Consegna Data Analysis

Antonio Sessa   Angelo Molinario   Pietro Martano
Massimiliano Ranauro

30 January 2025

Università degli Studi di Salerno

Name	**Surname**	**Serial**	**E-mail**
Antonio	Sessa	0622702305	a.sessa108@studenti.unisa.it
Angelo	Molinario	0622702311	a.molinario3@studenti.unisa.it
Massimiliano	Ranauro	0622702373	m.ranauro2@studenti.unisa.it
Pietro	Martano	0622702402	p.martano@studenti.unisa.it

Contents

1 Classificazione

1.1 Calcolo della distribuzione a posteriori

Dal teorema di Bayes sappiamo che la probabilità a posteriori ha la seguente forma:

$$P(+1|x) = \frac{\pi(+1) \cdot l(x|y=+1)}{\pi(+1) \cdot l(x|y=+1) + \pi(-1) \cdot l(x|y=-1)}$$

Essendo le probabilità a priori uniformi allora possiamo eliminarle dal calcolo.

$$P(+1|x) = \frac{l(x|y=+1)}{l(x|y=+1) + l(x|y=-1)}$$

$$P(+1|x) = \frac{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-1)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-1)^2}{2\sigma^2}} + \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x+1)^2}{2\sigma^2}}}$$

$$P(+1|x) = \frac{e^{-\frac{(x-1)^2}{2\sigma^2}}}{e^{-\frac{(x-1)^2}{2\sigma^2}} + e^{-\frac{(x+1)^2}{2\sigma^2}}}$$

Per semplificare la scrittura possiamo moltiplicare e dividere per $e^{-\frac{(x+1)^2}{2\sigma^2}}$

$$P(+1|x) = \frac{1}{1 + \frac{e^{-\frac{(x+1)^2}{2\sigma^2}}}{e^{-\frac{(x-1)^2}{2\sigma^2}}}}$$

$$P(+1|x) = \frac{1}{1 + e^{\frac{-2x}{\sigma^2}}}$$

La pmf di P(-1|x) può essere calcolata semplicemente a partire dalla posteriori P(+1|x) come di seguito

$$P(-1|x) = 1 - P(1|x) = \frac{1 + e^{\frac{-2x}{\sigma^2}} - 1}{1 + e^{\frac{-2x}{\sigma^2}}} = \frac{1}{1 + e^{\frac{2x}{\sigma^2}}}$$

Dato che le due label considerate per la y $\in \{-1, +1\}$ possiamo scrivere la pmf a posteriore P(y|x) in forma compatta.

$$P(y|x) = \frac{1}{1 + e^{\frac{-y(2x)}{\sigma^2}}}$$

Possiamo anche calcolare il classificatore MAP:

$$\hat{y}_{MAP} = \arg\max \left(\frac{1}{1 + e^{-\frac{y(2x)}{\sigma^2}}} \right)$$

che equivale a verificare il segno della x, se positivo assegno la classe +1 altrimenti -1, questo è confermato anche osservando come 0 è il punto equidistante tra le due medie dei label e che delimita le due aree di decisione, come si può vedere in figura 3.

1.2 Scelta delle deviazioni standard

Come deviazione standard ”buona”, abbiamo scelta una deviazione standard pari a **0.8**. Come deviazione standard ”cattiva”, abbiamo scelta una deviazione standard pari a **3**.

Presentiamo ora i grafici delle likelihood per le due deviazioni standard in figura 1 e 2

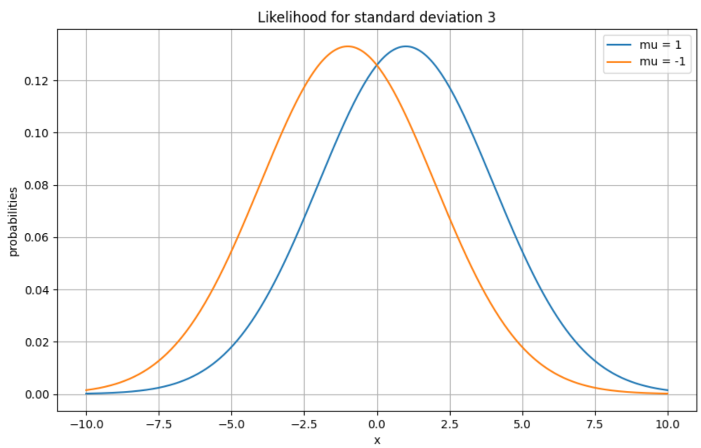


Figure 1: Deviazione standard = 3

Dalle figure, è possibile notare come le likelihood con una deviazione standard di 3 presentano una sovrapposizione maggiore rispetto al caso 0.8. Più i dati sono ”sovrapposti” più un classificatore farà difficoltà a seperarli. Dato che le medie del label 1 e label -1 sono releativamente vicine, anche nel caso ”buono” con deviazione standard 0.8 è presente una sovrapposizione ”relativamente” importante. Dato che la traccia specificava di non scegliere voler estremi per le varianze, confermiamo le nostre scelte.

Possiamo mettere in relazione la difficoltà del problema di classificazione con la forme delle curve in figura 3, notando come nel caso in cui la deviazione standard è di 0.8, abbiamo un netto incremento della probabilità a posteriori a nell'intorno di x = 0, ovvero il valore che separa le due aree di decisione (che sono identitiche per entrambi i problemi dato che le mediee rimangono invariate nei due casi), mentre nel caso di deviaizone standard uguale a 3, c'è un incremento graduale. Significativo notare come per un valore di x uguale alla media, cioè x = 1, la posteriori con sigma buono raggiunge un valore di quasi 1, mentre la posteriori con varianza alta, presenta un valore inferiore a 0.6 .

In figura 3 è graficata le probabilità a posteriori per il label di y uguale a 1 per le due varianze.

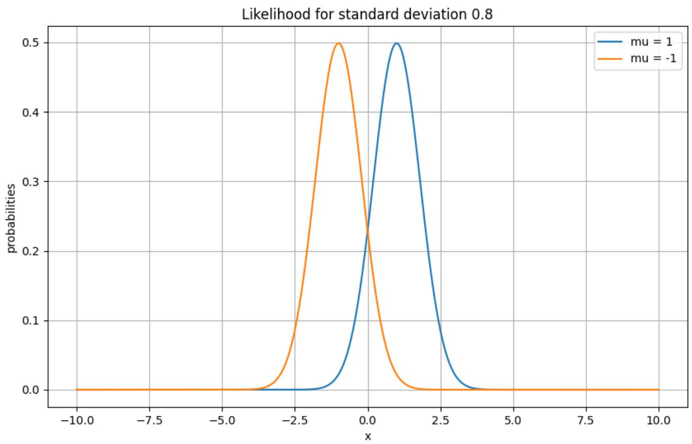


Figure 2: Deviazione standard = 0.8

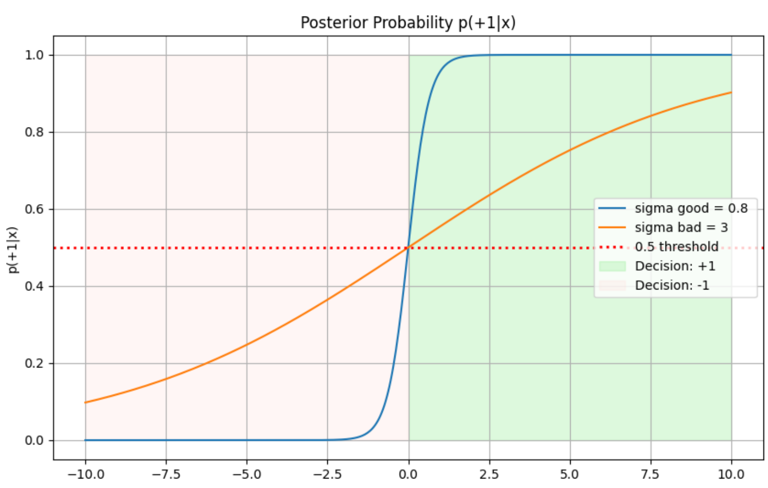


Figure 3: Confronto posteriori con le diverse varianze

1.3 Prove montecarlo

Probabilità di errore	**Deviazione Standard**
0.370	3
0.105	0.8

Table 1: Risultati delle simulazioni Montecarlo

Come ci aspettavamo, maggiore è la varianza nei dati, maggiore è l'errore del classificatore. Infatti, dato che conosciamo le priori e le posteriori, possiamo analicamente calcolare la probabilità di errore teorica come:

$$P(\text{err}) = P(1) \cdot P(x < 0 \mid y = 1) + P(-1) \cdot P(x > 0 \mid y = -1)$$

$$P(\text{err}) = P(y = 1) \cdot P(X < 0 \mid X \sim \mathcal{N}(1, 0.8^2)) + P(y = -1) \cdot P(X > 0 \mid X \sim \mathcal{N}(-1, 0.8^2)) = 0.1056$$

$$P(\text{err}) = P(y = 1) \cdot P(X < 0 \mid X \sim \mathcal{N}(1, 3^2)) + P(y = -1) \cdot P(X > 0 \mid X \sim \mathcal{N}(-1, 3^2)) = 0.3694$$

In poche parole, una varianza maggiore aumenta la probabilità che un campione con media 1 (o -1) abbia un valore inferiore (o maggiore) di 0, causando una classificazione sbagliata.

1.4 Logistic

Per la regressione logistica abbiamo scelto di generare un training set di 100 mila campioni, e calcolare la probabilità di errore empirica su un test set di 100 mila campioni. Il learning rate scelto è di 0.001 ed è rimasto costante durante l'addestramento. Per valutare le perfomance dell'addestramento, dato che conoscosciamo la likelihood e la priori, e dato che il logaritmo del loro rapporto ha forma lineare, possiamo calcolare i beta veri.

$$\beta_0 + \beta_1 x = \frac{2}{\sigma^2} x$$

$$\beta_0 = 0 \ , \ \beta_1 = \frac{2}{\sigma^2}$$

Date queste favorevoli circostanze (forma lineare della statistica di decisione ottima e varianza bassa) ci aspettiamo una buona perfomance del classificatore logistico, che dovrebbe essere solamente di poco inferiore rispetto al classificatore ottimo (MAP).

Infatti confrontando le perfomance[1] sullo stesso test-set (n=100000) tra il classificatore logistico e quello map, le nostre aspettative sono rispettate.

[1]Riportiamo i valori fino alla terza cifra decimale, dato che i successivi non hanno una grande importanza con il numero di campioni scelto

- Probabilità di errore logistico: 0.106
- Probabilità di errore map: 0.105
- Valore dei beta: [-0.165, 2.47]
- Valore dei beta ”veri” : [0, 3.125]

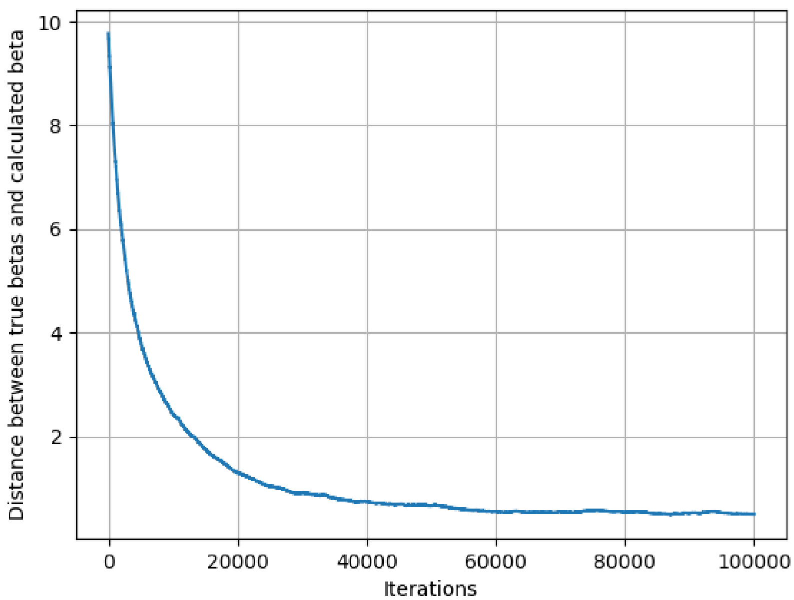


Figure 4: Distanza tra i beta veri e beta dell’addestramento

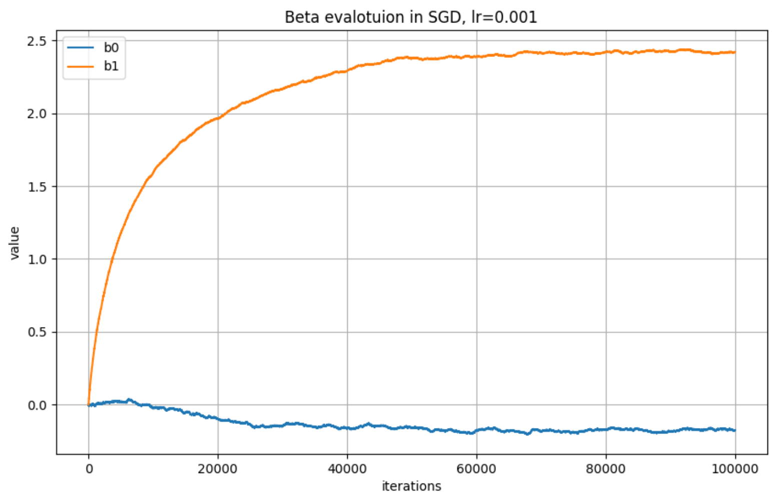
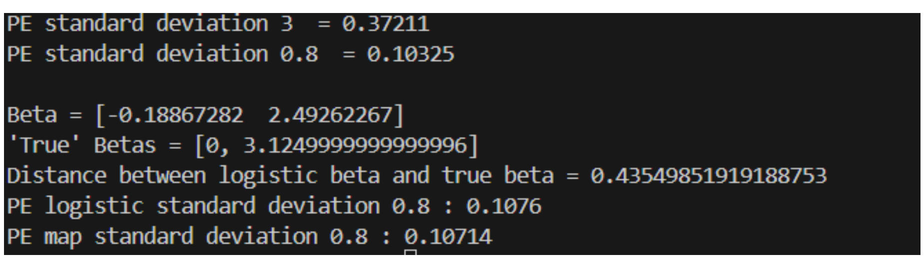


Figure 5: Valore di beta durante l'addestramento



```
PE standard deviation 3  = 0.37211
PE standard deviation 0.8  = 0.10325

Beta = [-0.18867282  2.49262267]
'True' Betas = [0, 3.1249999999999996]
Distance between logistic beta and true beta = 0.43549851919188753
PE logistic standard deviation 0.8 : 0.1076
PE map standard deviation 0.8 : 0.10714
```

Figure 6: Output del codice python