Metodi Statistici per le decisioni

2024-2025

Vincenzo Nardelli

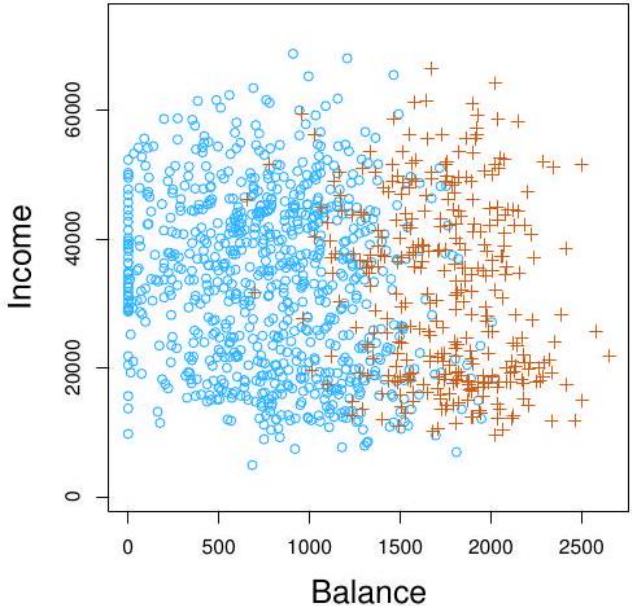
vincenzo.nardelli@unicatt.it

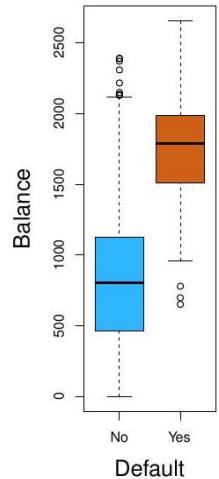
Classificazione

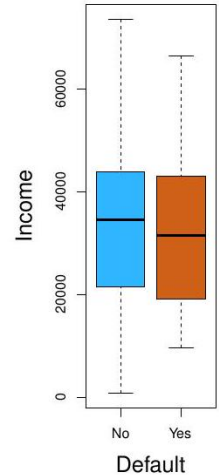
- Le variabili qualitative assumono valori in un insieme non ordinato $\mathcal{C}$, come ad esempio:
  colore degli occhi $\in$ { marrone, blu, verde } email $\in$ { spam, ham}.
- Dato un vettore di caratteristiche X e una risposta qualitativa Y che assume valori nell'insieme $\mathcal{C}$, il compito di classificazione consiste nel costruire una funzione C(X) che prenda in input il vettore X e ne predica il valore per Y; ovvero C(X) $\in \mathcal{C}$.
- Spesso siamo più interessati a stimare le probabilità che X appartenga a ciascuna categoria in $\mathcal{C}$.

Ad esempio, è più utile avere una stima della probabilità che una richiesta di assicurazione sia fraudolenta, piuttosto che una classificazione "fraudolenta" o "non fraudolenta".

Esempio Carta di Credito







Perchè la regressione lineare non basta?

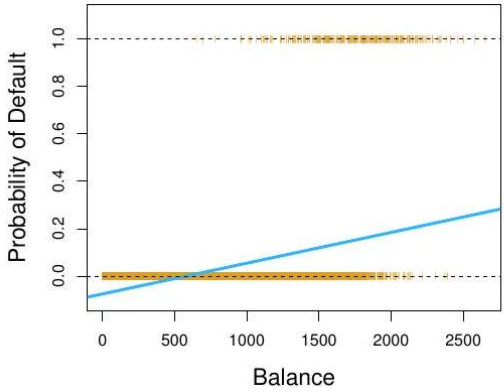
Supponiamo per il compito di classificazione Default di codificare

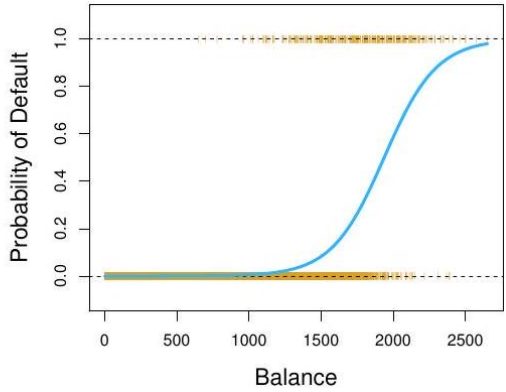
$$Y = \begin{cases} 0 & \text{se No} \\ 1 & \text{se Sì.} \end{cases}$$

Possiamo semplicemente eseguire una regressione lineare di Y su X e classificare come Sì se $\hat{Y} > 0.5$?

- In questo caso di risultato binario, la regressione lineare svolge un buon lavoro come classificatore ed è equivalente all'analisi discriminante lineare che discuteremo più avanti.
- Poiché nella popolazione $E(Y \mid X = x) = Pr(Y = 1 \mid X = x)$, si potrebbe pensare che la regressione sia perfetta per questo compito.
- Tuttavia, la regressione lineare potrebbe produrre probabilità inferiori a zero o superiori a uno. La regressione logistica è più appropriata.

Regressione Lineare VS Regressione Logistica





I segni arancioni indicano la risposta Y, 0 o 1. La regressione lineare non stima bene Pr(Y = 1 | X). La regressione logistica sembra adatta al compito.

Regressione Logistica

Scriviamo p(X) = Pr(Y = 1 | X) per semplicità e consideriamo l'uso del bilancio per prevedere il default. La regressione logistica usa la forma

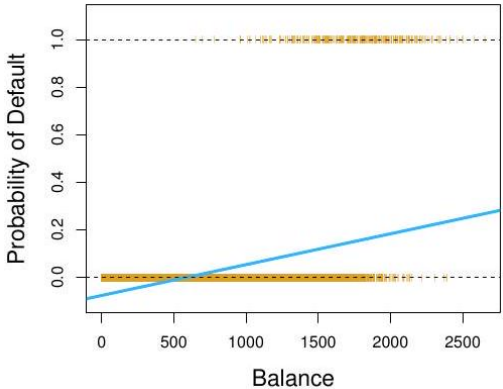
$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}}$$

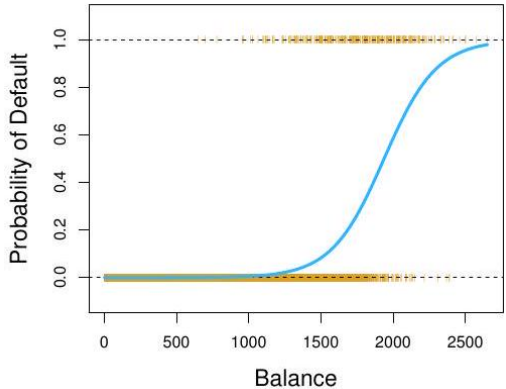
(e ≈ 2.71828 è una costante matematica [numero di Eulero]). È facile vedere che, indipendentemente dai valori di β_0, β_1 o X, p(X) avrà valori compresi tra 0 e 1.
Un po' di riorganizzazione dà

$$\log\left(\frac{p(X)}{1 - p(X)}\right) = \beta_0 + \beta_1 X$$

Questa trasformazione monotona è chiamata trasformazione log-odds o logit di p(X).
(con log intendiamo logaritmo naturale: ln).

Regressione Lineare VS Regressione Logistica





La regressione logistica garantisce che la nostra stima per p(X) sia compresa tra 0 e 1.

Massima Verosimiglianza

Usiamo la massima verosimiglianza per stimare i parametri.

$$\ell\left(\beta_0, \beta\right) = \prod_{i:y_i=1} p\left(x_i\right) \prod_{i:y_i=0} \left(1 - p\left(x_i\right)\right)$$

Questa funzione di verosimiglianza dà la probabilità degli zeri e degli uno osservati nei dati. Scegliamo β_0 e β_1 per massimizzare la probabilità dei dati osservati.

Massima Verosimiglianza

La maggior parte dei pacchetti statistici è in grado di fare il fit di modelli di regressione logistica lineare usando la massima verosimiglianza. In R, utilizziamo la funzione 'glm'.

	Coefficiente	Errore Std.	Statistica Z	Valore-P
Intercept	-10.6513	0.3612	-29.5	< 0.0001
balance	0.0055	0.0002	24.9	< 0.0001

Log-Odds: Coefficienti, Formula e Calcolo

Coefficienti:

$$\text{log-odds} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

- Cambiamenti nelle log-odds associati a incrementi unitari nelle variabili.

Log-odds:

$$\text{log-odds} = \ln\left(\frac{P(y = 1)}{1 - P(y = 1)}\right)$$

- Permettono di modellare la probabilità su una scala lineare.

Odds

$$\text{odds} = e^{\text{log-odds}}$$

- Il rapporto tra la probabilità di successo e di insuccesso.

Perché il modello logistico produce log-odds?

- Nella regressione logistica, il nostro obiettivo è modellare la **probabilità** che un evento accada, P(y = 1).
- Tuttavia, P(y = 1) è limitato tra 0 e 1, quindi non può essere modellata direttamente come combinazione lineare delle variabili.
- Per superare questa limitazione:
  - Si calcolano le **odds**, cioè il rapporto tra la probabilità di successo e quella di insuccesso:

$$\text{odds} = \frac{P(y = 1)}{1 - P(y = 1)}$$

  - Si applica il logaritmo naturale alle odds, ottenendo le **log-odds**, che hanno un intervallo illimitato:

$$\text{log-odds} = \ln\left(\frac{P(y = 1)}{1 - P(y = 1)}\right)$$

- La regressione logistica modella le log-odds come combinazione lineare delle variabili esplicative:

$$\text{log-odds} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots$$

Interpretazione dell'Intercetta nella Regressione Logistica

Definizione dell'Intercetta (β_0):

- L'intercetta rappresenta le **log-odds** dell'evento di successo (y = 1) quando tutte le variabili esplicative sono uguali a 0.
- Può essere trasformata in una probabilità:

$$P(y = 1) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$$

Interpretazione:

- $\beta_0 > 0$: L'evento è più probabile ($P(y = 1) > 0.5$) quando $x_1, x_2, \cdots = 0$.
- $\beta_0 = 0$: L'evento ha probabilità $P(y = 1) = 0.5$ quando $x_1, x_2, \cdots = 0$.
- $\beta_0 < 0$: L'evento è meno probabile ($P(y = 1) < 0.5$) quando $x_1, x_2, \cdots = 0$.

Considerazioni:

- L'intercetta ha senso solo se x = 0 è un valore interpretabile per tutte le variabili.
- Se x = 0 non ha significato pratico, l'intercetta è difficile da interpretare.

Interpretazione dei risultati

Supponiamo di avere i seguenti risultati di regressione logistica:

	Coefficiente	Errore Std.	Statistica Z	Valore-P
Intercept	-10.6513	0.3612	-29.5	< 0.0001
balance	0.0055	0.0002	24.9	< 0.0001

- **Intercept (-10.6513)**:
  - Rappresenta le log-odds quando tutte le variabili esplicative sono uguali a 0.
  - Essendo molto negativo, le probabilità di successo (P(y = 1)) sono molto basse quando balance = 0.
- **balance (0.0055)**:
  - Il coefficiente positivo indica che all'aumentare di balance, aumentano le log-odds, quindi la probabilità di successo.
  - Interpretazione delle odds: $e^{0.0055} \approx 1.0055$. Per ogni aumento unitario di balance, le odds aumentano dello 0.55%.
  - Il valore-p molto basso (< 0.0001) indica che l'effetto è statisticamente significativo.

Fare Previsioni per Balance = $1000

Log-odds:

$$\text{log-odds} = \hat{\beta}_0 + \hat{\beta}_1 \cdot X = -10.6513 + 0.0055 \cdot 1000 = -5.1524$$

Odds:

$$\text{odds} = e^{\text{log-odds}} = e^{-5.1524} \approx 0.0058$$

Probabilità:

$$\hat{p}(X) = \frac{\text{odds}}{1 + \text{odds}} = \frac{0.0058}{1 + 0.0058} \approx 0.006$$

Interpretazione:

- **Log-odds**: Le log-odds di default sono −5.1524, indicando una probabilità estremamente bassa.
- **Odds**: Le odds sono 0.0058, ovvero circa 1 possibilità su 173 che il default avvenga.
- **Probabilità**: La probabilità stimata di default è 0.006, cioè lo 0.6%.

Fare Previsioni per Balance = $2000

Log-odds:

$$\text{log-odds} = \hat{\beta}_0 + \hat{\beta}_1 \cdot X = -10.6513 + 0.0055 \cdot 2000 = 0.3465$$

Odds:

$$\text{odds} = e^{\text{log-odds}} = e^{0.3465} \approx 1.414$$

Probabilità:

$$\hat{p}(X) = \frac{\text{odds}}{1 + \text{odds}} = \frac{1.414}{1 + 1.414} \approx 0.586$$

Interpretazione:

- **Log-odds**: Le log-odds di default sono 0.3465, indicando una probabilità maggiore del 50%.
- **Odds**: Le odds sono 1.414, ovvero il default è circa 1.41 volte più probabile rispetto al non default.
- **Probabilità**: La probabilità stimata di default è 0.586, cioè il 58.6%.

Regressione logistica con più variabili

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

$$p(X) = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}$$

	Coefficiente	Errore Std.	Statistica Z	Valore-P
Intercept	-10.8690	0.4923	-22.08	< 0.0001
balance	0.0057	0.0002	24.74	< 0.0001
income	0.0030	0.0082	0.37	0.7115
student [Sì]	-0.6468	0.2362	-2.74	0.0062

La devianza in un modello logistico

Formula della devianza:

$$D = -2 \cdot \left(\ell_{\text{modello}} - \ell_{\text{modello di saturazione}}\right)$$

Dove:

- ℓ_{modello}: La log-verosimiglianza del modello stimato.
- $\ell_{\text{modello di saturazione}}$: La log-verosimiglianza di un modello perfetto che si adatta esattamente ai dati (massima possibile).

Spiegazione:

- La devianza misura quanto il modello corrente si discosta da un modello perfetto.
- Valori più bassi di devianza indicano un miglior adattamento ai dati.
- Si confrontano due tipi di devianza:
  - **Devianza nulla**: Modello con solo l'intercetta (nessun predittore).
  - **Devianza residua**: Modello con i predittori inclusi.
- La differenza tra devianza nulla e devianza residua indica quanto i predittori migliorano il modello.

Akaike Information Criterion (AIC)

Formula dell'AIC:

$$AIC = -2 \cdot \ell_{modello} + 2 \cdot k$$

Dove:

- $\ell_{modello}$: La log-verosimiglianza del modello stimato.
- k: Il numero di parametri stimati nel modello (inclusa l'intercetta).

Spiegazione:

- L'AIC valuta un modello bilanciando:
  - **Adattamento**: Quanto bene il modello si adatta ai dati (log-verosimiglianza).
  - **Semplicità**: Penalizza modelli troppo complessi (numero di parametri).
- Modelli con un AIC più basso sono preferibili.
- L'AIC è utile per confrontare modelli: un modello con AIC più basso è considerato migliore.

Nota: L'AIC non misura la qualità assoluta di un modello, ma è un criterio comparativo tra modelli con gli stessi dati.

Confusion Matrix e Accuracy

Definizione della Confusion Matrix:

	Predetto: No	**Predetto: Yes**
Effettivo: No	**True Negative (TN)**	**False Positive (FP)**
Effettivo: Yes	**False Negative (FN)**	**True Positive (TP)**

Terminologia chiave:

- **True Positive (TP):** Predetto "Yes" e l'evento è effettivamente "Yes".
- **False Positive (FP):** Predetto "Yes" ma l'evento è effettivamente "No".
- **True Negative (TN):** Predetto "No" e l'evento è effettivamente "No".
- **False Negative (FN):** Predetto "No" ma l'evento è effettivamente "Yes".

Accuracy:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{Totale delle osservazioni}}$$

Confusion Matrix e Accuracy

Risultati ottenuti:

	Effettivo: No	**Effettivo: Yes**
Predetto: No	9625	233
Predetto: Yes	42	100

Calcolo dell'Accuracy:

$$\text{Accuracy} = \frac{9625 + 100}{9625 + 233 + 42 + 100} = 0.9725\,(97.25\%)$$

La correttezza non è tutto

Problema:

- Un modello può avere un'alta accuratezza, ma potrebbe non essere ottimale in termini di falsi positivi (FP) e falsi negativi (FN).
- Dipende dal contesto: alcuni errori sono più costosi di altri.

Errori nel modello:

- **Falsi Positivi (FP):** Il modello predice positivi, ma sono negativi.
- **Falsi Negativi (FN):** Il modello predice negativi, ma sono positivi.

Obiettivo:

- In alcune situazioni, vogliamo minimizzare i FP.
- In altre, vogliamo minimizzare i FN.

Conclusione: La scelta tra ridurre FP e FN dipende dal contesto applicativo.

Quando ridurre FP e FN?

Ridurre i Falsi Positivi:

- Sistemi di raccomandazione:
  - Evitare di inviare offerte a utenti non interessati.
- Cybersecurity:
  - Evitare falsi allarmi che distraggono il team IT.
- Controllo qualità:
  - Evitare di scartare prodotti conformi.

Ridurre i Falsi Negativi:

- Diagnosi medica:
  - Identificare tutti i pazienti con una malattia grave.
- Customer Retention:
  - Identificare tutti i clienti a rischio di abbandono.
- Frodi:
  - Identificare tutte le transazioni fraudolente.

Cos'è la Sensitività?

Definizione:

$$\text{Sensitività (Recall)} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}}$$

Spiegazione:

- Misura la capacità del modello di identificare correttamente i positivi.
- Una sensibilità alta significa pochi falsi negativi.

Quando applicarla?

- **Diagnosi medica:** Identificare tutti i pazienti malati.
- **Frodi:** Assicurarsi che tutte le transazioni fraudolente siano segnalate.

Trade-off:

- Una sensitività alta può comportare un aumento dei falsi positivi.

Cos'è la Specificità?

Definizione:

$$\text{Specificità} = \frac{\text{True Negative (TN)}}{\text{True Negative (TN)} + \text{False Positive (FP)}}$$

Spiegazione:

- Misura la capacità del modello di identificare correttamente i negativi.
- Una specificità alta significa pochi falsi positivi.

Quando applicarla?

- **Sistemi di raccomandazione:** Evitare di disturbare utenti non interessati.
- **Cybersecurity:** Minimizzare gli allarmi inutili per attacchi inesistenti.

Trade-off:

- Una specificità alta può comportare un aumento dei falsi negativi.

Recap: Confusion Matrix e Metriche

Confusion Matrix:

	Predetto: No	**Predetto: Yes**
Effettivo: No	True Negative (TN)	False Positive (FP)
Effettivo: Yes	False Negative (FN)	True Positive (TP)

Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Sensitività:

$$\text{Sensitività} = \frac{TP}{TP + FN}$$

Specificità:

$$\text{Specificità} = \frac{TN}{TN + FP}$$

Trade-off:

- **Alta Sensitività:** Più FP, meno FN.
- **Alta Specificità:** Più FN, meno FP.

Spostare la Soglia: Bilanciare FP e FN

Spostare la soglia di classificazione:

- La soglia determina il punto in cui una probabilità viene classificata come positiva o negativa.

Effetti di soglie diverse:

- **Soglia bassa:**
  - Aumenta la sensibilità (meno falsi negativi).
  - Riduce la specificità (più falsi positivi).
- **Soglia alta:**
  - Aumenta la specificità (meno falsi positivi).
  - Riduce la sensibilità (più falsi negativi).

Esempio pratico: Diagnosi medica

- **Soglia bassa:** Identificare tutti i malati (anche con falsi allarmi).
- **Soglia alta:** Evitare di diagnosticare erroneamente pazienti sani.

Confronto tra diverse soglie di classificazione

Soglia = 0.2

$$\begin{bmatrix} 9404 & 134 \\ 263 & 199 \end{bmatrix}$$

- **Accuracy:** 0.9603
- **Sensitività:** 0.5976
- **Specificità:** 0.9728

Soglia = 0.5

$$\begin{bmatrix} 9625 & 233 \\ 42 & 100 \end{bmatrix}$$

- **Accuracy:** 0.9725
- **Sensitività:** 0.3003
- **Specificità:** 0.9957

Soglia = 0.8

$$\begin{bmatrix} 9662 & 307 \\ 5 & 26 \end{bmatrix}$$

- **Accuracy:** 0.9688
- **Sensitività:** 0.0781
- **Specificità:** 0.9995

- **Soglia 0.2**: Alta sensibilità, ma ridotta specificità (molti falsi positivi).
- **Soglia 0.5**: Bilanciamento tra sensibilità e specificità.
- **Soglia 0.8**: Alta specificità, ma ridotta sensibilità (molti falsi negativi).

LAB Puntualità delle Consegne nell'E-Commerce

La puntualità delle consegne è un aspetto fondamentale per garantire la soddisfazione del cliente nel settore dell'e-commerce. Attraverso l'analisi dei dati delle spedizioni, è possibile identificare i fattori che influenzano maggiormente i ritardi, consentendo all'azienda di ottimizzare le operazioni logistiche e migliorare l'efficienza.

Obiettivo dell'analisi:

- Identificare i fattori chiave che portano a ritardi nelle consegne.
- Creare un modello predittivo per anticipare i ritardi.
- Fornire raccomandazioni operative per mitigare i ritardi e migliorare la soddisfazione del cliente.

Descrizione delle Variabili - Dataset Shipping

Il dataset contiene informazioni raccolte su spedizioni e clienti per analizzare i ritardi nelle consegne. Le principali variabili includono:

- **mode_of_shipment**: Modalità di spedizione (Nave, Aereo, Strada).
- **customer_rating**: Valutazione del cliente (1 = peggiore, 5 = migliore).
- **cost_of_the_product**: Costo del prodotto (USD).
- **prior_purchases**: Numero di acquisti precedenti del cliente.
- **product_importance**: Importanza del prodotto (bassa, media, alta).
- **gender**: Genere del cliente (Maschio, Femmina).
- **discount_offered**: Sconto offerto sul prodotto.
- **weight_in_gms**: Peso del prodotto in grammi.
- **reached_time**: Variabile target (1 = in ritardo, 0 = puntuale).

LAB Puntualità delle Consegne: Attività

1. Esplora il dataset e verifica la distribuzione delle variabili. Controlla eventuali valori mancanti o incoerenti.
2. Costruisci un modello di regressione logistica con `Reached.on.Time_Y.N` come variabile dipendente e le altre variabili come predittori.
3. Analizza i coefficienti del modello per identificare i fattori significativi e il loro impatto sulla probabilità di ritardo.
4. Sviluppa raccomandazioni operative basate sui risultati del modello per ridurre i ritardi nelle consegne.

LAB: Analisi dell'Efficacia delle Campagne di Marketing Bancario

L'analisi dell'efficacia delle campagne di marketing bancario si concentra su come alcune variabili chiave influenzano la decisione dei clienti di sottoscrivere un deposito a termine. Questo caso studio utilizza dati raccolti durante campagne di marketing telefonico.

Obiettivi dell'analisi:

- Identificare l'impatto della durata della chiamata e delle condizioni finanziarie del cliente sulla decisione.
- Creare un modello predittivo basato sulla regressione logistica per stimare la probabilità di sottoscrizione.
- Valutare l'efficacia del modello tramite metriche come sensibilità e specificità.

Descrizione delle Variabili - Dataset Bank

Il modello utilizza le seguenti variabili:

- **duration**: Durata dell'ultimo contatto in secondi (variabile numerica).
- **housing**: Possesso di un mutuo sulla casa (1 = "yes", 0 = "no").
- **loan**: Possesso di un prestito personale (1 = "yes", 0 = "no").
- **y**: Variabile target che indica se il cliente ha sottoscritto un deposito a termine (1 = "yes", 0 = "no").

Obiettivo: Stimare la probabilità che un cliente sottoscriva un deposito a termine.

LAB Bank Marketing Analysis: Attività

1. Esplora il dataset per verificare la distribuzione delle variabili.
2. Trasforma le variabili categoriali 'housing', 'loan', e 'y' in valori numerici (1 e 0).
3. Costruisci un modello di regressione logistica utilizzando:

$$y \sim \text{duration} + \text{housing} + \text{loan}$$

4. Analizza i coefficienti del modello per interpretare l'impatto di ciascuna variabile sulla probabilità di sottoscrizione.
5. Valuta il modello calcolando:
   - Accuratezza,
   - Sensibilità,
   - Specificità.
6. Prova soglie diverse per ottimizzare le metriche del modello.

Utilità di Specificità e Sensibilità

Sensibilità (Recall):

- Indica la capacità del modello di identificare correttamente i clienti che sottoscriveranno un deposito.
- Una sensibilità alta riduce le opportunità mancate, identificando la maggior parte dei clienti propensi alla sottoscrizione.

Specificità:

- Misura la capacità del modello di identificare correttamente i clienti che non sottoscriveranno il deposito.
- Una specificità alta garantisce che le risorse non vengano sprecate su clienti non interessati.

Bilanciare sensibilità e specificità è cruciale per massimizzare l'efficacia delle campagne di marketing bancario.

LAB Analisi dell'Accettazione dei Coupon da Parte degli Autisti

L'analisi dell'accettazione dei coupon da parte degli autisti è cruciale per comprendere come variabili come destinazione, meteo, presenza di passeggeri e distanza influiscano sulle decisioni. Questo consente alle aziende di ottimizzare le strategie promozionali, ridurre gli sprechi e migliorare i profitti. Il caso studio si basa su dati reali raccolti in diversi scenari di guida, dove si valuta se gli autisti accettano o meno i coupon offerti.

Obiettivo dell'analisi:

- Identificare i fattori chiave che influenzano la decisione degli autisti di accettare un coupon durante la guida.
- Creare un modello predittivo basato sulla regressione logistica per stimare la probabilità di accettazione dei coupon.
- Analizzare i costi aziendali associati agli errori del modello, come falsi positivi (offerta di coupon inutili) e falsi negativi (perdita di opportunità), al fine di migliorare la strategia promozionale.

Descrizione delle Variabili - Dataset Coupon

Il dataset analizza scenari di guida e preferenze dei clienti per prevedere l'accettazione dei coupon. Le principali variabili includono:

- **destination**: Destinazione attuale (No Urgent Place, Home, Work).
- **passenger**: Passeggeri in macchina (Alone, Friend(s), Kid(s), Partner).
- **weather**: Condizioni meteo (Sunny, Rainy, Snowy).
- **coupon**: Tipo di coupon offerto (Restaurant, Coffee House, Bar, etc.).
- **income**: Fascia di reddito (<12.5k, 12.5k–25k, >100k, etc.).

Descrizione delle Variabili - Dataset Coupon

- **RestaurantLessThan20**: Frequenza di visite a ristoranti con spesa media inferiore a $20 (4-8, 1-3, less1, gt8, never).
- **Restaurant20To50**: Frequenza di visite a ristoranti con spesa media tra $20-$50 (1-3, less1, never, gt8, 4-8, nan).
- **toCoupon_GEQ15min**: Distanza in macchina >15 minuti (1 = sì, 0 = no).
- **toCoupon_GEQ25min**: Distanza in macchina >25 minuti (1 = sì, 0 = no).
- **direction_same**: Il coupon è nella stessa direzione della destinazione (1 = sì, 0 = no).
- **direction_opp**: Il coupon è nella direzione opposta alla destinazione (1 = sì, 0 = no).
- **Y**: Variabile target (1 = coupon accettato, 0 = non accettato).

LAB Coupon Analysis: Attività

1. Esplora il dataset per comprendere la distribuzione delle variabili e identificare valori mancanti o incoerenti.
2. Costruisci un modello di regressione logistica con `Y` come variabile dipendente e le altre variabili come predittori.
3. Analizza i coefficienti del modello per identificare i fattori significativi e il loro impatto sulla probabilità di accettazione del coupon.
4. Valuta l'errore del modello, calcolando l'impatto economico di falsi positivi e falsi negativi.

Valutazione dell'Errore e Impatti Economici

In un contesto aziendale, i costi associati a falsi positivi e falsi negativi sono fondamentali:

- **Falsi Positivi (FP)**: Il modello prevede che il cliente accetterà il coupon, ma in realtà non lo fa.
  - Costo aziendale: Spese di marketing per l'invio di un coupon (0.3€ per coupon inviato).
- **Falsi Negativi (FN)**: Il modello prevede che il cliente non accetterà il coupon, ma in realtà lo avrebbe fatto.
  - Costo aziendale: Perdita di opportunità di guadagno (4.5€ per coupon non accettato ma potenzialmente valido).

Calcolo dei costi totali:

- Costo totale FP = FP $\times$ 0.3
- Costo totale FN = FN $\times$ 4.5