

Statistica e Modelli Stocastici (Modulo II)

Prova intermedia n°0

March 19, 2017

Per lo svolgimento della prova seguire le seguenti istruzioni:

1. Aprire in Matlab lo script `genera_dati.m` ed inserire matricola, nome e cognome dove richiesto.
2. Eseguire in Matlab lo script `genera_dati.m`. Lo script produrrà nella workspace la variabile `data` necessaria per lo svolgimento della prova. Se la variabile dovesse corrompersi eseguire nuovamente lo script tenendo presente che il contenuto della workspace sarà prima eliminato.
3. Aprire in Matlab lo script `soluzione.m` e produrre il codice necessario per generare l'output richiesto dal testo degli esercizi. Seguire le istruzioni riportate nello script.
4. Completato lo script `soluzione.m`, eseguirlo in Matlab al fine di produrre il file di output `SMS2_prova0_[matricola].mat`
5. Caricare in ILIAS il file `.mat` generato e lo script `soluzione.m`.

1 Esercizio 1

La matrice `data` contiene come colonne le variabili velocità del vento (m/s), temperatura ($^{\circ}C$) e concentrazione di PM10 ($\mu g/m^3$) rilevate in giorni consecutivi da una stazione di rilevamento. Si assuma che le tre variabili possano essere congiuntamente modellizzate tramite una variabile casuale Normale multivariata.

Domanda 1 Tramite i dati osservati stimare il vettore delle medie e la matrice di varianza-covarianza della Normale multivariata.

Domanda 2 Dire per quale coppia di variabili la correlazione lineare è più forte.

Domanda 3 Condizionatamente ai parametri della Normale multivariata stimati nella Domanda 1, prevedere la concentrazione di PM10 tramite la sola variabile velocità del vento e poi tramite entrambe le variabili velocità del vento e temperatura. Calcolare il root mean squared error di previsione confrontando la previsione con i dati di concentrazione PM10 nella matrice **data**. Specificare se per la previsione è meglio sfruttare la sola informazione sulla velocità del vento oppure anche l'informazione sulla temperatura.

Nota Se y_i è una quantità osservata e \hat{y}_i è la quantità prevista, allora il root mean squared error è definito come $\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$.

Domanda 4 Dire come cambierebbe la matrice di varianza-covarianza se la velocità del vento fosse espressa in km/h .

2 Esercizio 2

Si supponga che velocità del vento e temperatura siano osservate senza errore e che la concentrazione di PM10 sia la variabile dipendente.

Domanda 1 Stimare i parametri del modello di regressione lineare multipla (con costante) che spiega la concentrazione di PM10 in funzione delle altre variabili.

Domanda 2 Stimare la varianza della componente di errore ε .

Domanda 3 Calcolare l'intervallo di confidenza sulla previsione in caso di velocità del vento pari a 10 m/s e temperatura 20 $^{\circ}C$.

Domanda 4 La varianza di ε può essere più grande della varianza di uno qualsiasi dei regressori?