

0.1 Regressione Logistica

La regressione logistica é un modello di regressione non lineare utilizzata quando la variabile dipendente é di tipo binario $[0, 1]$. Con questo modello stabiliamo la probabilità con cui un'osservazione può generare uno o l'altro valore della variabile dipendente; può essere utilizzato inoltre per classificare le osservazioni in due categorie(come nel nostro caso), facendo riferimento a un problema di classificazione. A differenza dei modelli di regressione lineare che sono utilizzati per identificare la relazione tra una variabile dipendente continua e una o più variabili indipendenti, la regressione logistica stima la probabilità del verificarsi di un evento, sulla base di uno specifico dataset di variabili indipendenti; verso un insieme di dominio della variabile dipendente $[0, 1]$. Una volta calcolato il vettore $\hat{\beta}$ ossia la stima del vettore dei parametri β , é possibile procedere alla stima della probabilità p . Per definizione questa probabilità é anche il valore atteso di Y .

$$\hat{p} = E[Y|X] = \Lambda(X^T \hat{\beta}) = \frac{e^{X^T \hat{\beta}}}{1 + e^{X^T \hat{\beta}}}$$

dove:

- Y é la variabile dipendente binaria
- X il vettore di variabili indipendenti(o regressori) X_1, \dots, X_k
- β il vettore di parametri β_0, \dots, β_k
- Λ é la funzione di ripartizione della distribuzione logistica standard
- e il numero di eulero

La regressione logistica appartiene alla famiglia di modelli supervisionati, inoltre, é considerato un modello discriminante che prova a distinguere tra classi(o categorie). La regressione logistica può essere incline all'**overfitting** in particolare quando é presente un numero elevato di variabili predittive nel modello.