Statistical Learning

Prova d'esame

24 Giugno 2022

Tempo a disposizione: 180 minuti

Problema 1

Si risponda alle seguenti domande.

a) Si consideri un modello di regressione con matrice del disegno X e vettore di risposta y dati da

$$X = \begin{bmatrix} 10^{-9} & -1 \\ -1 & 10^{-5} \end{bmatrix}, \quad y = c(3 \times 10^9, -2.99997)$$

Si calcoli la stima dei minimi quadrati $\hat{\beta}$. Si riporti il primo elemento di $\hat{\beta}$, arrotondando al primo decimale.

```
hat_beta = solve(X,y)
round(hat_beta,1)[1]
```

```
## [1] 3
```

b) Siano $X_1, \ldots, X_n$ i.i.d. ad una variabile X con distribuzione $N(\mu, 1)$. Si consideri lo stimatore media campionaria $\bar{X}$. Quanto vale il rapporto tra errore di previsione atteso ed errore di stima atteso, ovvero

$$\frac{\mathrm{E}[(\bar{X} - X)^2]}{\mathrm{E}[(\bar{X} - \mu)^2]}$$

per $n = 5$? Riportare il risultato arrotondando al primo decimale. Potete rispondere alla domanda in modo analitico oppure utilizzando una simulazione con un gran numero di ricampionamenti (e.g. 10^5).

```
(1/n + 1)/(1/n)
```

```
## [1] 6
```

```
smallsim <- function(n){
  mu <- 1
  x <- rnorm(n, mean=mu)
  x_new <- rnorm(1, mean=mu)
  x_bar <- mean(x)
  ep <- (x_bar - x_new)^2
  es <- (x_bar - mu)^2
  return(c(ep,es))
}
set.seed(123)
B <- 10^5
res <- replicate(B,smallsim(n=n))
ratio = mean( res[1,] ) / mean( res[2,] )
round(ratio,1)
```

```
## [1] 6
```

c) Sia

$$x = (0.24, 1.61, 4.61, 3.95)^t$$

la realizzazione di una variabile casuale X con distribuzione Gaussiana p-variata $N_p(\mu, \Sigma)$ con vettore delle medie μ incognito e matrice di varianze/covarianze

$$\Sigma = \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

con $\rho = -1/4$. La stima di massima verosimiglianza per μ è data da $\hat{\mu}_{\mathrm{MLE}} = x$. Calcolare la stima secondo James-Stein $\hat{\mu}_{\mathrm{JS}}$ (con $c = 1$). Riportare il valore del primo elemento di $\hat{\mu}_{\mathrm{JS}}$, arrotondando il risultato al quarto decimale.

```
p = 4
mu = 1:p
rho = -1/4
R = matrix(rep(rho,p*p),ncol=p) + diag(rep(1-rho,p))
R_inv = solve(R)
c = 1
lambda_max = max(eigen(R)$values)
p_tilde = sum(diag(R)) / lambda_max
x = matrix(x,nrow=p)
k = as.numeric(1 - c*(p_tilde-2)/(t(x) %*% R_inv %*% x))
mu_js = k * x
round(mu_js,4)[1]
```

```
## [1] 0.2376
```

Problema 2

Si risponda alle seguenti domande.

a) Si consideri il modello lineare semplice con intercetta e una covariata x_i. I dati sono

$$\{(y_i, x_i)_{i=1}^4\} = \{(1.4, 0), (1.4, -2), (0.8, 0), (0.4, 2)\}$$

Riportare il valore di λ che corrisponde alla stima ridge $\hat{\beta}_\lambda = (1, -1/24)^t$ senza penalizzare l'intercetta.

```
# a.
x = c(0,-2,0,2)
y = c(1.4,1.4,0.8,0.4)
beta = -1/m
xtx = crossprod(x)
xty = crossprod(x,y)
l = c(xty/beta - xtx)
a = round(l,2)
a
```

```
## [1] 40
```

b) Si supponga di avere una matrice del disegno X tale che $X^t X = I_p$. Sia $X^t y = (-0.6, -0.2, 1.6, 0.1)^t$. Calcolare la stima $\tilde{\beta}_\lambda$ dello stimatore *best subset selection*

$$\tilde{\beta}_\lambda = \min_\beta \frac{1}{2}\|y - X\beta\|^2 + \lambda\|\beta\|_0$$

con il valore di $\lambda = 0.1$. Riportare il valore del primo elemento di $\tilde{\beta}_\lambda$.

```
lambda = 0.1
beta_tilde = Xty * (abs(Xty) > sqrt(2*lambda))
beta_tilde[1]
```

```
## [1] -0.6
```

c) Con i dati del punto precedente, calcolare la stima James-Stein $\hat{\beta}_{\mathrm{JS}}$ per β, ipotizzando di conoscere il vero valore di $\sigma = 1$. Riportare il valore del primo elemento di $\hat{\beta}_{\mathrm{JS}}$.

```
p = 4
sigma = 1
beta_ls = matrix(Xty,ncol=1)
beta_js = c(1 - ((p-2)*sigma^2)/(t(beta_ls)%*% beta_ls)) * beta_ls
beta_js[1]
```

```
## [1] -0.1959596
```

Problema 3

Si consegni il file .R che produce le risposte alle domande richieste. Il codice deve essere **riproducibile** e, se eseguito, deve stampare in output **solo** il risultati richiesti dalle domande a) e b).

Si consideri il dataset `longley` presente nella libreria `datasets`. La variabile risposta è `Employed`, i predittori sono `GNP.deflator`, `GNP`, `Unemployed`, `Armed.Forces`, `Population` e `Year`.

Si consideri come *training set* le prime 15 osservazioni (righe) del dataset `longley`, e come *test point* la sedicesima osservazione (anno 1962).

Si utilizzi l'algoritmo *split conformal* considerando come *Learning set* le osservazioni del training set con indici pari, i.e. $L = \{2, 4, 6, 8, 10, 12, 14\}$ e come *Inference set* le osservazioni del training set con indici dispari, i.e. $I = \{1, 3, 5, 7, 9, 11, 13, 15\}$.

Si costruisca l'intervallo di previsione a livello $1 - \alpha$ con $\alpha = 1/3$ per il test point con la regressione *ridge* per $\lambda = 1$, utilizzando la funzione `lm.ridge` presente nella libreria `MASS`.

Riportare

a) gli estremi dell'intervallo di previsione;
b) il valore `TRUE` se il test point si trova all'interno dell'intervallo di previsione, `FALSE` altrimenti.

```
train = longley[-16,]
test = longley[16,]

x = as.matrix(train[,-7])
y = as.numeric(train[,7])
n = nrow(x)
p = ncol(x)
x_new = as.matrix(test[,-7],ncol=p)
y_true = as.numeric(test[,7])

L = c(2,4,6,8,10,12,14)
I = c(1,3,5,7,9,11,13,15)
m = length(I)
alpha = 1/3

library(MASS)
fit_L <- lm.ridge(Employed ~ ., train, subset=L, lambda = 1)
```

```
beta_L = matrix(coef(fit_L), ncol=1)
y_hat = cbind(1,x[I,,drop=F]) %*% beta_L
res =  abs(y[I] - y_hat)
o = order(res)
c = ceiling((1-alpha)*(m+1))
r = res[o][c]

y_new = c(1,x_new) %*% beta_L
lo = y_new - r
up = y_new + r
c(lo,up)
```

```
## [1] 70.76016 71.65629
```

```
y_true
```

```
## [1] 70.551
```

```
(lo <= y_true & y_true <= up)
```

```
##       [,1]
## [1,] FALSE
```

Problema 4

a) Sia X una variabile causale con distribuzione $N_p(\mu, \Sigma)$. La matrice Σ ha elementi diagonali pari a 1 ed elementi fuori dalla diagonale pari a $r > 0$. In questo caso l'autovalore più grande di Σ è pari a

$$\lambda_1 = 1 + (p-1)r$$

Per quale valori di p è preferibile utilizzare lo stimatore di James-Stein come alternativa allo stimatore di massima verosimiglianza? Cosa succede se $r = 0.5$?

b) Per gli intervalli di previsione, qual è la relazione tra *conditional coverage* e *marginal coverage*?

c) Si consideri un insieme di n punti (x_i, y_i), $i = 1, \ldots, n$. Cosa succede se stimiamo una *natural cubic spline* con un n nodi in corrispondenza dei valori $x_1, \ldots, x_n$?