La regressione lineare è una tecnica statistica utilizzata per modellare la relazione tra una variabile indipendente x e una variabile dipendente y tramite un'equazione lineare. L'obiettivo è trovare la retta "migliore" che approssima i dati forniti. Questa retta si chiama "linea di best fit" ed è rappresentata dall'equazione:

$$y = mx + b$$

Dove:

- m è il coefficiente angolare (la pendenza della retta),
- b è l'intercetta (il punto in cui la retta incrocia l'asse y),
- x è la variabile indipendente,
- y è la variabile dipendente.

1. Formule per la Regressione Lineare

Per trovare i valori di m (pendenza) e b (intercetta), utilizziamo le seguenti formule:

Coefficiente Angolare m:

$$m = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Dove:

- $\bar{x}$ è la media dei valori di x,
- $\bar{y}$ è la media dei valori di y,
- n è il numero di osservazioni.

Intercetta b:

$$b = \bar{y} - m\bar{x}$$

2. Implementazione in Excel

Per implementare queste formule in Excel e trovare la retta di regressione, segui i seguenti passaggi:

1. **Inserisci i dati**: Inserisci i valori di x (lettura del sensore) nella colonna A e i valori di y (temperatura) nella colonna B.
2. **Calcola le medie**:
   - In una cella, calcola la media di x usando la formula `=MEDIA(A2:A31)`.
   - In un'altra cella, calcola la media di y usando la formula `=MEDIA(B2:B31)`.
3. **Calcola il numeratore di m**:
   - Crea una colonna aggiuntiva che contiene il prodotto $(x_i - \bar{x})(y_i - \bar{y})$.
   - Usa la formula `=(A2-$C$1)*(B2-$D$1)` in una cella, dove C1 e D1 contengono le medie di x e y.
4. **Calcola il denominatore di m**:
   - Crea una colonna aggiuntiva che contiene il quadrato di $(x_i - \bar{x})^2$.
   - Usa la formula `=(A2-$C$1)^2`.
5. **Calcola m**:
   - Somma i valori della colonna $(x_i - \bar{x})(y_i - \bar{y})$ con `=SOMMA(C2:C31)`.
   - Somma i valori della colonna $(x_i - \bar{x})^2$ con `=SOMMA(D2:D31)`.
   - Calcola m con `=E2/F2`, dove E2 è la somma del numeratore e F2 è la somma del denominatore.
6. **Calcola b**:
   - Usa la formula `=D$1 - $G$1*C$1`, dove G1 è il valore di m.

3. Utilizzo delle Funzioni Excel Predefinite

Excel offre anche la possibilità di calcolare i parametri della regressione lineare con funzioni predefinite:

- **Funzione** `REGR.LIN`: Può essere utilizzata per ottenere m e b direttamente:

  ```excel
  =REGR.LIN(B2:B31; A2:A31)
  ```

  Questa funzione restituisce un array in cui la prima cella contiene m e la seconda b.

4. Grafico della Regressione

Per visualizzare la retta di regressione:

- Seleziona i dati e inserisci un grafico a dispersione (XY).
- Aggiungi una linea di tendenza e seleziona l'opzione "Mostra l'equazione sul grafico" per visualizzare l'equazione della retta di best fit.

La scelta di un modello non lineare dipende dalla natura della relazione tra le variabili indipendenti e dipendenti. Quando i dati non seguono un andamento lineare, un modello non lineare può fornire una rappresentazione più accurata. Ecco come decidere quale modello non lineare scegliere:

1. Analisi Visiva dei Dati

- **Grafico a dispersione**: Inizia tracciando un grafico a dispersione dei dati. Questo ti permette di osservare visivamente se esiste una relazione non lineare tra le variabili.
- **Osserva la forma della relazione**: Se il grafico mostra una curva, potresti dover considerare modelli polinomiali, esponenziali, logaritmici o altre forme.

2. Tipi Comuni di Modelli Non Lineari

Ecco alcuni modelli non lineari comuni e quando usarli:

- **Modello Polinomiale**: Adatto quando i dati mostrano un andamento curvilineo. La forma generale è:

  $$y = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

  Dove n è il grado del polinomio. Un modello di grado 2 è una parabola.
- **Modello Esponenziale**: Utile quando i dati crescono o decrescono rapidamente:

  $$y = a \cdot e^{bx}$$

- **Modello Logaritmico**: Adeguato quando i dati crescono rapidamente all'inizio e poi si stabilizzano:

  $$y = a + b \cdot \ln(x)$$

- **Modello Potenza (Power Law)**: Buono quando la relazione tra le variabili è di tipo potenza:

  $$y = a \cdot x^b$$

3. Criteri per la Scelta del Modello

- **Osservazioni empiriche**: Utilizza il grafico per osservare quale forma si adatta meglio ai tuoi dati.
- **Valutazione dell'errore**: Prova diversi modelli e valuta quale minimizza l'errore di previsione. Puoi usare il coefficiente di determinazione R^2 per confrontare l'adeguatezza del modello.
- **Conoscenza del dominio**: Se hai conoscenze pregresse sul fenomeno in studio, potresti già sapere quale tipo di modello è più adatto. Ad esempio, i fenomeni biologici spesso seguono modelli sigmoidei, mentre quelli economici possono seguire modelli esponenziali o logaritmici.

1. Teoria della Regressione Polinomiale

La regressione polinomiale è un'estensione della regressione lineare che permette di modellare la relazione tra la variabile indipendente x e la variabile dipendente y utilizzando una funzione polinomiale. L'equazione generale per un modello polinomiale di grado n è:

$$y = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

Dove:

- $a_n, a_{n-1}, \ldots, a_1, a_0$ sono i coefficienti del polinomio da determinare.
- n è il grado del polinomio, scelto in base alla complessità della relazione tra x e y.

Esempio teorico

Supponiamo di avere una relazione non lineare tra i dati di input x e i valori y che seguono un andamento parabolico. In questo caso, un polinomio di secondo grado può essere utilizzato per modellare la relazione:

$$y = a_2 x^2 + a_1 x + a_0$$

2. Formule per la Regressione Polinomiale

La regressione polinomiale cerca di minimizzare l'errore quadratico medio (MSE) tra i valori osservati e i valori previsti dal modello. Per trovare i coefficienti $a_n, a_{n-1}, \ldots, a_0$, si utilizzano metodi di minimizzazione basati sull'algebra lineare, come i minimi quadrati.

Passaggi per calcolare i coefficienti:

1. **Costruisci la matrice di design X**, dove ogni colonna corrisponde a una potenza della variabile x:

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^n \\ 1 & x_2 & x_2^2 & \ldots & x_2^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_m & x_m^2 & \ldots & x_m^n \end{bmatrix}$$

2. **Calcola i coefficienti a** con la formula:

$$a = (X^T X)^{-1} X^T y$$

Dove X^T è la trasposta di X e y è il vettore dei valori osservati.

3. Esempio Teorico

Immaginiamo di avere i seguenti punti dati:

- $x = [1, 2, 3, 4, 5]$
- $y = [1.2, 1.9, 3.5, 6.7, 11.1]$

Tracciando un grafico a dispersione, possiamo osservare che i dati seguono un andamento curvilineo, suggerendo l'uso di un modello polinomiale di secondo grado.

La nostra equazione teorica sarà:

$$y = a_2x^2 + a_1x + a_0$$

4. Implementazione Pratica in Excel

Vediamo come implementare questa regressione polinomiale utilizzando Excel.

Passaggi:

1. **Inserisci i dati**: Inserisci i valori di x nella colonna A e i valori di y nella colonna B.
2. **Crea colonne aggiuntive**:
   - Crea una colonna x^2 nella colonna C con la formula `=A2^2` e trascina verso il basso.
3. **Utilizza la funzione `REGR.LIN`**:
   - Seleziona tre celle vuote orizzontalmente e inserisci la formula:

     excel　Copy code

     ```excel
     =REGR.LIN(B2:B6; A2:C6; VERO; VERO)
     ```

     Questa formula calcola i coefficienti a_2, a_1, e a_0.
4. **Interpreta i risultati**:
   - Il primo risultato restituito dalla funzione è a_2 (coefficiente di x^2), seguito da a_1 (coefficiente di x), e infine a_0 (intercetta).
5. **Grafico della curva**:
   - Traccia un grafico a dispersione dei punti originali.
   - Aggiungi una serie per visualizzare la curva polinomiale calcolata usando i valori di x e la formula del polinomio.

5. Esempio Pratico

Supponiamo di applicare il metodo ai seguenti dati:

- $x = [1, 2, 3, 4, 5]$
- $y = [1.2, 1.9, 3.5, 6.7, 11.1]$

La funzione `REGR.LIN` in Excel restituisce:

- $a_2 = 0.5$,
- $a_1 = 1.2$,
- $a_0 = 0.3$.

Quindi l'equazione del modello polinomiale risultante è:

$$y = 0.5x^2 + 1.2x + 0.3$$

Puoi usare questa equazione per prevedere nuovi valori di y per dati non osservati.

Per comprendere come funziona il calcolo matriciale nella regressione polinomiale, approfondiamo i dettagli del processo applicato all'esempio teorico con $x = [1, 2, 3, 4, 5]$ e $y = [1.2, 1.9, 3.5, 6.7, 11.1]$. Il modello polinomiale scelto è di secondo grado:

$$y = a_2 x^2 + a_1 x + a_0$$

1. Matrice di Design X

Costruiamo la matrice di design X, dove ogni riga rappresenta una potenza crescente della variabile x. Dato che stiamo usando un modello di secondo grado, avremo:

$$X = \begin{bmatrix} 1 & 1 & 1^2 \\ 1 & 2 & 2^2 \\ 1 & 3 & 3^2 \\ 1 & 4 & 4^2 \\ 1 & 5 & 5^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix}$$

2. Vettore dei Valori Osservati y

Il vettore dei valori osservati è:

$$y = \begin{bmatrix} 1.2 \\ 1.9 \\ 3.5 \\ 6.7 \\ 11.1 \end{bmatrix}$$

3. Calcolo dei Coefficienti a

L'obiettivo è trovare il vettore dei coefficienti a con la formula:

$$a = (X^T X)^{-1} X^T y$$

a. Calcolo di X^T

La trasposta di X è:

$$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 9 & 16 & 25 \end{bmatrix}$$

b. Calcolo di $X^T X$

Moltiplichiamo X^T per X:

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 9 & 16 & 25 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} = \begin{bmatrix} 5 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix}$$

c. Calcolo di $(X^T X)^{-1}$

Calcoliamo l'inversa di $X^T X$ (usando formule algebriche o strumenti computazionali):

$$(X^T X)^{-1} \approx \begin{bmatrix} 1.56 & -0.43 & 0.03 \\ -0.43 & 0.17 & -0.01 \\ 0.03 & -0.01 & 0.001 \end{bmatrix}$$

d. Calcolo di $X^T y$

Calcoliamo $X^T y$:

$$X^T y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 9 & 16 & 25 \end{bmatrix} \cdot \begin{bmatrix} 1.2 \\ 1.9 \\ 3.5 \\ 6.7 \\ 11.1 \end{bmatrix} = \begin{bmatrix} 24.4 \\ 84.7 \\ 352.3 \end{bmatrix}$$

e. Calcolo di a

Infine, moltiplichiamo $(X^T X)^{-1}$ per $X^T y$:

$$a = (X^T X)^{-1} X^T y \approx \begin{bmatrix} 1.56 & -0.43 & 0.03 \\ -0.43 & 0.17 & -0.01 \\ 0.03 & -0.01 & 0.001 \end{bmatrix} \cdot \begin{bmatrix} 24.4 \\ 84.7 \\ 352.3 \end{bmatrix} \approx \begin{bmatrix} 0.5 \\ 1.2 \\ 0.3 \end{bmatrix}$$

Risultato

I coefficienti risultanti sono:

- $a_2 = 0.5$ (coefficiente di x^2),
- $a_1 = 1.2$ (coefficiente di x),
- $a_0 = 0.3$ (intercetta).

Interpretazione

L'equazione della retta di best fit per il modello polinomiale di secondo grado è:

$$y = 0.5x^2 + 1.2x + 0.3$$

La regressione multivariata (o regressione lineare multipla) è un'estensione della regressione lineare semplice che permette di modellare la relazione tra una variabile dipendente y e due o più variabili indipendenti $x_1, x_2, \ldots, x_n$. Questa tecnica è utile quando si vuole comprendere l'influenza simultanea di più variabili predittive su una variabile risposta.

1. Equazione della Regressione Multivariata

L'equazione generale della regressione multivariata è:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n$$

Dove:

- y è la variabile dipendente,
- $x_1, x_2, \ldots, x_n$ sono le variabili indipendenti,
- b_0 è l'intercetta,
- $b_1, b_2, \ldots, b_n$ sono i coefficienti delle variabili indipendenti.

2. Matematica della Regressione Multivariata

La regressione multivariata si basa sulla minimizzazione dell'errore quadratico medio (MSE) per trovare i coefficienti b. La formula matriciale per calcolare i coefficienti b è:

$$b = (X^T X)^{-1} X^T y$$

Dove:

- X è la matrice di design delle variabili indipendenti (con una colonna di 1 per l'intercetta),
- X^T è la trasposta di X,
- y è il vettore dei valori osservati della variabile dipendente,
- $(X^T X)^{-1}$ è l'inversa del prodotto di X^T e X.

3. Esempio Teorico

Supponiamo di avere un dataset con due variabili indipendenti (x_1 e x_2) e una variabile dipendente (y):

x_1 (ore di studio)	x_2 (ore di sonno)	y (punteggio esame)
2	7	75
3	6	80
4	5	85
5	8	90
6	9	95

Costruzione della Matrice X

La matrice di design X includerà una colonna di 1 per l'intercetta e le colonne per x_1 e x_2:

$$X = \begin{bmatrix} 1 & 2 & 7 \\ 1 & 3 & 6 \\ 1 & 4 & 5 \\ 1 & 5 & 8 \\ 1 & 6 & 9 \end{bmatrix}$$

Vettore dei Valori Osservati y

$$y = \begin{bmatrix} 75 \\ 80 \\ 85 \\ 90 \\ 95 \end{bmatrix}$$

Calcolo dei Coefficienti b

Usiamo la formula $b = (X^T X)^{-1} X^T y$.

1. **Calcola** $X^T X$:

$$X^T X = \begin{bmatrix} 5 & 20 & 35 \\ 20 & 90 & 153 \\ 35 & 153 & 279 \end{bmatrix}$$

2. **Calcola** $(X^T X)^{-1}$: Trova l'inversa di $X^T X$ (può essere calcolata manualmente o con strumenti software).
3. **Calcola** $X^T y$:

$$X^T y = \begin{bmatrix} 425 \\ 1875 \\ 3275 \end{bmatrix}$$

4. **Calcola** b: Moltiplica $(X^T X)^{-1}$ per $X^T y$ per ottenere i coefficienti b_0, b_1, b_2.

4. Implementazione Pratica in Excel

Per calcolare la regressione multivariata in Excel, puoi seguire questi passaggi:

1. **Inserisci i dati**: Metti le variabili indipendenti in colonne adiacenti (es. x_1 in A, x_2 in B) e la variabile dipendente in un'altra colonna (es. y in C).
2. **Utilizza** `REGR.LIN`:
   - Seleziona celle vuote per visualizzare i coefficienti e inserisci la formula:

     excel Copy code

     ```excel
     =REGR.LIN(C2:C6; A2:B6; VERO; VERO)
     ```

     Questa formula restituisce i coefficienti b_2, b_1 e b_0 (l'intercetta).
3. **Interpretazione dei Coefficienti**:
   - I coefficienti b_1 e b_2 rappresentano l'effetto delle variabili indipendenti x_1 e x_2 su y.
   - b_0 è il valore di y quando tutte le variabili indipendenti sono zero.

5. Considerazioni Finali

La regressione multivariata può essere estesa a più di due variabili indipendenti, ma i passaggi per il calcolo rimangono gli stessi. È importante verificare l'adeguatezza del modello attraverso l'analisi dei residui e altre metriche di validità come il coefficiente di determinazione R^2. Se i dati mostrano relazioni non lineari, si può considerare l'uso di trasformazioni delle variabili o modelli non lineari.