Homework2

December 22, 2024

- La consegna deve includere: i) il codice commentato (File1) e ii) un file PDF con i commenti agli output (File2).
- I nomi dei file devono essere nel formato “Homework2_Matricola_Cognome_NomeFile”.
- L’uso di chat-bot o intelligenza artificiale è ammesso solo come strumento per il controllo del testo, dei calcoli e del codice, ma non per la loro scrittura. L’uso di chat-bot o IA va dichiarato.
- È possibile lavorare in gruppo, ma il codice e la teoria vanno scritte autonomamente. E’ comunque necessario dichiarare con chi si è collaborato nel “File2”
- Non verrà assegnato un voto per l’homework, ma in caso di elaborati particolarmente meritevoli, ne verrà tenuto conto per la valutazione finale dell’esame.

File1

- deve essere un file sorgente del linguaggio di programmazione usato, eseguibile nella sua console. Quindi, nel caso di R, serve un file con estensione .R, se avete usato Julia serve un file con estensione jl, se avete usato python serve un file con estensione “.py”. Il file non può essere un notebook (quindi estensione tipo .ipynb o .md non vanno bene)
- va impostato il seed a inizio codice (la vostra matricola), in modo tale che eseguendo il codice, si ottengano gli stessi risultati presenti nel File2
- I commenti nel codice devono spiegare sinteticamente cosa fanno le diverse parti del codice. Potete fare riferimento a formule presenti nel File2, purché siano numerate in modo appropriato.

File2

Deve essere un pdf contenente

- la teoria utilizzata per risolvere gli esercizi
- i calcoli svolti, con spiegazione dei passaggi usati per arrivare al risultato (il solo risultato finale, senza spiegazioni, non viene accettato)
- i risultati numerici e/o grafici che rispondono alle domande degli esercizi.

Scadenza

La scadenza ultima è il 19 Gennaio per chi vuole fare l’esame il 22 Gennaio, e il 31 Gennaio per tutti gli altri. Potete risottomettere l’elaborato quante volte volete, ma le scadenza garantite sono

- Sottomissione entro 24 Dicembre -> Correzione entro il 30 Gennaio
- Sottomissione entro 3 Gennaio -> Correzione entro il 7 Gennaio
- Sottomissione entro 10 Gennaio -> Correzione entro il 14 Gennaio
- Sottomissione entro 20 Gennaio -> Correzione entro il 26 Gennaio

Informazioni Pratiche

- Per le stime a posteriori sono necessari almeno 1000 campioni a convergenza
- i campioni non devono per forza essere indipendenti, ma dovete almeno verificare e commentare la dipendenza
- se implementate un passo Metropolis, la varianza della proposta deve essere adattiva.
- Non potete usare software come STAN, Nimble, JAGS, etc, ma gli algoritmi MCMC vanno scritti

NOTA BENE: Potete o

- fare l'esercizio 1 per intero, o
- fare i primi 3 punti dell'esercizio 1, e l'esercizio 2.
- New Considerando la complessità nell'implementazione del modello richiesto per l'esercizio 1 (punto 3), se riscontrate difficoltà a farlo funzionare correttamente, potete comunque includere i risultati ottenuti e fornire una descrizione dettagliata dell'algoritmo riportato nel File2. Se gli errori presenti nel codice sono esclusivamente typo o di natura sintattica, e non teorica, il punto potrà essere considerato completato. Tuttavia, in questo caso, dato che non sarà possibile svolgere i punti 4 e 5, sarà necessario procedere con l'esercizio 2

Se fate tutti i punti dell'esercizio 1 e l'esercizio 2, ne terrò conto durante l'esame.

Alcuni Consigli

- Quando calcolate il rapporto presente nell'α del Metropolis, è consigliabile scriverlo come l'esponenziale della differenza dei logaritmi. Questo approccio permette di evitare problemi di cancellazione numerica. Prestate attenzione a calcolare direttamente la densità di in scala logaritmica, senza passare per il calcolo esplicito della densità e il successivo logaritmo (quindi, evitando di calcolare l'esponenziale, ma lavorando esclusivamente con il suo argomento). Per il determinante, potete utilizzare la funzione **determinant(Matrice_Covarianze, logarithm=T)$modulus**, che restituisce il determinante direttamente in scala logaritmica, riducendo il rischio di cancellazione numerica.
- Nella parte adattiva della varianza, come coefficiente moltiplicativo potete utilizzare $1/\text{iterazione}^\rho$, con $\rho \in [0.1, 0.5]$, oppure $A/(2A + \text{iterazione})$, con $A \in [100, 1000]$.
- Per campionare $\phi \in [a, b]$, potete procedere in uno dei seguenti modi:
  1. Effettuare una trasformazione del tipo
  $$\eta = \log\left(\frac{\phi - a}{b - \phi}\right)$$
  e lavorare proponendo η da una normale. In questo caso, tuttavia, è necessario trovare e utilizzare la prior di η indotta da quella di ϕ.
  2. Proporre direttamente ϕ con una normale, e rifiutare ogni volta che si esce dal dominio, i.e. $\phi \notin [a, b]$.
- Vi ricordo che vi avevo mostrato a lezione che, con una covarianza esponenziale, conviene assumere $\phi \in [3/\text{max dist}, 3/\text{min dist}]$.
- La matrice delle varianze e covarianze è sempre definita positiva nell'intervallo dei valori di $\phi \in [3/\text{max dist}, 3/\text{min dist}]$. Non è quindi necessario aggiungere nulla alla diagonale, né effettuare inversioni generalizzate. Se, per qualche motivo, la matrice risultasse non definita positiva, significa che c'è un errore nel codice.

1 Esercizio 1

Assumete che

$$\mathbf{s} = (s_1, s_2) \in \mathbb{R}^2$$

e assumete di avere il seguente modello

$$Y(\mathbf{s})|W(\mathbf{s}) \overset{iid}{\sim} GP(W(\mathbf{s}), \tau^2)$$

$$W(\mathbf{s}) \sim GP(m(\mathbf{s}), C(||\mathbf{s} - \mathbf{s}'||; \phi, \sigma^2))$$

con

$$m(\mathbf{s}) = \beta_0 + \beta_1 s_1$$

e

$$C(||\mathbf{s} - \mathbf{s}'||; \phi, \sigma^2) = \sigma^2 \exp(-\phi||\mathbf{s} - \mathbf{s}'||)$$

I parametri sono

$$\tau^2 = 0.5, \quad \sigma^2 = 5, \quad \phi = \frac{3}{10} \quad \beta_0 = 2 \quad \beta_1 = 0.1$$

1- Simulate dal modello 100 osservazioni $\mathbf{y} = (y_1, \dots, y_{100})$ nei punti $\mathbf{D} = (\mathbf{s}_1, \dots, \mathbf{s}_{100})$, dove

$$\mathbf{s}_i \sim U(0, 10) \times U(0, 10)$$

(le coordinate sono simulate uniformemente in $[0, 10]^2$)

2- fate uno scatterplot delle coordinate, con i punti che possono assumere 4 colori. Il primo gruppo (primo colore) deve contenere le osservazioni $< 25\%$ quantile di $\mathbf{y}$, il secondo gruppo (secondo colore) quelle $\in (25\%, 50\%]$, il terzo quelle $\in (50\%, 75\%]$, l'ultimo contiene le rimanenti.

3- scegliete un n qualsiasi $\in [10, 90]$. Tra le 100 osservazioni simulate al punto uno, prendetene in maniera casuale n tra le 100 e indicate il loro vettore con $\mathbf{y}_o$ e le restanti con $\mathbf{y}_u$. Il set delle coordinate delle osservazioni $\mathbf{y}_o$ è indicato con D_o, mentre quello di $\mathbf{y}_u$ con D_u. Definite delle prior appropriate, giustificando la scelta, e scrivete un MCMC per ottenere campioni da

$$f(\mathbf{w}, \beta_0, \beta_1, \tau^2, \sigma^2, \phi|\mathbf{y}_o)$$

Nuova Opzione Alternativamente, potete anche ottenere campioni da

$$f(\beta_0, \beta_1, \tau^2, \sigma^2, \phi|\mathbf{y}_o, \mathbf{w}_o)$$

dove $\mathbf{w}_o$ sono i valori di $\mathbf{w}$ associati ai valori $\mathbf{y}_o$

4- Ottenete campioni dalla a posteriori di

$$\epsilon(\mathbf{s}) = Y(\mathbf{s}) - (\beta_0 + \beta_1 s_1)$$

con $\mathbf{s} \in D_o$ e mostrate la distribuzione spaziale, in maniera simile al punto 2, della media a posteriori.

5- Ottenete campioni dalla a posteriori

$$f(y(\mathbf{s})|\mathbf{y}_o)$$

con $\mathbf{s} \in D_u^* \subseteq D_u$, con D_u^* compreso di almeno 10 elementi, e valutate nella maniera che ritenete migliore, se il valor vero è all'interno dell'intervallo di credibilità del 95%.

2 Esercizio 2

Abbiamo un modello mistura

$$Y_i|z_i \sim P(\lambda_{z_i})$$

con

$$P(z_i = k) = \pi_k$$

$n = 200$, $z_i \in \{1, 2, 3\}$, $\pi_k = 1/3$, k = 1,2,3, e

$$\lambda_1 = 1, \quad \lambda_2 = 10 \quad \lambda_3 = 25$$

1- Simulate da modello e rappresentate graficametne, nel modo che ritenete più appropriato, i dati. Dai plot deve essere chiara la differenza tra le osservazioni appartenenti ai diversi gruppi 2- Assumete delle prior appropriate e implementate un MCMC per ottenere stime a posteriori di

$$f(\lambda_1, \lambda_2, \lambda_3, \pi_1, \pi_2, \pi_3, z_1, \dots, z_n|\mathbf{y})$$

Come prior per λ vi suggerisco delle Gamma.