A otteniamo il 2 cioè la proiezione ortogonale di b nel range di A. aa PLUS è DETTO PROIETTORE ORTOGONALE su range di A.

an **LS** solution

$$\min_x \|Ax - b\|_2$$

can be denoted as

1

$$\boxed{x_{LS} = A^+ b}$$

$$\rho_{LS} = \|(AA^+ - I)b\|_2$$

2

$$\boxed{Ax_{LS} = AA^+ b}$$

is the orthogonal projection of b onto the range of A

$$\boxed{AA^+}$$

is the **orthogonal projector** onto the range of A

overdetermined systems of linear equations

Se il rank di A è n cioè è massimo, A è m x n, A trasposto * A è quadrata, simmetrica e positiva (positiva definita vuol dire che la matrice ha l'eigenvalues reale con valori tutti positiiv). In poche parole se sono in questo caso devo moltiplicare ambo i membri per la trasposta, in questo modo ho il sistema quadrato e lo posso risolvere.

$$\min_x \|Ax - b\|_2$$

case $rank(A) = n$ (**LS full rank**)

$$A^T A x = A^T b$$

normal equations system

$A^T A$ is square, symmetric and positive definite

Simmetrica che la trasposta è = all'originale, cioè At = A.

Cholesky factorization

Possiamo risolvere il sistema di equazioni normale con questa fattorizzazione, la cui complessità è scirtta T(m,n)

$$T(m,n) = (n^2/2)(m + n/3)$$

time complexity

overdetermined systems of linear equations

$$\min_{x} \|Ax - b\|_2$$

case $rank(A)=n$ (**LS full rank**)

$$A^T A x = A^T b$$

normal equations system

$A^T A$ is square, symmetric and positive definite

In questa slide che in questo caso analizzato, (lo stesso della slide precednete) c'è un indice di condizionamento molto grande. Perché l'indice di condizionamento della matrice (At * Ax) è = al quadrato dell'indice di condizionamento di A. Se A è mal condizionato il problema radodpia questo mal condizionamtno.

accuracy (relative error) proportional to

$$\kappa_2(A)^2$$

l'indicie di cond di una matrice rett. è la norma 2 * la norma 2 della matrice pseudo inversa

$$\kappa_2(A) = \|A\|_2 \|A^+\|_2$$