UNIVERSITÀ DEGLI STUDI DI SALERNO

Dipartimento di Ingegneria dell'Informazione ed Elettrica e Matematica applicata

Corso di Laurea Magistrale in Ingegneria Informatica

APPUNTI DI DATA SCIENCE

DI FRANCESCO PIO CIRILLO

https://github.com/francescopiocirillo

"Sii sempre forte"

😄 Ehi, un attimo prima di iniziare!

Hai appena aperto una raccolta di appunti che ho deciso di condividere **gratuitamente** su GitHub, se ti sono utili fai **una buona azione digitale**:

- ⭐ **Lascia una stellina alla repo**: è gratis, indolore e fa super piacere!
- 📤 **Condividila con amici**, compagni di corso, o chiunque possa averne bisogno.

Insomma, se questi appunti ti salvano anche solo una giornata di studio... fammelo sapere con una **stellina**!

Grazie di cuore ❤️

Dimostrazione formula del calcolo di Beta cappello

Prima parte, gradiente

Nel contesto:

MULTIPLE LINEAR REGRESSION

$X^T = (X_1, \dots, X_p)$ potenzialmente può influenzare la dinamica di y

$$Y = f(X) + \varepsilon = \beta_0 + \sum_{j=1}^{p} \beta_1 X_j + \varepsilon$$ ipotizziamo linearità

Usando Least Squares, quindi puntando alla minimizzazione di:

$$RSS(\beta) = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

Valgono le formule:

$i = 1, \dots, n$

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i$$

molt. tra matrici righe per colonne

$$\underset{n \times 1}{\underline{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}} = \underset{n \times (p+1)}{\begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & & & \\ 1 & x_{n1} & x_{n2} & & x_{np} \end{bmatrix}} \underset{(p+1) \times 1}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}} + \underset{n \times 1}{\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}}$$

moltiplicando β_0 non cambia

$n \times (p+1)$ DESIGN MATRIX

$$\underline{y} = \underline{\underline{X}}\,\underline{\beta} + \underline{\varepsilon}$$

est $\hat{\underline{\beta}} \Rightarrow \hat{\underline{y}} = \underline{\underline{X}}\,\hat{\underline{\beta}} \Rightarrow$ dobbiamo trovare $\hat{\underline{\beta}}$

Essenzialmente abbiamo preso l'espressione e l'abbiamo scritta come matrici e vettori.

Nella matrice delle x, o matrice di design, le x hanno due pedici, il primo distingue i diversi campionamenti, il secondo indica il numero del regressore.

Dobbiamo trovare $\hat{\beta}$ che minimizza l'RSS:

$$RSS(\underline{\beta}) = \sum_{i=1}^{n} \big(y_i - \underbrace{(\underline{\underline{X}}\,\underline{\beta})_i}_{a_i}\big)^2 = \underbrace{(\underline{y} - \underline{\underline{X}}\,\underline{\beta})^T (\underline{y} - \underline{\underline{X}}\,\underline{\beta})}_{\text{calcoli i quadrati}}$$

Se ho un vettore e voglio calcolarne i quadrati facciamo riga per colonna, la sommatoria la riscriviamo.

Ricordando che:

RICORDA: gradiente

① $\nabla_x\, a^T x = \nabla_x\, x^T a = a$

② $\nabla_x\, \underbrace{x^T A x}_{\sum_i \sum_j x_i a_{ij} x_j} = A^T x + A x$ — A mat quadrata

info da sapere per la dim che segue

$\underline{\beta}$ $\underline{\underline{X}}$ $\underline{y}$ sottintesi gli _

Dimostriamo la formula:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}.$$

Per trovare il minimo RSS iniziamo ponendo il gradiente a 0:

$$\nabla_{\underline{\beta}}\, RSS(\underline{\beta}) = \underline{0} = \nabla_{\underline{\beta}} \left[(\underline{y} - \underline{\underline{X}}\,\underline{\beta})^T (\underline{y} - \underline{\underline{X}}\,\underline{\beta}) \right]$$

Eseguiamo il prodotto:

$$= \nabla_{\underline{\beta}} \left[\underline{y}^T \underline{y} - \underbrace{\underline{y}^T \underline{\underline{X}}\,\underline{\beta}}_{\mathrm{I}} - \underbrace{\underline{\beta}^T \underline{\underline{X}}^T \underline{y}}_{\mathrm{II}} + \underline{\beta}^T \underline{\underline{X}}^T \underline{\underline{X}}\,\underline{\beta} \right]$$

II trasposto del I

dim: $1 \times n \quad n \times (p+1) \quad (p+1) \times 1$

$1 \times (p+1)$

1×1

trasposto dello scalare è lo scalare stesso

quindi II = I

Sommiamo i due termini rosa:

$$= \nabla_{\beta}\left[y^T y - 2\beta^T X^T y + \beta^T X^T X \beta \right]$$

Svolgiamo il gradiente:

deriviamo

$$= 0 - 2\overset{1}{X^T} y + \underbrace{(X^T X)^T}_{X^T (X^T)^T = X^T X} \beta + \overset{2}{X^T X} \beta$$

uguali

$$= -2X^T y + 2X^T X \beta$$

Ricordando che quest'ultima espressione va posta a 0 ricaviamo il valore $\hat{\beta}$ che minimizza l'RSS:

$$\Rightarrow \underbrace{\boxed{\hat{\beta}_{LS} = (X^T X)^{-1} X^T y}}_{\text{least squar}} \rightarrow$$

LSE
stimatore che useremo

Invertibilità

Nella formula che abbiamo ottenuto c'è l'inversione di una matrice ma non tutte le matrici sono invertibili (se il rango è 0 non si può).

Risulta che la matrice è invertibile se la matrice di design è a rango pieno (full rank).

💡 La matrice di design è full rank se $n > p$, cioè il numero di punti del dataset deve essere maggiore del numero di regressori.

Questa affermazione può essere interpretata come il fatto che ci serve avere abbastanza punti per permettere la stima dei parametri che vogliamo stimare.

Ortogonalità di errore e span generato dalle x

Ripartiamo dall'espressione:

$$-2\underline{\underline{X}}^T\underline{y} + 2\underline{\underline{X}}^T\underline{\underline{X}}\,\underline{\beta}$$

Mettiamo in evidenza -2 e specializziamo l'equazione per $\hat{\beta}$.

$$-2X^T\underbrace{(y - X\hat{\beta})}_{y - \hat{y} = e} = 0$$

💡 Il fatto che il risultato dell'equazione sia 0 ci dice che l'errore è ortogonale allo SPAN generato dai vettori x (visto che moltiplicato per la trasposta di x fa 0).

Il sottospazio generato dai vettori x si chiama SPAN. L'errore è ortogonale a quello spazio, dove? nel punto che minimizza l'RSS. Questa è una proprietà interessante.

Seconda parte, matrice Hessiana

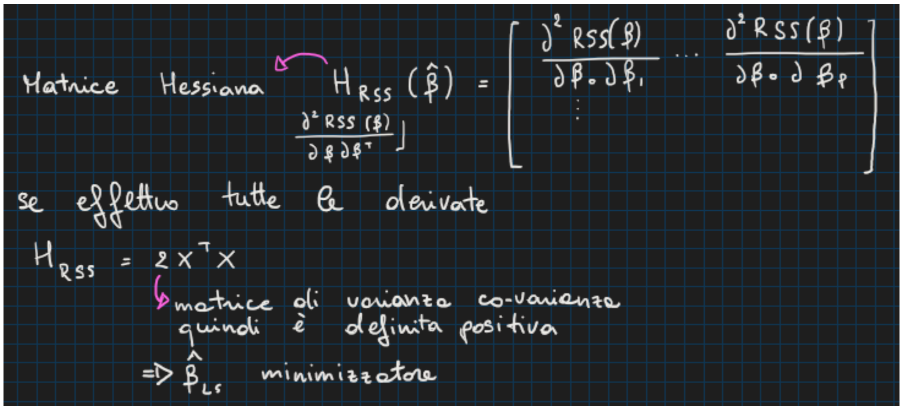
Ripartiamo dall'espressione:

$$-2\underline{\underline{X}}^T\underline{y} + 2\underline{\underline{X}}^T\underline{\underline{X}}\,\underline{\beta}$$

Abbiamo detto che a noi serve lo stimatore di β che minimizza l'RSS, per ottenere questo non basta porre il gradiente pari a 0 ma bisogna anche fare considerazioni sulla Matrice Hessiana.

Valutiamo la Matrice Hessiana nel punto $\hat{\beta}$.

Nello specifico la Matrice Hessiana deve essere definita positiva, quindi visto che è simmetrica basta che gli autovalori siano tutti positivi.

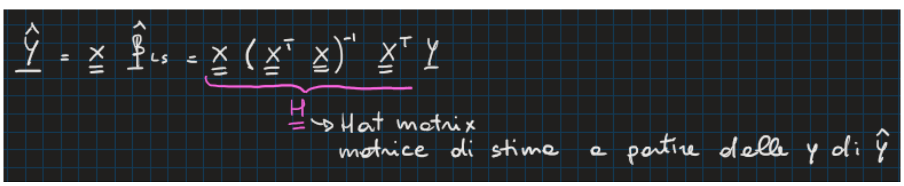


La matrice $H_{RSS} = 2X^TX$ è una somma di quadrati, è quindi proporzionale alla matrice di varianza-covarianza dei dati x. Si può quindi dimostrare che questa matrice è una stima della matrice di var-covar e sapendo che quest'ultima ha per definizione tutti i valori positivi questo varrà anche per la matrice che stiamo cercando.

Possiamo quindi confermare che il nostro $\hat{\beta}_{LS}$ è un punto di minimo.

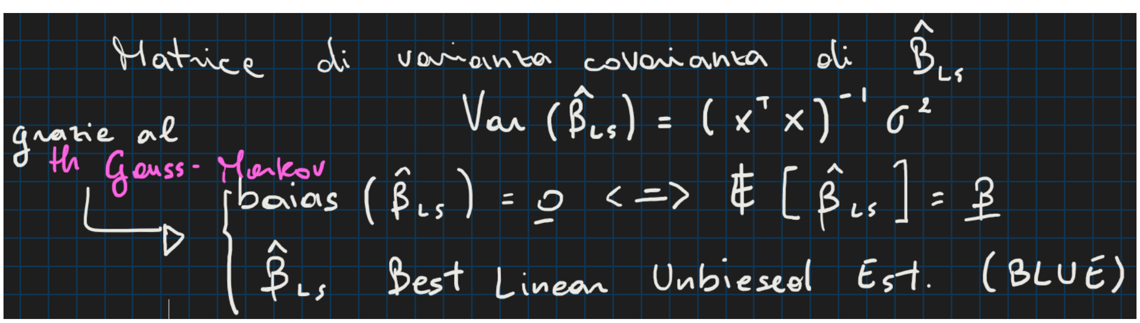
Ultime considerazioni

Avendo confermato la validità di $\hat{\beta}_{LS}$ possiamo sostituire l'espressione dello stimatore nell'espressione.



La matrice H, Hat Matrix, è una matrice che prende i dati y e li trasforma in $\hat{y}$, il modo in cui effettua questa azione è proiettando la y nello SPAN generato dalle x.

Constatiamo che la varianza del nostro stimatore dipende da σ^2, questo ha senso perché la varianza dell'errore implica la dispersione più o meno grande delle mie stime.



In virtù del teorema di Gauss-Markov ha media che coincide con il valore vero, quindi bias nullo. Questo è il miglior stimatore unbiased tra gli stimatori lineari, BLUE.

La stima di σ^2 è un problema affrontato nel continuo di 2 - Linear Regression B.

DISCLAIMER

Questi appunti sono stati realizzati a scopo puramente educativo e di condivisione della conoscenza. Non hanno alcun fine commerciale e non intendono violare alcun diritto d'autore o di proprietà intellettuale.

I contenuti di questo documento sono una rielaborazione personale di lezioni universitarie, materiali di studio e concetti appresi, espressi in modo originale ove possibile. Tuttavia, potrebbero includere riferimenti a fonti esterne, concetti accademici o traduzioni di materiale didattico fornito dai docenti o presente in libri di testo.

Se ritieni che questo documento contenga materiale di tua proprietà intellettuale e desideri richiederne la modifica o la rimozione, ti invito a contattarmi. Sarò disponibile a risolvere la questione nel minor tempo possibile.

In quanto autore di questi appunti non posso garantire l'accuratezza, la completezza o l'aggiornamento dei contenuti e non mi assumo alcuna responsabilità per eventuali errori, omissioni o danni derivanti dall’uso di queste informazioni. L’uso di questo materiale è a totale discrezione e responsabilità dell’utente.