

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea in Matematica

**PROBLEMA AI MINIMI QUADRATI
APPLICATO ALLA
REGRESSIONE LINEARE:
STUDIO DI CASI APPLICATIVI**

Tesi di Laurea in Calcolo Numerico

Relatrice:
Chiar.ma Prof.ssa
VALERIA SIMONCINI

Presentata da:
SILVIA CASERIO

II Sessione
Anno Accademico 2015/2016

... ai miei nonni.

Indice

Introduzione	iii
1 Il modello di regressione lineare	1
1.1 Il modello di regressione lineare classico	1
1.2 La stima ai minimi quadrati	2
1.2.1 La decomposizione della somma di quadrati	5
1.2.2 Proprietà di campionamento degli stimatori ai minimi quadrati classici	6
1.3 Stima della funzione di regressione e predizione di una nuova osservazione in \mathbf{z}_0	8
2 Il problema ai minimi quadrati e metodi di risoluzione	11
2.1 L'equazione normale	11
2.2 La fattorizzazione QR	12
2.3 La decomposizione in valori singolari	13
2.3.1 La decomposizione in valori singolari ed il problema ai minimi quadrati	15
2.4 Considerazioni sul condizionamento dei metodi di risoluzione trattati	18
3 Studio di due casi applicativi	21
3.1 LONGLEY dataset	22
3.2 DETROIT dataset	29

Conclusioni	37
A Appendice	39
B Appendice	41
B.1 Listati dei programmi	41
B.1.1 Funzione MATLAB REGRESSLIN	41
B.1.2 Funzione MATLAB regresslin095	42
B.1.3 LONGLEY script	43
B.1.4 DETROIT script	45
Bibliografia	47

Introduzione

La regressione lineare è un metodo statistico utilizzato per predire i valori di una o più variabili dipendenti, dette *risposte*, da una collezione di valori di variabili indipendenti, dette *predittori* e può essere usato anche per valutare gli effetti delle variabili predittrici sulle risposte ([1]). Trova applicazione in svariati ambiti, quali, ad esempio, l'ingegneria, la biologia, l'economia, ed è così largamente diffuso in quanto si traduce in un normale problema ai minimi quadrati.

Nel primo capitolo, verrà presentato il modello di regressione lineare multivariata e verranno esposti i suoi aspetti teorici, evidenziandone le proprietà qualitative e la sua riconducibilità ad un problema di minimi quadrati.

Nel secondo capitolo, verrà presentato il problema ai minimi quadrati ed alcuni suoi risultati generali, seguiti dalla descrizione dei metodi utilizzati per la sua risoluzione dal punto di vista numerico.

Nel terzo capitolo, infine, verranno analizzati sperimentalmente due set di dati noti in letteratura, ricorrendo ai metodi numerici descritti nel capitolo precedente, sottolineando quale sia quello più efficiente.

Capitolo 1

Il modello di regressione lineare

1.1 Il modello di regressione lineare classico

Siano z_1, z_2, \dots, z_m con $m \in \mathbb{N}$ variabili predittrici associate alla risposta Y . Il modello di regressione lineare classico afferma che Y è data da una combinazione, che dipende linearmente dagli z_j , e da un errore random, r , che rappresenta l'errore di misura e gli effetti di altre variabili non esplicitamente considerate nel modello. I valori dei predittori, in quanto dati sperimentali acquisiti da un'osservazione, sono considerati fissati. L'errore, e quindi la risposta, è visto come una variabile aleatoria il cui comportamento è determinato da un insieme di ipotesi distributive ([1]). In particolare, il modello di regressione lineare con risposta singola è dato da

$$Y = \beta_0 + \beta_1 z_1 + \dots + \beta_m z_m + r \quad (1.1)$$

[Risposta]=[media (dipendente da z_1, z_2, \dots, z_m)]+[errore]

dove il termine 'lineare' si deve alla media, in quanto funzione lineare delle incognite $\beta_0, \beta_1, \dots, \beta_m$. I predittori possono o meno influenzare il modello come termini di primo ordine ([1]). Analogamente, date n osservazioni indipendenti di Y e valori associati di z_k , denotati

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{Z} = \begin{bmatrix} 1 & z_{1,1} & z_{1,2} & \cdots & z_{1,m} \\ 1 & z_{2,1} & z_{2,2} & \cdots & z_{2,m} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & z_{n,1} & z_{n,2} & \cdots & z_{n,m} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}, \mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}$$

tali che gli errori r_k soddisfino le seguenti proprietà

1. $E(r_k) = 0$
2. $Var(r_k) = \sigma^2$ (costante)
3. $Cov(r_j, r_k) = 0$ se $j \neq k$,

allora il modello diventa in notazione matriciale

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{r} \quad (1.2)$$

con $E(\mathbf{r}) = \mathbf{0}$, $Cov(\mathbf{r}) = E(\mathbf{r}\mathbf{r}^T) = \sigma^2\mathbf{I}_n$ e dove $\boldsymbol{\beta}$ e σ^2 sono parametri sconosciuti. Ogni colonna di \mathbf{Z} consta di n valori della corrispondente variabile predittrice, mentre la k -esima riga di \mathbf{Z} è formata dai valori di tutti i predittori del k -esimo esperimento ([1]).

1.2 La stima ai minimi quadrati

Uno degli obiettivi della regressione lineare è quello di determinare un'equazione che permetta di predire la risposta in base a particolari valori assunti dai predittori ([1]). A tal fine è necessario adattare il modello in (1.2) alle quantità osservate y_k in corrispondenza di quelle note $1, z_{k,1}, \dots, z_{k,m}$ per $k \in \{1, \dots, n\}$, ed individuare i valori dei coefficienti di regressione $\boldsymbol{\beta}$ e della varianza σ^2 relativa ad \mathbf{r} in accordo con i dati disponibili ([1]).

Siano

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n \text{ e } \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^{m+1} \text{ i vettori, rispettivamente, delle } n \text{ risposte}$$

osservate e degli $m+1$ valori sperimentali di β . Sia $\mathbf{Z} \in \mathbb{R}^{n,m+1}$ come in (1.2), con $n \geq m+1$. Il *metodo dei minimi quadrati* consiste nel determinare il vettore $\mathbf{b} \in \mathbb{R}^{m+1}$ che risolve il seguente problema di ottimizzazione:

$$\min_{\mathbf{b} \in \mathbb{R}^{m+1}} S(\mathbf{b}) := \min_{\mathbf{b} \in \mathbb{R}^{m+1}} \|\mathbf{y} - \mathbf{Z}\mathbf{b}\|_2^2. \quad (1.3)$$

I coefficienti \mathbf{b} , soluzione di (1.3), sono chiamati *stime ai minimi quadrati* dei parametri di regressione β .

Osservazione 1. Si denotino gli elementi di

$$\operatorname{argmin}_{\mathbf{b} \in \mathbb{R}^{m+1}} S(\mathbf{b})$$

con $\hat{\beta} \in \mathbb{R}^{m+1}$ e $\mathbf{y} - \mathbf{Z}\hat{\beta}$ con $\hat{\mathbf{r}} \in \mathbb{R}^n$. Quest'ultimo sarà il vettore dei *residui* e contiene informazioni sul parametro ignoto σ^2 che compare tra le proprietà di \mathbf{r} in (1.2).

Il lemma ed il teorema seguenti costituiscono un risultato generale che permette di stabilire condizioni di esistenza ed unicità della soluzione di (1.2).

Lemma 1.2.1. ([2]) *Una matrice $\mathbf{A} \in \mathbb{R}^{n,m}$ con $n \geq m$ ha rango massimo se e solo se la matrice $\mathbf{A}^T \mathbf{A}$ è non singolare.*

Dimostrazione. Sia $\mathbf{A} = \mathbf{Q}_1 \mathbf{R}_1$ la fattorizzazione QR di \mathbf{A} , con \mathbf{R}_1 matrice $m \times m$ e \mathbf{Q}_1 avente colonne ortonormali. Dalla procedura di Gram-Schmidt, per esempio, segue che le colonne di \mathbf{A} sono linearmente indipendenti se e solo se \mathbf{R}_1 è non singolare. Si ha, inoltre, che $\mathbf{A}^T \mathbf{A} = \mathbf{R}_1^T \mathbf{Q}_1^T \mathbf{Q}_1 \mathbf{R}_1 = \mathbf{R}_1^T \mathbf{R}_1 \geq 0$. Segue che $\mathbf{A}^T \mathbf{A}$ è non singolare, ed in particolare $\mathbf{A}^T \mathbf{A} > 0$, se e solo se $\mathbf{R}_1^T \mathbf{R}_1$ è non singolare, se e solo se $\mathbf{R}_1^T \mathbf{R}_1 > 0$, se e solo se \mathbf{R}_1 è non singolare. \square

Teorema 1.2.2. ([2]) *Sia \mathcal{X} l'insieme dei vettori $x \in \mathbb{R}^m$ che risolvono*

$$\min_{x \in \mathbb{R}^m} \|c - \mathbf{A}x\|_2.$$

Allora valgono le seguenti affermazioni:

1. $x \in \mathcal{X}$ se e solo se x è soluzione dell'equazione normale

$$\mathbf{A}^T \mathbf{A}x = \mathbf{A}^T c \quad (1.4)$$

2. \mathcal{X} ha un solo elemento se e solo se \mathbf{A} ha rango massimo.

Dimostrazione. Siano

$$R = \text{range}(\mathbf{A}) = \{y \in \mathbb{R}^n \text{ tale che } y = \mathbf{A}x, x \in \mathbb{R}^m\}$$

e il suo ortogonale

$$R^\perp = \{z \in \mathbb{R}^n \text{ tale che } z^T y = 0, y \in R\}.$$

Si può quindi scrivere $c = c_1 + c_2$, con $c_1 \in R$ e $c_2 \in R^\perp$.

1. Si ha

$$r = c - \mathbf{A}x = c_1 - \mathbf{A}x + c_2$$

con $c_1 - \mathbf{A}x \in R$ e $c_2 \in R^\perp$. Poichè $c_2^T(c_1 - \mathbf{A}x) = 0$, si ha

$$\|r\|_2^2 = \|c_1 - \mathbf{A}x\|_2^2 + \|c_2\|_2^2.$$

Per ottenere il minimo di $\|r\|$, si può operare solo sul primo termine, in quanto nel secondo non compare x . Quindi la minima $\|r\|_2$ è raggiunta se e solo se x risolve $c_1 = \mathbf{A}x$. Seguirà che $r = c_2$. Allora $r \in R^\perp$. In particolare, il residuo r risulta ortogonale alle colonne di \mathbf{A} , cioè

$$0 = \mathbf{A}^T r = \mathbf{A}^T(c - \mathbf{A}x) \Leftrightarrow \mathbf{A}^T \mathbf{A}x = \mathbf{A}^T c.$$

2. Per il Lemma 1.2.1, la matrice \mathbf{A} ha rango massimo se e solo se $\mathbf{A}^T \mathbf{A}$ è invertibile se e solo se l'equazione normale (1.4) ammette un'unica soluzione.

□

Corollario 1.2.3. ([1]) Sia \mathbf{Z} in (1.2) di rango massimo $m + 1 \leq n$. La stima ai minimi quadrati di β è data da

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

Si denoti con $\hat{\mathbf{y}} = \mathbf{Z} \hat{\beta}$ il valore arrotondato di \mathbf{y} . I residui

$$\hat{\mathbf{r}} = \mathbf{y} - \hat{\mathbf{y}} = [\mathbf{I} - \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T] \mathbf{y}$$

soddisfano $\mathbf{Z}^T \hat{\mathbf{r}} = 0$ e $\hat{\mathbf{y}}^T \hat{\mathbf{r}} = 0$. Inoltre, si ha che

$$\begin{aligned} \hat{\mathbf{r}}^T \hat{\mathbf{r}} &= \|\hat{\mathbf{r}}\|_2^2 = \left\| \mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\beta}} \right\|_2^2 = \mathbf{y}^T [\mathbf{I} - \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T] \mathbf{y} = \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{Z}\hat{\boldsymbol{\beta}} \\ &= \|\mathbf{y}\|_2^2 - \mathbf{y}^T \mathbf{Z}\hat{\boldsymbol{\beta}}. \end{aligned} \quad (1.5)$$

Dimostrazione. Segue dal Teorema 1.2.2. \square

1.2.1 La decomposizione della somma di quadrati

Per il Corollario 1.2.3, dalla condizione $\hat{\mathbf{y}}^T \hat{\mathbf{r}} = 0$, segue che

$$\sum_{j=1}^n y_j^2 = \|\mathbf{y}\|_2^2 = \|\pm \hat{\mathbf{y}} + \mathbf{y}\|_2^2 = \|\hat{\mathbf{y}} + \hat{\mathbf{r}}\|_2^2 = \|\hat{\mathbf{y}}\|_2^2 + \|\hat{\mathbf{r}}\|_2^2. \quad (1.6)$$

Inoltre, per costruzione di \mathbf{Z} , dal risultato $\mathbf{Z}^T \hat{\mathbf{r}} = 0$ del Corollario 1.2.3 si ha:

$$0 = \mathbf{1}^T \hat{\mathbf{r}} = \sum_{j=1}^n \hat{r}_j = \sum_{j=1}^n (y_j - \hat{y}_j) \Leftrightarrow \bar{\mathbf{y}} = \bar{\hat{\mathbf{y}}}.$$

Sottraendo $n\bar{\mathbf{y}}^2 = n\bar{\hat{\mathbf{y}}}^2$ ad entrambi i membri di (1.6), si ottiene la decomposizione elementare della somma di quadrati attorno alla media ([1]):

$$\|\mathbf{y}\|_2^2 - n\bar{\mathbf{y}}^2 = \|\hat{\mathbf{y}}\|_2^2 + \|\hat{\mathbf{r}}\|_2^2 - n\bar{\hat{\mathbf{y}}}^2 \Leftrightarrow \sum_{j=1}^n (y_j - \bar{\mathbf{y}})^2 = \sum_{j=1}^n (\hat{y}_j - \bar{\mathbf{y}})^2 + \|\hat{\mathbf{r}}\|_2^2 \quad (1.7)$$

La (1.7) dà informazioni sulla qualità del modello adattato che può essere misurata dal *coefficiente di determinazione*

$$R^2 = \frac{\sum_{j=1}^n (\hat{y}_j - \bar{\mathbf{y}})^2}{\sum_{j=1}^n (y_j - \bar{\mathbf{y}})^2} = 1 - \frac{\sum_{j=1}^n \hat{r}_j^2}{\sum_{j=1}^n (y_j - \bar{\mathbf{y}})^2}. \quad (1.8)$$

In particolare, la quantità R^2 mostra la variazione totale nelle y_j attribuibile alle variabili predittrici z_1, z_2, \dots, z_m ([1]). Si ha:

- $R^2 = 1$ se $\hat{\mathbf{r}} = \mathbf{0}$, cioè se l'equazione adattata del modello di regressione lineare passa per tutti i punti prestabiliti;
- $R^2 = 0$ se $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \bar{\mathbf{y}} & 0 & \dots & 0 \end{bmatrix}$, cioè se i predittori non influenzano la risposta.

1.2.2 Proprietà di campionamento degli stimatori ai minimi quadrati classici

Lo stimatore $\hat{\beta}$ di minimi quadrati ed i residui $\hat{\mathbf{r}}$ godono di proprietà di campionamento descritte dal teorema seguente ([1]).

Teorema 1.2.4. *Dato il modello generale di regressione lineare in (1.2), lo stimatore $\hat{\beta}$ ai minimi quadrati ed i residui $\hat{\mathbf{r}}$ soddisfano le seguenti condizioni:*

1. $E(\hat{\beta}) = \beta$
2. $Cov(\hat{\beta}) = \sigma^2 (\mathbf{Z}^T \mathbf{Z})^{-1}$
3. $E(\hat{\mathbf{r}}) = \mathbf{0}$
4. $Cov(\hat{\mathbf{r}}) = \sigma^2 [\mathbf{I} - \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T]$
5. $E(\|\hat{\mathbf{r}}\|_2^2) = (n - m - 1)\sigma^2$

Allora, definito

$$s^2 := \frac{\|\hat{\mathbf{r}}\|_2^2}{n - m - 1} = \frac{\mathbf{Y}^T [\mathbf{I} - \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T] \mathbf{Y}}{n - m - 1}$$

si ha $E(s^2) = \sigma^2$. Inoltre, $\hat{\beta}$ e $\hat{\mathbf{r}}$ non sono correlati ([1]).

Dimostrazione. Prima dell'osservazione della risposta $\mathbf{Y} = \mathbf{Z}\beta + \mathbf{r}$, essa è una variabile aleatoria ([1]). Si ha

$$\hat{\beta} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Y} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T (\mathbf{Z}\beta + \mathbf{r}) = \beta + (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{r}$$

e

$$\hat{\mathbf{r}} = [\mathbf{I}_n - \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T] \mathbf{Y}. \quad (1.9)$$

Sia $\mathcal{P} := \mathbf{I}_n - \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T$. Poichè $\mathcal{P}\mathbf{Z} = \mathbf{Z} - \mathbf{Z} = \mathbf{0}$, segue che

$$\hat{\mathbf{r}} = \mathcal{P} [\mathbf{Z}\beta + \mathbf{r}] = \mathcal{P}\mathbf{r}.$$

Per la linearità del valore atteso di una variabile aleatoria e per le proprietà distributive caratterizzanti (1.2), si ottiene $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T E(\mathbf{r}) = \boldsymbol{\beta}$.

Inoltre, per le proprietà A.0.3 e A.0.4,

$Cov(\hat{\boldsymbol{\beta}}) = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T Cov(\mathbf{r}) \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} = \sigma^2 (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} = \sigma^2 (\mathbf{Z}^T \mathbf{Z})^{-1}$, $E(\hat{\mathbf{r}}) = \mathcal{P}E(\mathbf{r}) = \mathbf{0}$ e $Cov(\hat{\mathbf{r}}) = \mathcal{P}Cov(\mathbf{r})\mathcal{P}^T = \sigma^2 \mathcal{P}$ dove l'ultima uguaglianza segue dall'idempotenza di \mathcal{P} . Infatti $\mathcal{P}^2 = \mathcal{P}$. In più, poichè $\mathbf{Z}^T \mathcal{P} = \mathbf{0}$ e $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{r}$, vale $Cov(\hat{\boldsymbol{\beta}}, \hat{\mathbf{r}}) = E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \hat{\mathbf{r}}^T] = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T E(\mathbf{r} \mathbf{r}^T) \mathcal{P} = \sigma^2 (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathcal{P} = \mathbf{0}$. Dalla (1.9), dall'idempotenza di \mathcal{P} e dal risultato A.0.2, $\|\hat{\mathbf{r}}\|_2^2 = \mathbf{r}^T \mathcal{P} \mathcal{P} \mathbf{r} = \mathbf{r}^T \mathcal{P} \mathbf{r} = tr[\mathbf{r}^T \mathcal{P} \mathbf{r}] = tr(\mathcal{P} \mathbf{r} \mathbf{r}^T)$. Ora, per un'arbitraria matrice \mathbf{W} $n \times n$ a coefficienti random $E(tr(\mathbf{W})) = E(W_{1,1} + \dots + W_{n,n}) = E(W_{1,1}) + \dots + E(W_{n,n}) = tr[E(\mathbf{W})]$ ([1]). Di conseguenza, per il risultato A.0.1, si ottiene $E(\|\hat{\mathbf{r}}\|_2^2) = tr(\mathcal{P}E(\mathbf{r} \mathbf{r}^T)) = \sigma^2 tr(\mathcal{P}) = \sigma^2 tr(\mathbf{I}_n) - \sigma^2 tr[\mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T] = \sigma^2 n - \sigma^2 tr(\mathbf{I}_{m+1}) = \sigma^2 (n - m - 1)$. Da cui segue il risultato per

$$s^2 := \frac{\|\hat{\mathbf{r}}\|_2^2}{n - m - 1}.$$

□

Lo stimatore $\hat{\boldsymbol{\beta}}$ di minimi quadrati soddisfa delle proprietà di minima varianza dimostrate da Gauss e che vertono sulla 'migliore' stima delle funzioni parametriche lineari della forma $\mathbf{c}^T \boldsymbol{\beta} = c_0 \beta_0 + \dots + c_m \beta_m$ ([1]), $\forall \mathbf{c}$.

Teorema 1.2.5 (Teorema ai minimi quadrati di Gauss). ([1])

Sia $\mathbf{Y} = \mathbf{Z} \boldsymbol{\beta} + \mathbf{r}$ come in (1.2), cioè tale che $E(\mathbf{r}) = \mathbf{0}$, $Cov(\mathbf{r}) = \sigma^2 \mathbf{I}$ e $\mathbf{Z} \in \mathbb{R}^{n, m+1}$ sia di rango massimo $m + 1$. Per ogni $\mathbf{c} \in \mathbb{R}^{m+1}$, lo stimatore $\mathbf{c}^T \hat{\boldsymbol{\beta}} = c_0 \hat{\beta}_0 + \dots + c_m \hat{\beta}_m$ di $\mathbf{c}^T \boldsymbol{\beta}$ gode della più piccola varianza possibile tra tutti gli stimatori lineari non distorti della forma $\mathbf{a}^T \mathbf{Y} = a_1 Y_1 + \dots + a_n Y_n$ per $\mathbf{c}^T \boldsymbol{\beta}$.

Dimostrazione. $\forall \mathbf{c} \in \mathbb{R}^{m+1}$ fissato, sia $\mathbf{a}^T \mathbf{Y}$ uno stimatore non distorto di $\mathbf{c}^T \boldsymbol{\beta}$. Allora, $E(\mathbf{a}^T \mathbf{Y}) = \mathbf{c}^T \boldsymbol{\beta}$, qualsiasi sia il valore di $\boldsymbol{\beta}$. Inoltre, per le ipotesi su \mathbf{Y} , si avrà $E(\mathbf{a}^T \mathbf{Y}) = E(\mathbf{a}^T \mathbf{Z} \boldsymbol{\beta} + \mathbf{a}^T \mathbf{r}) = \mathbf{a}^T \mathbf{Z} \boldsymbol{\beta}$. Uguagliando le due

espressioni, segue che: $\mathbf{c}^T \boldsymbol{\beta} = \mathbf{a}^T \mathbf{Z} \boldsymbol{\beta} \Leftrightarrow (\mathbf{c}^T - \mathbf{a}^T \mathbf{Z}) \boldsymbol{\beta} = \mathbf{0}, \forall \boldsymbol{\beta}$. In particolare, si può scegliere $\boldsymbol{\beta} = (\mathbf{c}^T - \mathbf{a}^T \mathbf{Z})^T$. Questo implica che $\mathbf{c}^T = \mathbf{a}^T \mathbf{Z}$ per qualsiasi stimatore non distorto ([1]). Ora, ponendo $\mathbf{a}^* = \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{c}$, si ha $\mathbf{c}^T \hat{\boldsymbol{\beta}} = \mathbf{c}^T (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Y} = \mathbf{a}^{*T} \mathbf{Y}$. In più, dal Teorema 1.2.4, $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, così $\mathbf{c}^T \hat{\boldsymbol{\beta}} = \mathbf{a}^{*T} \mathbf{Y}$ è uno stimatore non distorto di $\mathbf{c}^T \boldsymbol{\beta}$. Per ogni \mathbf{a} tale che soddisfi il requisito di non distorsione $\mathbf{c}^T = \mathbf{a}^T \mathbf{Z}$ ([1]), poichè dalla condizione $(\mathbf{a} - \mathbf{a}^*)^T \mathbf{Z} = \mathbf{a}^T \mathbf{Z} - \mathbf{a}^{*T} \mathbf{Z} = \mathbf{c}^T - \mathbf{c}^T = \mathbf{0}^T$ si ricava che $(\mathbf{a} - \mathbf{a}^*)^T \mathbf{a}^* = (\mathbf{a} - \mathbf{a}^*)^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{c} = \mathbf{0}$, allora $Var(\mathbf{a}^T \mathbf{Y}) = Var(\mathbf{a}^T \mathbf{Z} \boldsymbol{\beta} + \mathbf{a}^T \mathbf{r}) = Var(\mathbf{a}^T \mathbf{r}) = \mathbf{a}^T \mathbf{I} \sigma^2 \mathbf{a} = \sigma^2 \|\mathbf{a} \pm \mathbf{a}^*\|_2^2 = \sigma^2 [\|\mathbf{a} - \mathbf{a}^*\|_2^2 + \|\mathbf{a}^*\|_2^2]$. Segue che $Var(\mathbf{a}^T \mathbf{Y})$ è minimizzata scegliendo $\mathbf{a}^{*T} \mathbf{Y} = \mathbf{c}^T (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Y} = \mathbf{c}^T \hat{\boldsymbol{\beta}}$, in quanto \mathbf{a}^* è fissato e $\|\mathbf{a} - \mathbf{a}^*\|_2^2 > 0 \forall \mathbf{a} \neq \mathbf{a}^*$. \square

Questo importante risultato enuncia che, con la sostituzione di $\boldsymbol{\beta}$ con $\hat{\boldsymbol{\beta}}$, si ottiene il miglior stimatore di $\mathbf{c}^T \boldsymbol{\beta}$ per ogni $\mathbf{c} \in \mathbb{R}^{m+1}$ di interesse. In termini statistici, lo stimatore $\mathbf{c}^T \hat{\boldsymbol{\beta}}$ è chiamato il *miglior stimatore lineare non distorto (di minima varianza)*, in inglese *Best linear unbiased estimator - B.L.U.E.*, di $\mathbf{c}^T \boldsymbol{\beta}$ ([1]).

1.3 Stima della funzione di regressione e predizione di una nuova osservazione in \mathbf{z}_0

([1]) Una volta che il modello di regressione adattato è soddisfacente, può essere usato per risolvere due problemi di predizione.

Sia $\mathbf{z}_0 = [1, z_{0,1}, \dots, z_{0,m}]^T$ il vettore costituito da valori selezionati delle variabili predittrici. Allora \mathbf{z}_0 e $\hat{\boldsymbol{\beta}}$ possono essere sfruttati per stimare

1. la funzione di regressione, $\mathbf{z}_0^T \boldsymbol{\beta} = \beta_0 + \beta_1 z_{0,1} + \dots + \beta_m z_{0,m}$, in \mathbf{z}_0 ;
2. il valore della risposta, \mathbf{Y} , in \mathbf{z}_0 .

Stimare la funzione di regressione in z_0

([1]) Sia \mathbf{Y}_0 il valore della risposta quando le variabili predittrici assumono i valori di $\mathbf{z}_0 = [1, z_{0,1}, \dots, z_{0,m}]^T$. Coerentemente col modello in (1.2), il valore atteso di \mathbf{Y}_0 è

$$E(\mathbf{Y}_0|\mathbf{z}_0) = \beta_0 + \beta_1 z_{0,1} + \dots + \beta_m z_{0,m} = \mathbf{z}_0^T \boldsymbol{\beta}. \quad (1.10)$$

E la sua stima ai minimi quadrati è $\mathbf{z}_0^T \hat{\boldsymbol{\beta}}$.

Predire una nuova osservazione in z_0

La predizione di una nuova osservazione, come per esempio \mathbf{Y}_0 , in $\mathbf{z}_0 = [1, z_{0,1}, \dots, z_{0,m}]^T$ comporta una maggiore incertezza rispetto alla stima del *valore atteso* di \mathbf{Y}_0 . In accordo con il modello in (1.2), $\mathbf{Y}_0 = \mathbf{z}_0^T \boldsymbol{\beta} + \mathbf{r}_0$ ossia

(nuova risposta \mathbf{Y}_0)=(valore atteso di \mathbf{Y}_0 in z_0)+(nuovo errore)

dove \mathbf{r}_0 è distribuito come $N(0, \sigma^2)$ ed è indipendente da \mathbf{r} e, dunque, da $\hat{\boldsymbol{\beta}}$ e da s^2 . Gli errori \mathbf{r} influenzano gli stimatori $\hat{\boldsymbol{\beta}}$ e s^2 attraverso le risposte \mathbf{Y} , ma non \mathbf{r}_0 ([1]).

Capitolo 2

Il problema ai minimi quadrati e metodi di risoluzione

Nel paragrafo 1.2 è stato utilizzato il metodo dei minimi quadrati per stimare i coefficienti di regressione del modello (1.2).

In generale, dati $\mathbf{A} \in \mathbb{R}^{n,m}$, con $n \geq m$ e $\mathbf{b} \in \mathbb{R}^n$, il problema dei minimi quadrati consiste nel determinare il vettore $\mathbf{x} \in \mathbb{R}^m$ che renda minimo il residuo $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$, tipicamente in norma Euclidea, cioè

$$\min_{\mathbf{x} \in \mathbb{R}^m} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2. \quad (2.1)$$

Supposto \mathbf{A} di rango massimo, dal punto di vista numerico sono tre i metodi per risolvere questo problema, basati rispettivamente su:

- l'equazione normale e la fattorizzazione di Cholesky per $\mathbf{A}^T \mathbf{A}$;
- la fattorizzazione QR di \mathbf{A} ;
- la decomposizione in valori singolari di \mathbf{A} .

2.1 L'equazione normale

Il primo metodo risolutivo è suggerito dal Teorema 1.2.2 di esistenza ed unicità della soluzione di (2.1). Infatti, dall'ipotesi di massimo rango di

\mathbf{A} , segue la non singolarità di $\mathbf{A}^T \mathbf{A}$ per il Lemma 1.2.1. Di conseguenza $\mathbf{A}^T \mathbf{A}$ è una matrice $m \times m$ simmetrica definita positiva e, dunque, ammette la sua fattorizzazione di Cholesky $\mathbf{L}\mathbf{L}^T$. Pertanto trovare una soluzione del problema (2.1) è equivalente a risolvere i seguenti sistemi lineare in ordine:

1. $\mathbf{L}y = \mathbf{A}^T b$
2. $\mathbf{L}^T x = y$.

con costo computazionale pari a

- $\frac{1}{2}(2n-1)m^2$ flops per la moltiplicazione $\mathbf{A}^T \mathbf{A}$
- $(2n-1)m$ flops per il prodotto $\mathbf{A}^T b$
- $\frac{m^3}{3} + \mathcal{O}(m^2)$ flops per la fattorizzazione di Cholesky $\mathbf{A}^T \mathbf{A} = \mathbf{L}\mathbf{L}^T$

per un totale di $(m+2)mn + \frac{m^3}{3} + \mathcal{O}(m^2)$ flops.

Da un punto di vista numerico, il metodo dell'equazione normale risulta essere inadeguato per risolvere il problema (2.1) se le colonne di \mathbf{A} sono 'quasi' linearmente dipendenti: infatti, passando in aritmetica con precisione finita, si rischia che questo tipo di matrici dia problemi di stabilità numerica, anche se matematicamente le colonne non sono combinate tra loro linearmente.

2.2 La fattorizzazione QR

Il secondo metodo di risoluzione del problema (2.1) si basa sulla fattorizzazione QR di \mathbf{A} :

$$\mathbf{A} = \mathbf{Q}\mathbf{R} = [\mathbf{Q}_1, \mathbf{Q}_2] \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix} = \mathbf{Q}_1 \mathbf{R}_1$$

con $\mathbf{Q} \in \mathbb{R}^{n,n}$ ortogonale e $\mathbf{R}_1 \in \mathbb{R}^{m,m}$ triangolare superiore invertibile. Sfruttando le proprietà delle matrici ortogonali, si ha che

1. $\mathbf{Q}_2^T \mathbf{Q}_1 = \mathbf{0}$

$$2. \text{Range}(\mathbf{A}) = \text{Range}(\mathbf{Q}_1)$$

$$3. \text{Range}(\mathbf{A})^\perp = \text{Range}(\mathbf{Q}_2)$$

$$4. \|\mathbf{Q}\|_2 = 1$$

risultati che permettono di trovare un'espressione equivalente della funzione che viene minimizzata in (2.1). Nello specifico

$$\begin{aligned} \|b - \mathbf{A}x\|_2^2 &= \|b - \mathbf{Q}\mathbf{R}x\|_2^2 = \|\mathbf{Q}(\mathbf{Q}^T b - \mathbf{R}x)\|_2^2 = \|\mathbf{Q}^T b - \mathbf{R}x\|_2^2 \\ &= \left\| \mathbf{Q}^T b - \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix} \right\|_2^2 = \left\| \begin{bmatrix} \mathbf{Q}_1^T b - \mathbf{R}_1 x \\ \mathbf{Q}_2^T b \end{bmatrix} \right\|_2^2 \\ &= \|\mathbf{Q}_1^T b - \mathbf{R}_1 x\|_2^2 + \|\mathbf{Q}_2^T b\|_2^2. \end{aligned}$$

La minima norma del residuo si ha quando

$$\|\mathbf{Q}_1^T b - \mathbf{R}_1 x\|_2 = 0$$

cioè per l'unica $x = \mathbf{R}_1^{-1} \mathbf{Q}_1^T b$ soluzione del sistema lineare $\mathbf{R}_1 x = \mathbf{Q}_1^T b$. Per tale x si ha

$$\min_{x \in \mathbb{R}^m} \|b - \mathbf{A}x\|_2 = \|\mathbf{Q}_2^T b\|_2$$

dove $\|\mathbf{Q}_2^T b\|_2 = \|\mathbf{Q}_2 \mathbf{Q}_2^T b\|_2$ e $\mathbf{Q}_2 \mathbf{Q}_2^T b$ coincide con la proiezione di b sullo spazio ortogonale alle colonne di \mathbf{A} . Inoltre, la fattorizzazione QR di \mathbf{A} offre una soluzione per ovviare al problema della 'quasi' dipendenza lineare delle colonne di una matrice, che si può riscontrare usando il metodo dell'equazione normale.

2.3 La decomposizione in valori singolari

Il terzo ed ultimo metodo numerico per la risoluzione del problema (2.1) è basato sulla decomposizione a valori singolari (in inglese, *Singular Values Decomposition*, SVD) della matrice \mathbf{A} , un tipo di fattorizzazione che permette di estendere il concetto di diagonalizzazione ad una qualunque matrice rettangolare, ricorrendo a due differenti matrici ortogonali di trasformazione.

Data $\mathbf{A} \in \mathbb{R}^{n,m}$, non è restrittivo supporre $n \geq m$: infatti, se valesse $n \leq m$, basterebbe prendere \mathbf{A}^T . L'esistenza della decomposizione a valori singolari è data dal seguente teorema.

Teorema 2.3.1. ([2]) Sia $\mathbf{A} \in \mathbb{R}^{n,m}$, con $n \geq m$. Allora esistono $\mathbf{U} = [u_1, \dots, u_n] \in \mathbb{R}^{n,n}$, $\mathbf{V} = [v_1, \dots, v_m] \in \mathbb{R}^{m,m}$ ortogonali e $\mathbf{\Sigma} \in \mathbb{R}^{n,m}$ tali che

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (2.2)$$

con

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \\ 0 & \cdots & \sigma_m \\ \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{\Sigma}_1 \\ \mathbf{0} \end{bmatrix}$$

dove $\mathbf{\Sigma}_1 = \text{diag}(\sigma_1, \dots, \sigma_m)$ e $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m \geq 0$.

Definizione 2.1. Si chiamano *vettori singolari sinistri* le colonne $\{u_i\}_{i=1,\dots,n}$ di \mathbf{U} , *vettori singolari destri* le colonne $\{v_j\}_{j=1,\dots,m}$ di \mathbf{V} e *valori singolari* gli scalari $\{\sigma_k\}_{k=1,\dots,m}$.

Osservazione 2. I due seguenti modi equivalenti di scrivere la (2.2)

$$\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma} \Leftrightarrow \mathbf{A}v_i = u_i\sigma_i \quad \forall i = 1, \dots, m \quad (2.3)$$

e

$$\mathbf{A}^T\mathbf{U} = \mathbf{V}\mathbf{\Sigma}^T \Leftrightarrow \mathbf{A}^T u_i = v_i\sigma_i \quad \forall i = 1, \dots, m \quad (2.4)$$

suggeriscono delle proprietà caratterizzanti le matrici \mathbf{A} , \mathbf{U} , \mathbf{V} , di cui è possibile fruire per la risoluzione del problema (2.1) col metodo SVD ed esplicitate nella proposizione seguente.

Proposizione 2.3.2. Sia $\mathbf{A} \in \mathbb{R}^{n,m}$ con $n \geq m$ e sia $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ la sua decomposizione in valori singolari. Sia $p \in \{1, \dots, m\}$ tale che $\sigma_p = \min \{\sigma_i | \sigma_i > 0\}$. Definendo $\mathbf{U}_1 := [u_1, \dots, u_p]$, $\mathbf{U}_2 := [u_{p+1}, \dots, u_n]$, $\mathbf{V}_1 := [v_1, \dots, v_p]$ e $\mathbf{V}_2 := [v_{p+1}, \dots, v_m]$, si hanno le uguaglianze seguenti:

1. \mathbf{A} ha rango massimo se e solo se $p = m$
2. $\text{Range}(\mathbf{A}) = \text{Range}(\mathbf{U}_1)$
3. $\text{Range}(\mathbf{A}^T) = \text{Range}(\mathbf{V}_1)$
4. $\text{Ker}(\mathbf{A}) = \text{Range}(\mathbf{V}_2)$
5. $\text{Ker}(\mathbf{A}^T) = \text{Range}(\mathbf{U}_2)$
6. $\mathbf{U}_1^T \mathbf{U}_2 = \mathbf{0}$

Dimostrazione. Per provare queste proprietà è sufficiente osservare che, date $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ e $\mathbf{A}^T = \mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T$, si ha

$$\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_1 \\ \mathbf{0} \end{bmatrix} = \mathbf{U}_1\mathbf{\Sigma}_1 \quad (2.5)$$

e

$$\mathbf{A}^T\mathbf{U} = \mathbf{V}\mathbf{\Sigma}^T = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \end{bmatrix} = \mathbf{V}_1\mathbf{\Sigma}_1 \quad (2.6)$$

ossia lo spazio generato dalle colonne di \mathbf{A} (rispettivamente \mathbf{A}^T) coincide con lo spazio generato dalle colonne di \mathbf{U}_1 (rispettivamente \mathbf{V}_1). Inoltre, poichè \mathbf{U} è una matrice ortogonale per ipotesi, allora le sue colonne sono vettori a due a due ortogonali tra loro, da cui segue che $\mathbf{U}_1^T \mathbf{U}_2 = \mathbf{0}$. \square

2.3.1 La decomposizione in valori singolari ed il problema ai minimi quadrati

Sia \mathbf{A} di rango massimo, ossia $p = m$ per la Proposizione 2.3.2. Il Teorema 2.3.1 assicura l'esistenza della sua fattorizzazione SVD, utile per risolvere il problema (2.1) di minimi quadrati. In particolare, usando la relazione (2.5) e l'ipotesi di ortogonalità di \mathbf{U} , si ottiene una scrittura alternativa della

funzione da minimizzare:

$$\begin{aligned} \|b - \mathbf{A}x\|_2^2 &= \|b - \mathbf{U}\Sigma\mathbf{V}^T x\|_2^2 = \|\mathbf{U}(\mathbf{U}^T b - \Sigma\mathbf{V}^T x)\|_2^2 = \left\| \mathbf{U}^T b - \underbrace{\Sigma\mathbf{V}^T x}_{=:y} \right\|_2^2 \\ &= \left\| \begin{pmatrix} \mathbf{U}_1^T b \\ \mathbf{U}_2^T b \end{pmatrix} - \begin{pmatrix} \Sigma_1 \\ \mathbf{0} \end{pmatrix} y \right\|_2^2 = \|\mathbf{U}_1^T b - \Sigma_1 y\|_2^2 + \|\mathbf{U}_2^T b\|_2^2. \end{aligned}$$

Osservando che, per minimizzare la norma del residuo, si può agire solo sul primo dei due addendi, si ha l'annullamento di tale termine per $y = \Sigma_1^{-1}\mathbf{U}_1^T b$. Per l'ortogonalità di \mathbf{V} e per definizione di y , segue che $x = \mathbf{V}y = \mathbf{V}\Sigma_1^{-1}\mathbf{U}_1^T b$, corrispondente alla soluzione cercata, e

$$\min_{x \in \mathbb{R}^m} \|b - \mathbf{A}x\|_2 = \|\mathbf{U}_2^T b\|_2^2.$$

Osservazione 3. La soluzione $x = \mathbf{V}y = \mathbf{V}\Sigma_1^{-1}\mathbf{U}_1^T b$ può essere espressa anche come combinazione lineare dei vettori singolari destri: $x = \sum_{k=1}^m \left(\frac{u_k^T b}{\sigma_k} \right) v_k$.

Osservazione 4. Data una matrice $\mathbf{A} \in \mathbb{R}^{n,m}$, con $n \geq m$, la norma-2 è definita come

$$\|\mathbf{A}\|_2^2 := \max_{\|x\|_2=1} \|\mathbf{A}x\|_2^2 = \max_{\|x\|_2=1} x^T \mathbf{A}^T \mathbf{A} x = \lambda_{\max}(\mathbf{A}^T \mathbf{A}),$$

dove nell'ultima uguaglianza sono state usate le proprietà del quoziente di Rayleigh.

L'osservazione 4 ed alcune proprietà elementari delle norme permettono di dare la definizione di *numero di condizionamento di una matrice rettangolare*, concetto che servirà per stimare dall'alto l'errore relativo della soluzione del problema ai minimi quadrati nel caso in cui si ricorra alla decomposizione in valori singolari.

Osservazione 5. Data $\mathbf{A} \in \mathbb{R}^{n,m}$, con $n \geq m$, valgono:

$$\|\mathbf{A}\|_2 = \max_{\|x\|_2=1} \|\mathbf{A}x\|_2 = \sigma_1$$

e

$$\min_{\|x\|_2=1} \|\mathbf{A}x\|_2 = \sigma_m.$$

Dimostrazione. Infatti, per l'ortogonalità di \mathbf{U} , $\forall x \in \mathbb{R}^m$ tale che $\|x\|_2 = 1$ si può scrivere

$$\begin{aligned} \|\mathbf{A}x\|_2 &= \|\mathbf{U}\Sigma\mathbf{V}^T x\|_2 = \left\| \underbrace{\Sigma\mathbf{V}^T x}_{=:y} \right\|_2 = \|\Sigma y\|_2 \\ &= \left\| \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \\ 0 & \cdots & \sigma_m \\ \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} y \right\|_2 = \left\| \begin{bmatrix} \Sigma_1 \\ \mathbf{0} \end{bmatrix} y \right\|_2 \\ &= \|\Sigma_1 y\|_2 \leq \|\Sigma_1\|_2 \|y\|_2 = \sigma_1 \end{aligned}$$

dove l'ultima uguaglianza è stata ottenuta osservando che $\|y\|_2 = 1$ e dove σ_1 è il massimo valore singolare di \mathbf{A} . Inoltre, tale valore è assunto quando $x = v_1$: infatti $\|\mathbf{A}v_1\|_2 = \|u_1\sigma_1\|_2 = \sigma_1$, provando così la prima parte dell'enunciato. Analogamente, si mostra che

$\forall x \in \mathbb{R}^m$ tale che $\|x\|_2 = 1$ vale la disuguaglianza $\|\mathbf{A}x\|_2 \geq \sigma_m$ e $\|\mathbf{A}v_m\|_2 = \|u_m\sigma_m\|_2 = \sigma_m$, ossia

$$\min_{\|x\|_2=1} \|\mathbf{A}x\|_2 = \sigma_m.$$

□

Definizione 2.2. Se $\sigma_m \neq 0$, si definisce il *numero di condizionamento* di una generica matrice rettangolare $\mathbf{B} \in \mathbb{R}^{n,m}$ come

$$k_2(\mathbf{B}) := \frac{\sigma_1(\mathbf{B})}{\sigma_{\min\{n,m\}}(\mathbf{B})}, \quad \text{se } \sigma_{\min\{n,m\}}(\mathbf{B}) \neq 0.$$

Osservazione 6. La decomposizione in valori singolari è una generalizzazione di quella in autovalori e sono collegate tra loro: in particolare, per $\mathbf{A} \in \mathbb{R}^{n,m}$, con $n \geq m$, si ha $\mathbf{A}^T \mathbf{A} = (\mathbf{U}\Sigma\mathbf{V}^T)^T (\mathbf{U}\Sigma\mathbf{V}^T) = \mathbf{V}\Sigma_1^2 \mathbf{V}^T$, ossia la decomposizione spettrale di $\mathbf{A}^T \mathbf{A}$. In generale, vale anche la seguente relazione: $\sigma_i(\mathbf{A}) = \sqrt{\lambda_i(\mathbf{A}^T \mathbf{A})}$, $i = 1, \dots, m$ e, nel caso di \mathbf{A} normale, $\sigma_i(\mathbf{A}) = |\lambda_i(\mathbf{A})|$.

2.4 Considerazioni sul condizionamento dei metodi di risoluzione trattati

Dato un problema matematico di cui è possibile trovarne la soluzione esatta, uno degli interessi dell'analisi numerica è lo studio della *propagazione degli errori*, in particolare di come eventuali perturbazioni dei dati del problema possano influenzarne i risultati in maniera più o meno significativa: nel primo caso si parlerà di *mal condizionamento*, nel secondo caso, invece, di problema *ben condizionato*. Gli enunciati seguenti mostrano, rispettivamente, come la fattorizzazione QR permetta di risolvere, in pratica, un problema 'vicino' a quello originario (2.1), risultando dunque stabile all'indietro, e l'errore ottenuto nella soluzione del problema dei minimi quadrati quando viene usato l'approccio della SVD.

Teorema 2.4.1. ([2]) *Il vettore $\hat{x} \in \mathbb{R}^m$, approssimazione della soluzione del problema (2.1) ottenuta tramite fattorizzazione QR , è tale che*

$$\hat{x} = \operatorname{argmin}_{x \in \mathbb{R}^m} \|(\mathbf{A} + \delta\mathbf{A})x - (b + \delta b)\|_2,$$

dove

$$\|\delta\mathbf{A}\|_F \leq (6n - 3m + 40) m\mathbf{u} \|\mathbf{A}\|_F + \mathcal{O}(\mathbf{u}^2)$$

e

$$\|\delta b\|_2 \leq (6n - 3m + 40) m\mathbf{u} \|b\|_2 + \mathcal{O}(\mathbf{u}^2).$$

Teorema 2.4.2. ([2]) *Sia $\mathbf{A} \in \mathbb{R}^{n,m}$, con $n \geq m$, una matrice di rango massimo e sia $\delta\mathbf{A} \in \mathbb{R}^{n,m}$ una sua perturbazione tale che $\|\delta\mathbf{A}\|_2 < \sigma_m(\mathbf{A})$. Siano poi $b \in \mathbb{R}^n$, $b \neq 0$ e $\delta b \in \mathbb{R}^n$. Allora si ha che*

1. $\mathbf{A} + \delta\mathbf{A}$ ha rango massimo
2. detta $x + \delta x$ la soluzione del problema dei minimi quadrati perturbato

$$\min_{y \in \mathbb{R}^m} \|(b + \delta b) - (\mathbf{A} + \delta\mathbf{A})y\|_2,$$

risulta

$$\frac{\|\delta x\|_2}{\|x\|_2} \leq \frac{k_2(\mathbf{A})}{1 - \varepsilon_{\mathbf{A}} k_2(\mathbf{A})} \left(\left(1 + k_2(\mathbf{A}) \frac{\|r\|_2}{\|\mathbf{A}\|_2 \|x\|_2} \right) \varepsilon_{\mathbf{A}} + \frac{\|b\|_2}{\|\mathbf{A}\|_2 \|x\|_2} \varepsilon_b \right),$$

dove $r = b - \mathbf{A}x$ è il vettore residuo,

$$\varepsilon_{\mathbf{A}} = \frac{\|\delta \mathbf{A}\|_2}{\|\mathbf{A}\|_2} \quad e \quad \varepsilon_b = \frac{\|\delta b\|_2}{\|b\|_2}.$$

Capitolo 3

Studio di due casi applicativi

In questo capitolo verranno analizzati due insiemi di dati e mostrati alcuni risultati relativi alle stime di minimi quadrati dei coefficienti di regressione ed alla norma minimizzata dei residui, in base a tre diversi metodi:

1. l'equazione normale
2. la fattorizzazione QR della matrice \mathbf{Z} dei predittori
3. la costruzione di una nuova matrice \mathbf{Z}_{new} di predittori, ottenuti prendendo una particolare combinazione lineare delle variabili predittrici iniziali.

Più precisamente, per il nuovo k -esimo predittore, i 'pesi' sono le componenti del k -esimo autovettore dominante di $\mathbf{Z}^T \mathbf{Z}$ tale che il corrispondente autovalore dominante λ_k , considerati gli autovalori dominanti λ_i della matrice, soddisfa la disuguaglianza

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^m \lambda_i} \geq 0.8. \quad (3.1)$$

Inoltre, si considererà anche il caso in cui la disuguaglianza soddisfatta dall'autovalore dominante λ_k è

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^m \lambda_i} \geq 0.95. \quad (3.2)$$

Questa procedura equivale approssimativamente ad una pre-elaborazione dei dati mediante analisi delle componenti principali (vedi [1, Capitolo 8]) e corrisponde ad effettuare una SVD troncata della matrice delle variabili predittive, in cui vengono eliminati i vettori singolari corrispondenti ai valori singolari sotto ad una certa soglia. Per tale scopo sono stati scritti, in codice MATLAB, le due funzioni `REGRESSLIN` (in Appendice B.1.1) e `regresslin095` (in Appendice B.1.2), ed i due file ‘LONGLEY script’ (in Appendice B.1.3) e ‘DETROIT script’ (in Appendice B.1.4): le funzioni hanno la matrice \mathbf{X} dei dati ed il vettore risposta \mathbf{y} come argomenti in entrata e mostrano, in uscita, i numeri di condizionamento delle matrici $\mathbf{Z}^T \mathbf{Z}$ e $\mathbf{Z}_{new}^T \mathbf{Z}_{new}$, le stime ai minimi quadrati dei parametri di regressione, i valori approssimati delle risposte, i residui e le minime norme residue, ottenuti applicando i tre metodi di cui sopra, sia in valenza della (3.1) che della (3.2). Negli script, invece, sono presenti la matrice \mathbf{X} dei dati originali e la risposta \mathbf{y} , costituiti dai rispettivi valori dei due dataset da analizzare, e \mathbf{X}_{scaled} , che è la matrice dei dati \mathbf{X} standardizzata ed ottenuta prendendo i valori campionari dei predittori sottratti della rispettiva media aritmetica e divisi per la corrispondente deviazione standard. \mathbf{X} , \mathbf{y} e \mathbf{X}_{scaled} saranno gli argomenti in entrata nelle chiamate delle funzioni `REGRESSLIN` (in Appendice B.1.1) e `regresslin095` (in Appendice B.1.2). Infine, a differenza di ‘LONGLEY script’ (in Appendice B.1.3), nel file ‘DETROIT script’ (in Appendice B.1.4) è stato aggiunto un algoritmo per avere maggiori informazioni sulla dipendenza lineare di un particolare predittore rispetto agli altri.

3.1 LONGLEY dataset

Il primo insieme di dati analizzato è noto, in letteratura, con il nome di ‘LONGLEY dataset’, tratti dalla rivista scientifica ‘An appraisal of least-squares programs from the point of view of the user’, JASA, 62(1967) p. 819–841. Descrivono il numero di persone impiegate in funzione, rispettivamente, del deflatore del PIL (ossia il rapporto tra il PIL nominale di un Paese ed

il suo PIL reale, espresso in percentuali), del Prodotto Interno Lordo, del numero di persone disoccupate, delle forze armate, della popolazione non censita di età maggiore o uguale a 14 anni, e dell'anno. Eseguendo il file 'LONGLEY script' (in Appendice B.1.3), si ottengono i risultati seguenti:

```
*****
***** CONDIZIONE SUGLI AUTOVALORI: >= 0.8 *****
*****
***** Analisi con variabili originali *****
*****
num.cond. originale 2.36310e+019 num. cond ridotto 3.36598e+012 con m=1 termini
```

Stime dei coeff. di regressione coi primi due metodi:

```
-3.4823e+006 -3.4823e+006
 1.5062e+001  1.5062e+001
-3.5819e-002 -3.5819e-002
-2.0202e+000 -2.0202e+000
-1.0332e+000 -1.0332e+000
-5.1104e-002 -5.1104e-002
 1.8292e+003  1.8292e+003
```

Errore relativo tra i due coeff 1.092926e-008 :

Stime dei coeff. di regressione col terzo metodo:

```
5.0952e+004
3.5461e-002
```

Stime delle risposte coi 3 metodi:

```
6.0056e+004  6.0056e+004  5.9995e+004
6.1216e+004  6.1216e+004  6.0861e+004
6.0125e+004  6.0125e+004  6.0826e+004
6.1597e+004  6.1597e+004  6.1742e+004
6.2911e+004  6.2911e+004  6.3264e+004
6.3888e+004  6.3888e+004  6.3890e+004
6.5153e+004  6.5153e+004  6.4534e+004
6.3774e+004  6.3774e+004  6.4468e+004
6.6005e+004  6.6005e+004  6.5649e+004
6.7402e+004  6.7402e+004  6.6402e+004
6.8186e+004  6.8186e+004  6.7222e+004
6.6552e+004  6.6552e+004  6.7298e+004
6.8811e+004  6.8811e+004  6.8611e+004
6.9650e+004  6.9650e+004  6.9309e+004
6.8989e+004  6.8989e+004  6.9864e+004
7.0758e+004  7.0758e+004  7.1136e+004
```

Residui coi 3 metodi:

```

2.6734e+002  2.6734e+002  3.2767e+002
-9.4014e+001 -9.4014e+001  2.6057e+002
 4.6287e+001  4.6287e+001 -6.5539e+002
-4.1011e+002 -4.1011e+002 -5.5470e+002
 3.0971e+002  3.0971e+002 -4.3182e+001
-2.4931e+002 -2.4931e+002 -2.5085e+002
-1.6405e+002 -1.6405e+002  4.5504e+002
-1.3180e+001 -1.3180e+001 -7.0710e+002
 1.4305e+001  1.4305e+001  3.6961e+002
 4.5539e+002  4.5539e+002  1.4550e+003
-1.7269e+001 -1.7269e+001  9.4680e+002
-3.9055e+001 -3.9055e+001 -7.8503e+002
-1.5555e+002 -1.5555e+002  4.3799e+001
-8.5671e+001 -8.5671e+001  2.5542e+002
 3.4193e+002  3.4193e+002 -5.3267e+002
-2.0676e+002 -2.0676e+002 -5.8503e+002

```

Norma dei residui coi 3 metodi:

```
9.1456e+002  9.1456e+002  2.4787e+003
```

```
*****
```

```
***** normalizzazione delle variabili *****
```

```
*****
```

```
num.cond. originale 1.2220e+004 num. cond ridotto 4.31567e+000 con m=2 termini
```

Stime dei coeff. di regressione coi primi due metodi:

```

6.5317e+004  6.5317e+004
 1.6254e+002  1.6254e+002
-3.5602e+003 -3.5602e+003
-1.8878e+003 -1.8878e+003
-7.1904e+002 -7.1904e+002
-3.5549e+002 -3.5549e+002
 8.7085e+003  8.7085e+003

```

Errore relativo tra i due coeff 1.217864e-013 :

Stime dei coeff. di regressione col terzo metodo:

```

6.5317e+004
 1.5651e+003
 3.9183e+002

```

Stime delle risposte coi 3 metodi:

```

6.0056e+004  6.0056e+004  5.9578e+004
6.1216e+004  6.1216e+004  6.0273e+004
6.0125e+004  6.0125e+004  6.1046e+004
6.1597e+004  6.1597e+004  6.1540e+004
6.2911e+004  6.2911e+004  6.3550e+004

```



```
6.3888e+004 6.3888e+004 6.4489e+004
6.5153e+004 6.5153e+004 6.4966e+004
6.3774e+004 6.3774e+004 6.5600e+004
6.6005e+004 6.6005e+004 6.5744e+004
6.7402e+004 6.7402e+004 6.6235e+004
6.8186e+004 6.8186e+004 6.6980e+004
6.6552e+004 6.6552e+004 6.7823e+004
6.8811e+004 6.8811e+004 6.8202e+004
6.9650e+004 6.9650e+004 6.8817e+004
6.8989e+004 6.8989e+004 6.9747e+004
7.0758e+004 7.0758e+004 7.0483e+004
```

Residui coi 3 metodi:

```
2.6734e+002 2.6734e+002 7.4524e+002
-9.4014e+001 -9.4014e+001 8.4947e+002
4.6287e+001 4.6287e+001 -8.7506e+002
-4.1011e+002 -4.1011e+002 -3.5271e+002
3.0971e+002 3.0971e+002 -3.2921e+002
-2.4931e+002 -2.4931e+002 -8.4997e+002
-1.6405e+002 -1.6405e+002 2.2766e+001
-1.3180e+001 -1.3180e+001 -1.8392e+003
1.4305e+001 1.4305e+001 2.7452e+002
4.5539e+002 4.5539e+002 1.6219e+003
-1.7269e+001 -1.7269e+001 1.1889e+003
-3.9055e+001 -3.9055e+001 -1.3098e+003
-1.5555e+002 -1.5555e+002 4.5317e+002
-8.5671e+001 -8.5671e+001 7.4714e+002
3.4193e+002 3.4193e+002 -4.1555e+002
-2.0676e+002 -2.0676e+002 6.8293e+001
```

Norma dei residui coi 3 metodi:

```
9.1456e+002 9.1456e+002 3.6273e+003
```

```
*****
***** CONDIZIONE SUGLI AUTOVALORI: >= 0.95 *****
*****
***** Analisi con variabili originali *****
*****
numero cond ridotto 3.36598e+012 con m=1 termini
```

Stime dei coeff. di regressione col terzo metodo:

```
5.0952e+004
3.5461e-002
```

Stime delle risposte col terzo metodo:

```
5.9995e+004
6.0861e+004
```

```
6.0826e+004
6.1742e+004
6.3264e+004
6.3890e+004
6.4534e+004
6.4468e+004
6.5649e+004
6.6402e+004
6.7222e+004
6.7298e+004
6.8611e+004
6.9309e+004
6.9864e+004
7.1136e+004
```

Residui col terzo metodo:

```
3.2767e+002
2.6057e+002
-6.5539e+002
-5.5470e+002
-4.3182e+001
-2.5085e+002
4.5504e+002
-7.0710e+002
3.6961e+002
1.4550e+003
9.4680e+002
-7.8503e+002
4.3799e+001
2.5542e+002
-5.3267e+002
-5.8503e+002
```

Norma dei residui col terzo metodo:

```
2.4787e+003
```

```
*****
***** normalizzazione delle variabili *****
*****
numero cond ridotto 4.31567e+000 con m=2 termini
```

Stime dei coeff. di regressione col terzo metodo:

```
6.5317e+004
1.5651e+003
3.9183e+002
```

Stime delle risposte col terzo metodo:

```
5.9578e+004
6.0273e+004
6.1046e+004
6.1540e+004
6.3550e+004
6.4489e+004
6.4966e+004
6.5600e+004
6.5744e+004
6.6235e+004
6.6980e+004
6.7823e+004
6.8202e+004
6.8817e+004
6.9747e+004
7.0483e+004
```

Residui col terzo metodo:

```
7.4524e+002
8.4947e+002
-8.7506e+002
-3.5271e+002
-3.2921e+002
-8.4997e+002
2.2766e+001
-1.8392e+003
2.7452e+002
1.6219e+003
1.1889e+003
-1.3098e+003
4.5317e+002
7.4714e+002
-4.1555e+002
6.8293e+001
```

Norma dei residui col terzo metodo:

```
3.6273e+003
```

Esaminando i risultati ottenuti dall'analisi delle variabili originali e soffermandosi sui metodi, rispettivamente, dell'equazione normale e della fattorizzazione QR della matrice dei predittori, si può osservare un forte mal condizionamento del problema ai minimi quadrati corrispondente a tali dati: il numero di condizionamento della matrice $\mathbf{Z}^T \mathbf{Z}$ del sistema normale è

molto maggiore dell'unità, in particolare

$$k_2(\mathbf{Z}^T \mathbf{Z}) = 2.36 \cdot 10^{19} > \frac{1}{\epsilon_M},$$

dove $\epsilon_M = 2.22 \cdot 10^{-16}$ è l'epsilon di macchina. Inoltre, da risultati teorici relativi alla SVD di \mathbf{Z} ed alla sua fattorizzazione QR ridotta, si ha che

$$k_2(\mathbf{Z}^T \mathbf{Z}) = k_2(R_1^T R_1) = [k_2(R_1)]^2. \quad (3.3)$$

Questo comporta una minore accuratezza della soluzione dell'equazione normale rispetto ai coefficienti di regressione stimati usando il secondo metodo. Nello specifico, l'errore relativo commesso rispetto a questi ultimi è di $1.09 \cdot 10^{-8}$, ben otto ordini di grandezza in più rispetto ad ϵ_M . Con l'applicazione del terzo metodo nel caso in cui vale la condizione (3.1), invece, il numero di condizionamento della matrice $\mathbf{Z}_{new}^T \mathbf{Z}_{new}$ con $m = 1$ predittori risulta essere $3.37 \cdot 10^{12}$, sette ordini di grandezza in meno rispetto al primo caso; tuttavia, il mal condizionamento del problema permane. In termini di norme residue, non si osserva una grande differenza tra i tre metodi: ponendo l'attenzione sulle cifre decimali maggiormente significative, la norma dei residui, ottenuta con i primi due, è di $9.15 \cdot 10^2$ contro $2.48 \cdot 10^3$ del terzo. Uno degli obiettivi della regressione lineare è quello di predire il valore atteso della risposta in base a determinati valori dei predittori, come espresso dalla (1.10). A tal proposito, è possibile fare delle considerazioni sull'equazione del modello adattato per LONGLEY dataset: analizzando gli ordini di grandezza dei coefficienti di regressione stimati con i primi due metodi, si potrebbe ipotizzare che il PIL ed il numero della popolazione non censita con più di 14 anni non influiscano in maniera significativa sulla predizione del numero di persone impiegate, poichè i corrispettivi valori delle stime ai minimi quadrati dei parametri di regressione sono dell'ordine di 10^{-2} . Tuttavia si incorrerebbe in un errore: infatti, osservando i rispettivi risultati nel caso dei dati standardizzati, il loro ordine di grandezza è 10^2 , ed il 10^{-2} di cui sopra sarà bilanciato dagli ordini di grandezza dei valori assunti dal secondo e quinto predittore, coerentemente con quello del valore stimato della risposta. Col

passaggio dai dati originali a quelli standardizzati, si guadagna in termini di accuratezza della soluzione ai minimi quadrati: per i primi due metodi, si ha che il numero di condizionamento di $\mathbf{Z}^T \mathbf{Z}$ è $1.22 \cdot 10^4$, mentre quello di $\mathbf{Z}_{new}^T \mathbf{Z}_{new}$ con $m = 2$ termini è 4.32; l'errore relativo tra i due coefficienti di regressione stimati con i primi due metodi è di $1.22 \cdot 10^{-13}$ ed, in termini di minima norma residua, si ottiene $9.15 \cdot 10^2$ con i metodi dell'equazione normale e della fattorizzazione QR , $3.63 \cdot 10^3$ con il terzo metodo. Inoltre, applicando quest'ultimo, si osservano risultati identici indipendentemente che valga la condizione (3.1) o (3.2), sia nel caso dei dati originali che di quelli standardizzati.

3.2 DETROIT dataset

Il secondo set di dati è noto in letteratura con il nome di 'DETROIT' ed è stato oggetto di studio di diversi libri o riviste scientifiche, quali 'Subset selection in regression' di Alan J. Miller pubblicato nella collana Chapman & Hall di monografie sulla Statistica e la Probabilità applicata, no. 40., 'Regression analysis and its application: A data-oriented approach' di Gunst & Mason, Statistics textbooks and monographs no. 24, Marcel Dekker. I dati furono collezionati da J.C. Fisher ed usati nel suo paper: 'Homicide in Detroit: The Role of Firearms', Criminology, vol.14, 387 – 400 (1976). I dati descrivono il tasso di omicidi avvenuti nella città di Detroit tra il 1961 e il 1973, in funzione rispettivamente del numero di poliziotti a tempo pieno ogni 100,000 abitanti, la percentuale di abitanti disoccupati, il numero (espresso in migliaia) di operai manifatturieri, il numero di porto d'armi ogni 100,000 abitanti, il numero di pistole registrate ogni 100,000 abitanti, la percentuale di omicidi seguiti da arresto, il numero di uomini bianchi nella popolazione, il numero di operai non manifatturieri espresso in migliaia, il numero di lavoratori statali espresso in migliaia, i guadagni medi ad ora, i guadagni medi a settimana, il tasso di morte per incidenti ogni 100,000 abitanti ed il numero di assolti ogni 100,000 abitanti. Per studiare tali dati, è necessario prede-

re un suo sottoinsieme, poichè, altrimenti, darebbero luogo ad una matrice quasi singolare e molto mal condizionata: a tal fine, è sufficiente eliminare il settimo predittore, ossia il numero di uomini bianchi nella popolazione. Infatti, i valori della settima variabile predittrice sono ridondanti, in quanto linearmente dipendente dagli altri. Eseguendo il file 'DETROIT script' (in Appendice B.1.4), si ottengono i risultati seguenti:

```
*****
**studio dipendenza lineare colonna 7 dalle altre colonne**
*****
Columns 1 through 7

5.4739e+003  1.9480e+005  3.9492e+003  1.2611e+003  1.2632e+003  8.2409e+004  2.5166e+003

Columns 8 through 12

7.5718e+003  3.2353e+005  7.4233e+003  5.4702e+004  4.4493e+003

*****
***** CONDIZIONE SUGLI AUTOVALORI: >= 0.8 *****
*****
***** Analisi con variabili originali *****
*****
num.cond. originale 2.33397e+010 num. cond ridotto 2.49757e+007 con m=1 termini

Stime dei coeff. di regressione coi primi due metodi:
-3.3128e+001 -3.3128e+001
 1.6706e-002  1.6706e-002
 9.2858e-001  9.2858e-001
-2.7793e-002 -2.7793e-002
 1.7738e-002  1.7738e-002
-4.1702e-003 -4.1702e-003
-5.0400e-002 -5.0400e-002
-2.3252e-003 -2.3252e-003
 1.9582e-001  1.9582e-001
-5.8219e+000 -5.8219e+000
 2.5775e-001  2.5775e-001
-9.9188e-003 -9.9188e-003
 1.5797e-002  1.5797e-002

Errore relativo tra i due coeff 4.435342e-012 :

Stime dei coeff. di regressione col terzo metodo:
-2.4941e+001
 3.9481e-002
```

Stime delle risposte coi 3 metodi:

8.6000e+000	8.6000e+000	7.3381e+000
8.9000e+000	8.9000e+000	7.1702e+000
8.5200e+000	8.5200e+000	8.9337e+000
8.8900e+000	8.8900e+000	1.0493e+001
1.3070e+001	1.3070e+001	1.4592e+001
1.4570e+001	1.4570e+001	1.8397e+001
2.1360e+001	2.1360e+001	2.8678e+001
2.8030e+001	2.8030e+001	4.5290e+001
3.1490e+001	3.1490e+001	3.7310e+001
3.7390e+001	3.7390e+001	3.5352e+001
4.6260e+001	4.6260e+001	3.6653e+001
4.7240e+001	4.7240e+001	3.9311e+001
5.2330e+001	5.2330e+001	3.7133e+001

Residui coi 3 metodi:

-1.0481e-013	-1.7764e-015	1.2619e+000
1.6183e-012	-5.3291e-015	1.7298e+000
-3.2188e-012	3.5527e-015	-4.1369e-001
3.7179e-012	1.0658e-014	-1.6026e+000
-2.3270e-012	5.3291e-015	-1.5219e+000
7.8160e-013	-3.5527e-015	-3.8269e+000
-6.8212e-013	0	-7.3179e+000
6.4304e-013	-1.0658e-014	-1.7260e+001
-2.8422e-014	-3.5527e-015	-5.8199e+000
-8.8107e-013	7.1054e-015	2.0380e+000
9.9476e-014	-2.1316e-014	9.6072e+000
-1.2790e-013	-1.4211e-014	7.9290e+000
4.9027e-013	0	1.5197e+001

Norma dei residui coi 3 metodi:

5.8959e-012	3.2122e-014	2.8283e+001
-------------	-------------	-------------

```
*****
***** normalizzazione delle variabili *****
*****
```

num.cond. originale 3.88302e+004 num. cond ridotto 7.07838e+000 con m=2 termini

Stime dei coeff. di regressione coi primi due metodi:

2.5127e+001	2.5127e+001
7.8205e-001	7.8205e-001
2.1907e+000	2.1907e+000
-1.3847e+000	-1.3847e+000
5.6125e+000	5.6125e+000
-1.2970e+000	-1.2970e+000
-6.3803e-001	-6.3803e-001

```
-2.2036e-001 -2.2036e-001
 7.2525e+000 7.2525e+000
-5.6273e+000 -5.6273e+000
 1.0957e+001 1.0957e+001
-5.0979e-002 -5.0979e-002
 1.1546e+000 1.1546e+000
```

Errore relativo tra i due coeff 5.340349e-013 :

Stime dei coeff. di regressione col terzo metodo:

```
2.5127e+001
5.8780e+000
5.4357e-001
```

Stime delle risposte coi 3 metodi:

```
8.6000e+000 8.6000e+000 7.5783e+000
8.9000e+000 8.9000e+000 9.6086e+000
8.5200e+000 8.5200e+000 8.8999e+000
8.8900e+000 8.8900e+000 9.8113e+000
1.3070e+001 1.3070e+001 1.2883e+001
1.4570e+001 1.4570e+001 1.5728e+001
2.1360e+001 2.1360e+001 2.1471e+001
2.8030e+001 2.8030e+001 2.5616e+001
3.1490e+001 3.1490e+001 3.2999e+001
3.7390e+001 3.7390e+001 3.6374e+001
4.6260e+001 4.6260e+001 4.2808e+001
4.7240e+001 4.7240e+001 4.9694e+001
5.2330e+001 5.2330e+001 5.3179e+001
```

Residui coi 3 metodi:

```
4.2633e-014 -1.7764e-015 1.0217e+000
-2.8066e-013 -1.7764e-015 -7.0858e-001
5.6133e-013 -1.2434e-014 -3.7989e-001
-3.1619e-013 -5.3291e-015 -9.2126e-001
-1.5632e-013 -7.1054e-015 1.8742e-001
2.8422e-013 -1.2434e-014 -1.1582e+000
-1.5987e-013 -3.5527e-015 -1.1130e-001
8.8818e-014 -7.1054e-015 2.4141e+000
-4.6185e-014 -1.0658e-014 -1.5092e+000
-1.3500e-013 0 1.0157e+000
9.9476e-014 -7.1054e-015 3.4525e+000
0 1.4211e-014 -2.4540e+000
2.8422e-014 7.1054e-015 -8.4891e-001
```

Norma dei residui coi 3 metodi:

```
8.1571e-013 2.9564e-014 5.6327e+000
```



```
*****
***** CONDIZIONE SUGLI AUTOVALORI: >= 0.95 *****
*****
***** Analisi con variabili originali *****
*****
numero cond ridotto 2.49757e+007 con m=1 termini
```

Stime dei coeff. di regressione col terzo metodo:

```
-2.4941e+001
3.9481e-002
```

Stime delle risposte col terzo metodo:

```
7.3381e+000
7.1702e+000
8.9337e+000
1.0493e+001
1.4592e+001
1.8397e+001
2.8678e+001
4.5290e+001
3.7310e+001
3.5352e+001
3.6653e+001
3.9311e+001
3.7133e+001
```

Residui col terzo metodo:

```
1.2619e+000
1.7298e+000
-4.1369e-001
-1.6026e+000
-1.5219e+000
-3.8269e+000
-7.3179e+000
-1.7260e+001
-5.8199e+000
2.0380e+000
9.6072e+000
7.9290e+000
1.5197e+001
```

Norma dei residui col terzo metodo:

```
2.8283e+001
```

```
*****
***** normalizzazione delle variabili *****
*****
```

numero cond ridotto 9.05034e+000 con m=3 termini

Stime dei coeff. di regressione col terzo metodo:

2.5127e+001
5.8780e+000
5.4357e-001
1.0318e+000

Stime delle risposte col terzo metodo:

8.5364e+000
9.6103e+000
8.4797e+000
9.0149e+000
1.1777e+001
1.4824e+001
2.1764e+001
2.7817e+001
3.3237e+001
3.6971e+001
4.3146e+001
4.9551e+001
5.1922e+001

Residui col terzo metodo:

6.3553e-002
-7.1025e-001
4.0269e-002
-1.2491e-001
1.2927e+000
-2.5387e-001
-4.0416e-001
2.1295e-001
-1.7466e+000
4.1918e-001
3.1144e+000
-2.3110e+000
4.0780e-001

Norma dei residui col terzo metodo:

4.5719e+000

Osservando tali risultati, si possono aggiungere delle considerazioni sulla dipendenza lineare della settima variabile dalle altre: in particolare, essa risulta essere più linearmente dipendente con il secondo, sesto, decimo e dodicesimo predittore, come suggeriscono gli ordini di grandezza 10^4 e 10^5 dei numeri

di condizionamento delle matrici costituite dalla settima e, rispettivamente, dalla colonna 2, 6, 10 e 12 della matrice dei dati. Ponendo l'attenzione sui risultati relativi ai dati originali, si può osservare come, sia in valenza di (3.1) che di (3.2), i risultati ottenuti applicando la strategia di riduzione non varino e come il problema risulti mal condizionato con tutti e tre i metodi: il numero di condizionamento della matrice $\mathbf{Z}^T \mathbf{Z}$ del sistema normale è $2.33 \cdot 10^{10}$, mentre quello di $\mathbf{Z}_{new}^T \mathbf{Z}_{new}$ con $m = 1$ termini è $2.5 \cdot 10^7$. Inoltre, poichè vale (3.3), si ha una maggiore accuratezza della soluzione del problema ai minimi quadrati ottenuta con la fattorizzazione QR di \mathbf{Z} rispetto al primo metodo. L'errore relativo commesso ricorrendo a quest'ultimo è di $4.44 \cdot 10^{-12}$, quattro ordini di grandezza in più rispetto ad ϵ_M . In termini di norme residue, si ottiene rispettivamente $5.9 \cdot 10^{-12}$ con il primo metodo risolutivo, $3.21 \cdot 10^{-14}$ con il secondo e $2.83 \cdot 10$ con il terzo. Da ciò si può dedurre che la strategia di riduzione non è efficace in questo caso. Si possono, inoltre, aggiungere delle considerazioni sull'equazione del modello adattato per il sottoinsieme dei dati DETROIT studiato: esaminando gli ordini di grandezza dei coefficienti di regressione stimati con i primi due metodi, si potrebbe ipotizzare che il quinto, il settimo e l'undicesimo predittore siano quelli meno influenti nella predizione del numero di omicidi commessi nella città di Detroit ogni 100,000 abitanti, poichè i corrispettivi valori delle stime ai minimi quadrati dei coefficienti di regressione sono dell'ordine di 10^{-3} . Tuttavia si commetterebbe un errore: infatti, osservando i corrispondenti risultati nel caso dei dati standardizzati, il loro ordine di grandezza è rispettivamente 10^0 , 10^{-1} , 10^{-2} . Di conseguenza, l'unica variabile predittrice da cui dipende meno la risposta è l'undicesima, ossia il tasso di morte per incidenti ogni 100,000 abitanti, ed il 10^0 e 10^{-1} di cui sopra saranno bilanciato dagli ordini di grandezza dei valori assunti dal quinto e settimo predittore, in accordo con quello del valore stimato della risposta. Col passaggio ai dati standardizzati, si ha una drastica diminuzione del numero di condizionamento delle matrici dei tre metodi di risoluzione: il numero di condizionamento di $\mathbf{Z}^T \mathbf{Z}$ è $3.88 \cdot 10^4$, mentre quello della matrice $\mathbf{Z}_{new}^T \mathbf{Z}_{new}$ è 7.08 se vale la condizione (3.1) e 9.05 se vale

il requisito (3.2), dove \mathbf{Z}_{new} è stata costruita, rispettivamente, con $m = 2$ ed $m = 3$ termini. L'errore relativo che si commette ricorrendo al criterio dell'equazione normale, piuttosto che alla fattorizzazione QR di \mathbf{Z} , per stimare i coefficienti di regressione lineare è di $5.34 \cdot 10^{-13}$. Inoltre le norme dei residui ottenuti con i tre metodi risolutivi assumono i valori, rispettivamente, di $8.16 \cdot 10^{-13}$, $2.96 \cdot 10^{-14}$, 5.63 e 4.57, gli ultimi due a seconda che sia stata considerata la condizione (3.1) o (3.2) per la strategia di riduzione.

Conclusioni

In questo elaborato finale è stato descritto il modello di regressione lineare multivariata, mostrandone i risultati qualitativi più rilevanti ed il legame con il problema di minimi quadrati.

Nella seconda parte, è stato presentato il problema ai minimi quadrati nel suo aspetto più generale, ponendo l'attenzione su risultati e metodi numerici per trovarne la soluzione, quali l'equazione normale, la fattorizzazione QR e la decomposizione in valori singolari della matrice del problema.

Nell'ultima parte, infine, è stato usando il software scientifico MATLAB per analizzare due famosi set di dati, 'LONGLLEY' e 'DETROIT', applicando differenti metodi numerici risolutivi e constatando che la scelta della fattorizzazione QR per la matrice dei predittori risulta essere la migliore per determinare un modello di regressione lineare che meglio approssimi i dati studiati.

Appendice A

Appendice

Proposizione A.0.1. *Sia $c \in \mathbb{R}$ e $k \in \mathbb{N}$. Siano $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{k,k}$. Allora:*

1. $tr(c \mathbf{A}) = c tr(\mathbf{A})$
2. $tr(\mathbf{A} \pm \mathbf{B}) = tr(\mathbf{A}) \pm tr(\mathbf{B})$
3. $tr(\mathbf{AB}) = tr(\mathbf{BA})$
4. $tr(\mathbf{B}^{-1} \mathbf{AB}) = tr(\mathbf{A})$
5. $tr(\mathbf{AA}^T) = \sum_{i=1}^k \sum_{j=1}^k a_{i,j}^2$

Proposizione A.0.2. *Siano $k \in \mathbb{N}$, $\mathbf{A} \in \mathbb{R}^{k,k}$ una matrice simmetrica, $\mathbf{x} \in \mathbb{R}^k$ un vettore e $\Lambda(\mathbf{A})$ l'insieme degli autovalori di \mathbf{A} . Valgono le seguenti identità:*

1. $\mathbf{x}^T \mathbf{A} \mathbf{x} = tr(\mathbf{x}^T \mathbf{A} \mathbf{x}) = tr(\mathbf{A} \mathbf{x} \mathbf{x}^T)$
2. $tr(\mathbf{A}) = \sum_{i=1}^k \lambda_i$ dove $\lambda_i \in \Lambda(\mathbf{A})$.

Proposizione A.0.3. *Siano $k \in \mathbb{N}$, $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{k,k}$ due matrici random. Siano \mathbf{A}, \mathbf{B} due matrici conformi a coefficienti reali costanti. Allora valgono:*

- $E(\mathbf{X} + \mathbf{Y}) = E(\mathbf{X}) + E(\mathbf{Y})$
- $E(\mathbf{AXB}) = \mathbf{AE}(\mathbf{X})\mathbf{B}$

Proposizione A.0.4. *Siano $p, q \in \mathbb{N}$, $\mathbf{X} \in \mathbb{R}^p$ un vettore random e $\mathbf{C} \in \mathbb{R}^{q,p}$ una matrice costante. Allora valgono:*

- $E(\mathbf{C}\mathbf{X}) = \mathbf{C}E(\mathbf{X})$
- $Cov(\mathbf{C}\mathbf{X}) = \mathbf{C}Cov(\mathbf{X})\mathbf{C}^T$.

Appendice B

Appendice

B.1 Listati dei programmi

B.1.1 Funzione MATLAB REGRESSLIN

```
function [condB,condB3,b1,b2,b3,y1,y2,y3,normr1,normr2,normr3] =REGRESSLIN(X,y)

% METODO 1) equazione normale
% METODO 2) fattorizzazione QR della matrice dei predittori Z
% METODO 3) costruzione di una nuova matrice dei predittori Znew

% INPUT: X è la matrice dei dati relativi ai predittori;
% y è la risposta.
% OUTPUT: condB è il numero di condizionamento della matrice B=Z'*Z;
% condB3 è il numero di condizionamento della matrice B3=Znew'*Znew;
% b1 è la stima ai minimi quadrati dei coefficienti di regressione
% col metodo 1;
% b2 è la stima ai minimi quadrati dei coefficienti di regressione
% col metodo 2;
% b3 è la stima ai minimi quadrati dei coefficienti di regressione
% col metodo 3.

[n,p]= size(X);
Z= zeros(n,p+1);
Z(:,1)= 1;
Z(:,2:p+1)= X;
B= Z'*Z;
condB= cond(B);
b1= B\ (Z'*y);
b2= Z\y;
```

```

S= X'*X;
[V,L]= eig(S);
[e,index]= sort(diag(L),'descend');
Vsort= V(:,index);
sc= cumsum(e);
s= sc/sc(end);
m=find(s>=0.8,1);
Znew= zeros(n,m+1);
Znew(:,1)= 1;
Xnew = X*Vsort(:,1:m);
Znew(:,2:m+1)= Xnew;
B3=Znew'*Znew;
condB3= cond(B3);
fprintf(' num.cond. originale %5.5d num. cond ridotto %5.5d con m=%1d termini\n',condB,condB3,m)
fprintf('\n')
b3= B3\Znew'*y;
fprintf(' Stime dei coeff. di regressione coi primi due metodi:\n')
disp([b1,b2])
fprintf(' Errore relativo tra i due coeff %d :\n', norm(b1-b2)/norm(b2))
fprintf('\n')
fprintf(' Stime dei coeff. di regressione col terzo metodo:\n')
disp(b3)
y1= Z*b1; % valore approssimato di y col metodo 1
y2= Z*b2; % valore approssimato di y col metodo 2
y3= Znew*b3; % valore approssimato di y col metodo 3
fprintf(' Stime delle risposte coi 3 metodi:\n')
disp([y1,y2,y3])
r1= y-y1; % residuo metodo 1
r2= y-y2; % residuo metodo 2
r3= y-y3; % residuo metodo 3
fprintf(' Residui coi 3 metodi:\n')
disp([r1,r2,r3])
fprintf('\n')
fprintf(' Norma dei residui coi 3 metodi:\n')
normr1= norm(r1); % norma residuo metodo 1
normr2= norm(r2); % norma residuo metodo 2
normr3= norm(r3); % norma residuo metodo 3
disp([normr1,normr2,normr3])
end

```

B.1.2 Funzione MATLAB regresslin095

```

function [condB4,b4,y4,normr4] =regresslin095(X,y)

% METODO 3) costruzione di una nuova matrice dei predittori Znew

% INPUT: X è la matrice dei dati relativi ai predittori; y è la risposta.

```

```

% OUTPUT: condB4 è il numero di condizionamento della matrice B4=Znew'*Znew;
% r4 è il residuo ottenuto con il metodo 3.

[n,p]= size(X);
S= X'*X;
[V,L]= eig(S);
[e,index]= sort(diag(L),'descend');
Vsort= V(:,index);
sc= cumsum(e);
s= sc/sc(end);
m= find(s>=0.95,1); % verifica della condizione >=0.95
Znew= zeros(n,m+1);
Znew(:,1)= 1;
Xnew = X*Vsort(:,1:m);
Znew(:,2:m+1)= Xnew;
B4= Znew'*Znew;
condB4= cond(B4);
fprintf(' numero cond ridotto %5.5d con m=%1d termini\n',condB4,m)
fprintf('\n')
b4= B4\Znew'*y); % stima ai minimi quadrati dei coefficienti di regressione col metodo 3
fprintf('\n')
fprintf(' Stime dei coeff. di regressione col terzo metodo:\n')
disp(b4)
y4= Znew*b4; % valore approssimato di y col metodo 3
fprintf(' Stime delle risposte col terzo metodo:\n')
disp(y4)
r4= y-y4; % residuo metodo 3
fprintf(' Residui col terzo metodo:\n')
disp(r4)
fprintf('\n')
fprintf(' Norma dei residui col terzo metodo:\n')
normr4= norm(r4); % norma residuo metodo 3
disp(normr4)
end

```

B.1.3 LONGLEY script

```

format short e
%%
%%% # The infamous Longley data, "An appraisal of least-squares programs from
% # the point of view of the user", JASA, 62(1967) p819-841.
% #
% # Variables are:
% # Number of people employed (usually the y variable)
% # GNP (Gross National Product) implicit price
% # deflator (The ratio of a country's nominal GNP to its real GNP, expressed as a percentage.)
% # GNP

```

```

% # Unemployed
% # Armed forces
% # Non-institutionalized population >=14 years of age
% # Year

X=[60323  83.0  234289  2356  1590  107608  1947
   61122  88.5  259426  2325  1456  108632  1948
   60171  88.2  258054  3682  1616  109773  1949
   61187  89.5  284599  3351  1650  110929  1950
   63221  96.2  328975  2099  3099  112075  1951
   63639  98.1  346999  1932  3594  113270  1952
   64989  99.0  365385  1870  3547  115094  1953
   63761 100.0  363112  3578  3350  116219  1954
   66019 101.2  397469  2904  3048  117388  1955
   67857 104.6  419180  2822  2857  118734  1956
   68169 108.4  442769  2936  2798  120445  1957
   66513 110.8  444546  4681  2637  121950  1958
   68655 112.6  482704  3813  2552  123366  1959
   69564 114.2  502601  3931  2514  125368  1960
   69331 115.7  518173  4806  2572  127852  1961
   70551 116.9  554894  4007  2827  130081  1962];

fprintf('\n')
fprintf('*****\n')
fprintf('***** CONDIZIONE SUGLI AUTOVALORI: >= 0.8 *****\n')
fprintf('*****\n')
fprintf('***** Analisi con variabili originali *****\n')
fprintf('*****\n')
y= X(:,1);
X(:,1)=[];
[condB,condB3,b1,b2,b3,y1,y2,y3,normr1,normr2,normr3] =REGRESSLIN(X,y);

fprintf('\n')
fprintf('*****\n')
fprintf('***** normalizzazione delle variabili *****\n')
fprintf('*****\n')
Xscaled=zscore(X); % permette di avere tutti i risultati con lo stesso ordine di grandezza
[condB,condB3,b1,b2,b3,y1,y2,y3,normr1,normr2,normr3] =REGRESSLIN(Xscaled,y);

fprintf('\n')
fprintf('*****\n')
fprintf('***** CONDIZIONE SUGLI AUTOVALORI: >= 0.95 *****\n')
fprintf('*****\n')
fprintf('***** Analisi con variabili originali *****\n')
fprintf('*****\n')
[condB4,b4,y4,normr4] =regresslin095(X,y);

```

```
fprintf('\n')
fprintf('*****\n')
fprintf('***** normalizzazione delle variabili *****\n')
fprintf('*****\n')
[condB4,b4,y4,normr4] =regresslin095(Xscaled,y);
```

B.1.4 DETROIT script

```
format short e
```

```
% The data are on the homicide rate in Detroit for the years 1961-1973.
% FTP - Full-time police per 100,000 population
% UEMP - % unemployed in the population
% MAN - number of manufacturing workers in thousands
% LIC - Number of handgun licences per 100,000 population
% GR - Number of handgun registrations per 100,000 population
% CLEAR - % homicides cleared by arrests
% WM - Number of white males in the population
% NMAN - Number of non-manufacturing workers in thousands
% GOV - Number of government workers in thousands
% HE - Average hourly earnings
% WE - Average weekly earnings
% HOM - Number of homicides per 100,000 of population (the y variable)
% ACC - Death rate in accidents per 100,000 population
% ASR - Number of assaults per 100,000 population

X0=[260.35 11.0 455.5 178.15 215.98 93.4 558724. 538.1 133.9 2.98 117.18 8.60 39.17 306.18
269.80 7.0 480.2 156.41 180.48 88.5 538584. 547.6 137.6 3.09 134.02 8.90 40.27 315.16
272.04 5.2 506.1 198.02 209.57 94.4 519171. 562.8 143.6 3.23 141.68 8.52 45.31 277.53
272.96 4.3 535.8 222.10 231.67 92.0 500457. 591.0 150.3 3.33 147.98 8.89 49.51 234.07
272.51 3.5 576.0 301.92 297.65 91.0 482418. 626.1 164.3 3.46 159.85 13.07 55.05 230.84
261.34 3.2 601.7 391.22 367.62 87.4 465029. 659.8 179.5 3.60 157.19 14.57 53.90 217.99
268.89 4.1 577.3 665.56 616.54 88.3 448267. 686.2 187.5 3.73 155.29 21.36 50.62 286.11
295.99 3.9 596.9 1131.21 1029.75 86.1 432109. 699.6 195.4 2.91 131.75 28.03 51.47 291.59
319.87 3.6 613.5 837.60 786.23 79.0 416533. 729.9 210.3 4.25 178.74 31.49 49.16 320.39
341.43 7.1 569.3 794.90 713.77 73.9 401518. 757.8 223.8 4.47 178.30 37.39 45.80 323.03
356.59 8.4 548.8 817.74 750.43 63.4 387046. 755.3 227.7 5.04 209.54 46.26 44.54 357.38
376.69 7.7 563.4 583.17 1027.38 62.5 373095. 787.0 230.9 5.47 240.05 47.24 41.03 422.07
390.19 6.3 609.3 709.59 666.50 58.9 359647. 819.8 230.2 5.76 258.05 52.33 44.17 473.01];

X=X0;
y= X(:,12);
X(:,12)= [];
x7=X(:,7);
[n,1]=size(X);
diplin7=zeros(1,1);
for k=1:l
```

```

    diplin7(k)=cond([X(:,k),x7]);
end
diplin7(7)=[];
% diplin7 conterrà informazioni sulla dipendenza lineare di WM dagli altri predittori
X(:,7)=[];
fprintf('\n')
fprintf('*****\n')
fprintf('**studio dipendenza lineare colonna 7 dalle altre colonne**\n')
fprintf('*****\n')
disp(diplin7)
fprintf('\n')
fprintf('*****\n')
fprintf('***** CONDIZIONE SUGLI AUTOVALORI: >= 0.8 *****\n')
fprintf('*****\n')
fprintf('***** Analisi con variabili originali *****\n')
fprintf('*****\n')
[condB,condB3,b1,b2,b3,y1,y2,y3,normr1,normr2,normr3] =REGRESSLIN(X,y);
fprintf('\n')
fprintf('*****\n')
fprintf('***** normalizzazione delle variabili *****\n')
fprintf('*****\n')
Xscaled=zscore(X); % ora i risultati hanno lo stesso ordine di grandezza
[condB,condB3,b1,b2,b3,y1,y2,y3,normr1,normr2,normr3] =REGRESSLIN(Xscaled,y);

```

Bibliografia

- [1] Richard A. Johnson, Dean W. Wichern, '*Applied Multivariate Statistical Analysis*', Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632, 1982
- [2] Davide Palitta, Valeria Simoncini, '*Dispense del corso di Calcolo Numerico - Modulo di Algebra Lineare Numerica - A.A. 2014/2015*', II edizione, v.1.1, 12 Febbraio 2016

Ringraziamenti

Colgo l'occasione per ringraziare la Prof.ssa Valeria Simoncini, per la disponibilità e l'affabilità dimostratemi durante la stesura di questo elaborato. Ringrazio la mia famiglia, per avermi permesso di intraprendere questo percorso ed avermi sempre supportato.

Ed infine, ma non per importanza, ringrazio tutti i miei amici, di vecchia data e non, per avermi accompagnato in questo viaggio e condiviso quotidianamente con me soddisfazioni e delusioni, gioie e dolori, grandi risate e qualche lacrima..... Ad maiora!