Università degli Studi di Salerno

Dipartimento di Ingegneria dell'Informazione ed Elettrica e Matematica Applicata

Corso di Laurea Magistrale in Ingegneria Informatica

Data Science 30/01/2025
Canale A-H

Project Work

Quesito 2 - Classificazione

Gruppo n. **07 – AH**

Cognome e Nome	Matricola	e-mail
Apicella Antonio	0622702531	a.apicella97@studenti.unisa.it
Celano Benedetta Pia	0622702558	b.celano1@studenti.unisa.it
Cirillo Francesco Pio	0622702466	f.cirillo36@studenti.unisa.it
Fasolino Alessandra	0622702465	a.fasolino35@studenti.unisa.it

Sommario

1. Caso 1: modello statistico perfettamente noto 2
1.1. Calcolo della pmf a posteriori 2
1.2. Rappresentazione grafica di $\boldsymbol{p}(\boldsymbol{y} = +\mathbf{1} \mid \boldsymbol{x})$ in due casi a diversa varianza 3
1.2.1. Conclusioni tratte dall'osservazione del grafico 3
1.3. Calcolo dell'errore empirico per mezzo di simulazioni Montecarlo 4
2. Caso 2: classificazione supervisionata 5
2.1. Classificazione supervisionata per mezzo dell'utilizzo di Stochastic Gradient Descent 5
2.2. Calcolo empirico delle prestazioni dei classificatori 7
2.2.1. Confronti tra le probabilità d’errore relative ai diversi approcci 8

1. Caso 1: modello statistico perfettamente noto

1.1. Calcolo della pmf a posteriori

$$p(y|x) = \frac{\pi(y)\cdot \ell(x|y)}{\sum \pi(y')\,\ell(x|y')} = \frac{\pi(y)\,\ell(x|y)}{\pi(1)\,\ell(x|y=1) + \pi(-1)\,\ell(x|y=-1)} =$$

$$= \frac{\pi(y)\,\ell(x|y)}{\frac{1}{2}\left(\ell(x|y=-1) + \ell(x|y=1)\right)} \overset{(2)}{=} \frac{\ell(x|y)}{\ell(x|y=-1) + \ell(x|y=1)} =$$

$$= \frac{\ell(x|y)}{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x+1)^2}{2\sigma^2}\right\} + \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-1)^2}{2\sigma^2}\right\}} =$$

$$\overset{(1)}{=} \frac{\exp\left\{-\frac{(x-y)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(x+1)^2}{2\sigma^2}\right\} + \exp\left\{-\frac{(x-1)^2}{2\sigma^2}\right\}}$$

(1) $$\ell(x|y=y) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-y)^2}{2\sigma^2}\right\}$$

(2) $$\pi(y) = \pi(-1) = \pi(1) = \frac{1}{2}$$

Figura 1: Calcolo pmf a posteriori

Il calcolo mostrato in figura è semplicemente il calcolo della posteriori per mezzo della regola di Bayes.

1.2. Rappresentazione grafica di $p(y = +1 \mid x)$ in due casi a diversa varianza

Quanto più bassa è la varianza tanto più ci aspettiamo che il classificatore sia preciso dato che diminuisce la dispersione dei dati. Per non scegliere misure estreme, come richiesto, si sono assunti i valori variance_good=1 e variance_bad=2.

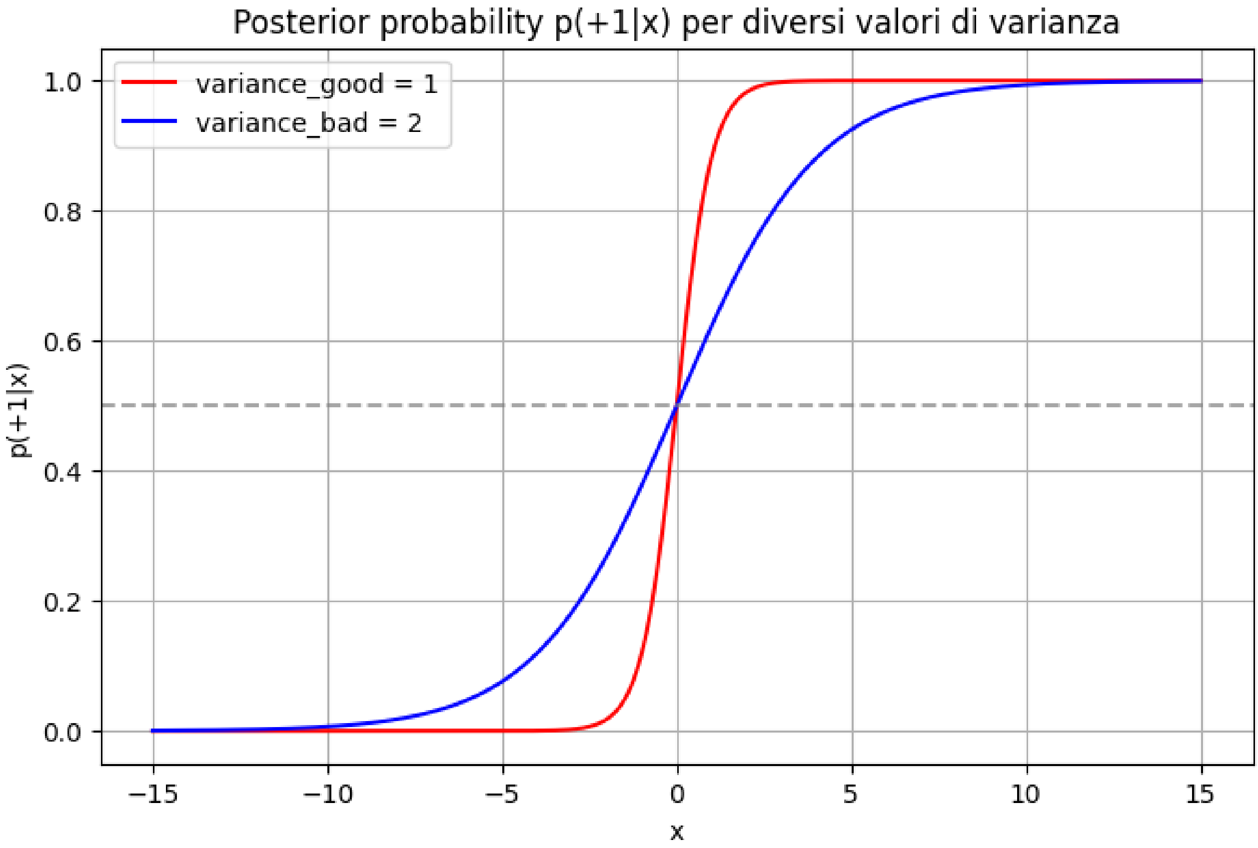

Figura 2: Grafico che mostra la posteriori nel caso in cui y=+1

1.2.1. Conclusioni tratte dall'osservazione del grafico

Possiamo interpretare il grafico $p(y = 1|x)$ (Figura 2) come il grafico che esprime la probabilità che la y sia uguale ad 1 data la specifica osservazione di x della quale si è in possesso.

Ciò che notiamo è che quando la varianza è più alta (grafico blu), e quindi le realizzazioni della gaussiana sono più disperse, è più difficile dire se una certa x appartiene alla classe -1 o alla classe 1, soprattutto quando le realizzazioni di x sono vicine al punto medio tra le due distribuzioni che generano i dati.

Infatti per il grafico rosso la probabilità che la y relativa alla x sia 1 passa rapidamente da essere 0 ad essere 1, lasciando quindi un piccolo range di valori della x (tra -1 e 1) per i quali c'è incertezza. Invece per il grafico blu c'è un range più alto di valori per i quali non è certa la loro appartenenza alla classe -1 o 1, all'incirca tra $x = -5$ e $x = 5$.

1.3. Calcolo dell'errore empirico per mezzo di simulazioni Montecarlo

Notiamo che nel caso di varianza più alta la probabilità d'errore stimata è circa 0.31, nel caso di varianza più bassa la probabilità d'errore stimata è invece circa 0.16 (circa la metà).

Questo risultato risulta in linea con le aspettative, era infatti prevedibile che nel caso in cui le realizzazioni delle due gaussiane sono meno disperse è più facile capire qual è la distribuzione che ha generato i dati, permettendo quindi una classificazione più agevole e con meno errori.

2. Caso 2: classificazione supervisionata

2.1. Classificazione supervisionata per mezzo dell'utilizzo di Stochastic Gradient Descent

Le beta sono calcolate senza l'ausilio di una tecnica Montecarlo (Figura 3) e per mezzo di una tecnica Montecarlo (Figura 4) al fine di confrontare la qualità dei risultati in queste due casistiche.

Nell'implementazione in cui le beta sono calcolate senza l'ausilio di una tecnica Montecarlo generiamo semplicemente un dataset e lo usiamo per il training. Ci aspettiamo in questo caso risultati meno precisi di quelli che otterremo nell'implementazione successiva.

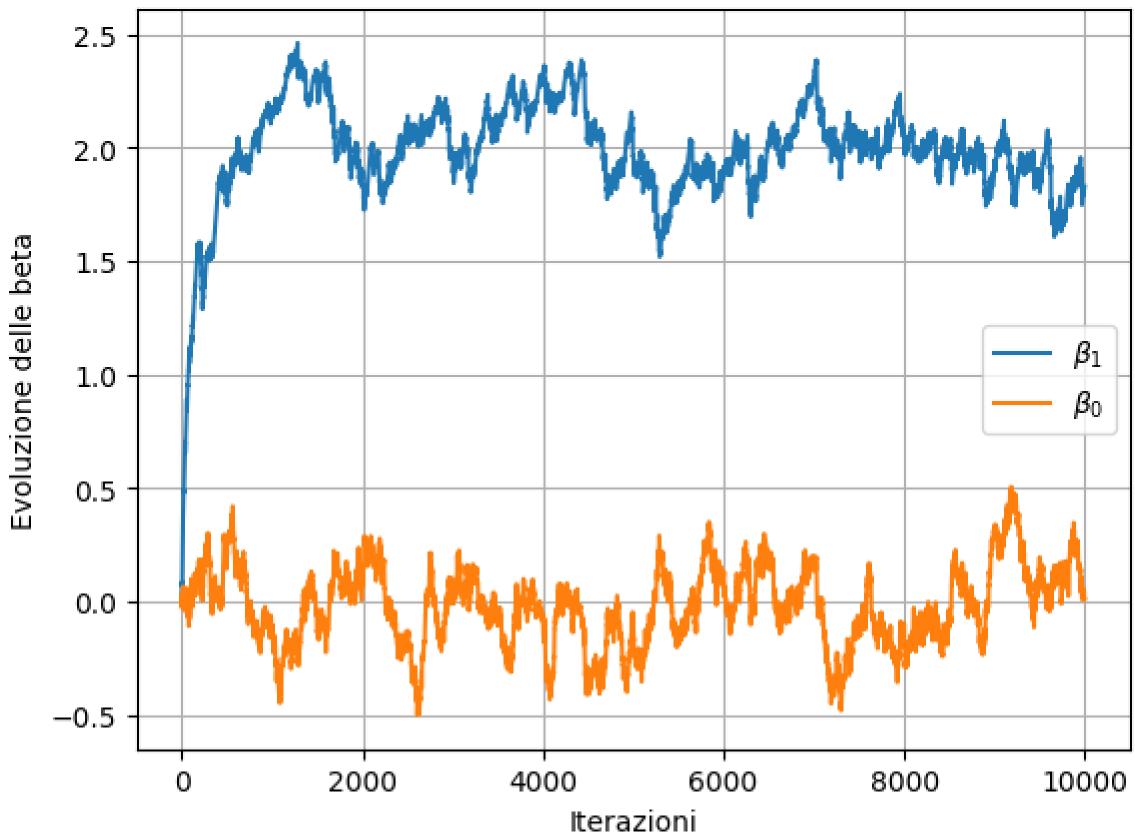

Figura 3: Evoluzione delle beta per ogni iterazione senza tecnica Montecarlo

Nell'implementazione successiva le beta sono calcolate per mezzo di una tecnica Montecarlo, ad ogni iterazione viene generato un dataset in accordo alle distribuzioni note e le beta risultanti sono le medie di quelle ottenute con le diverse realizzazioni del dataset.

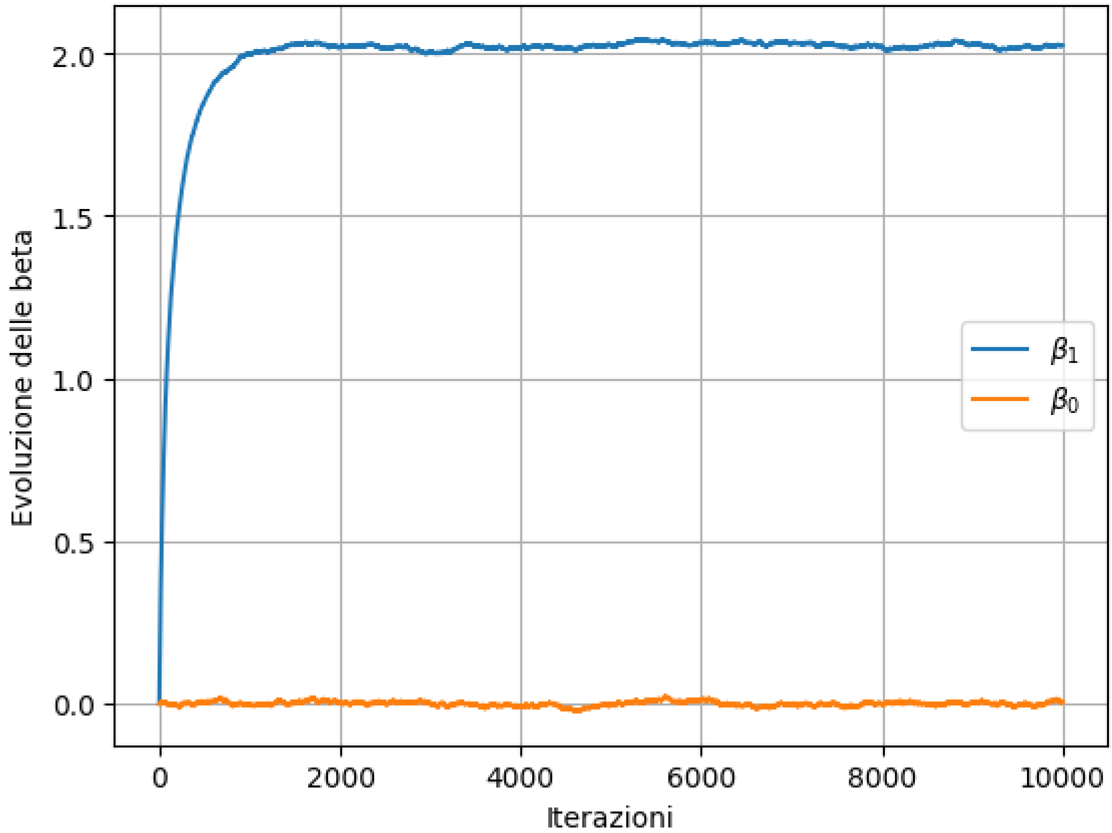

Figura 4: Evoluzione delle beta per ogni iterazione con tecnica Montecarlo

Notiamo dall'osservazioni dei grafici che mostrano l'evoluzione delle Beta che grazie alle iterazioni Montecarlo è stato possibile ottenere stime di beta molto meno rumorose.

2.2. Calcolo empirico delle prestazioni dei classificatori

È stato calcolato il test error utilizzando iterazioni Montecarlo per entrambi i classificatori ottenuti, in modo da assicurare una stima più precisa.
Di seguito sono messi a confronto i risultati ottenuti con i beta calcolati senza e con l'ausilio di iterazioni Montecarlo per il training.

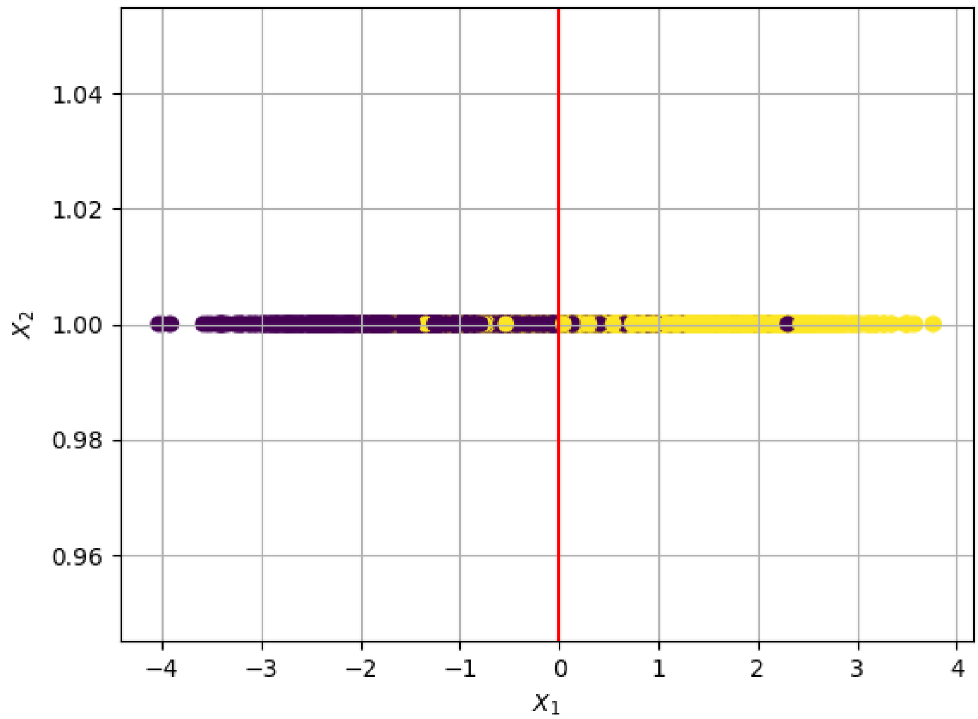

Figura 5: Scatterplot del modello ottenuto senza iterazioni Montecarlo

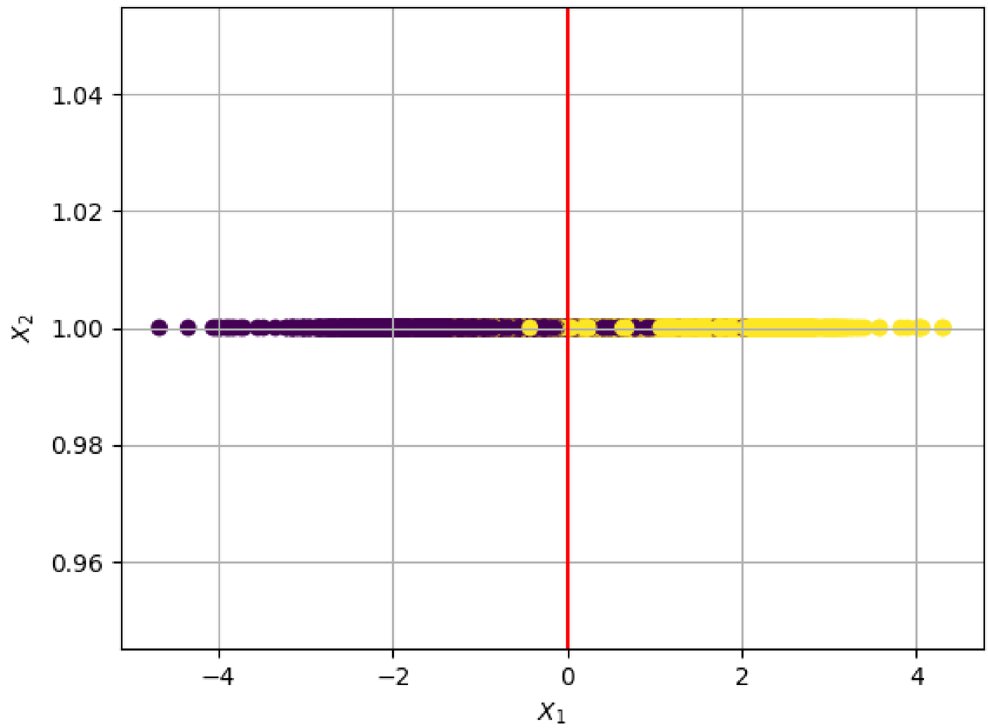

Figura 6: Scatterplot del modello ottenuto con iterazioni Montecarlo

Dagli scatter plot (Figura 5-6) non è possibile apprezzare una differenza tra i due approcci, questo sicuramente a causa della situazione data-rich che abbiamo creato. Resta tuttavia apprezzabile il fatto che le stime delle beta con l'ausilio di Montecarlo sono molto meno rumorose.

2.2.1. Confronti tra le probabilità d'errore relative ai diversi approcci

Tutti i seguenti confronti sono relativi, come richiesto dalla traccia, al caso con varianza più bassa, nello specifico è stata scelta varianza uguale ad 1.

Probabilità d'errore calcolata per il classificatore MAP con modello noto	0.15843
Probabilità d'errore calcolata per il classificatore ottenuto con SGD con constant step-size senza iterazioni Montecarlo	0.15866
Probabilità d'errore calcolata per il classificatore ottenuto con SGD con constant step-size con iterazioni Montecarlo	0.15847

Notiamo che, seppur di poco, la Probabilità d'errore ottenuta con MAP risulta più bassa di quelle ottenute con SGD, sicuramente il fatto che la differenza sia poca è dovuta all'alto numero di campioni che compongono i dataset utilizzati per SGD. In una situazione non "data-rich" la differenza sarebbe stata più pronunciata, MAP avvalendosi dei modelli è il criterio ottimo e privo dei compromessi che nascono dalla dipendenza da un dataset.

Notiamo inoltre che tra le due probabilità d'errore ottenute con SGD è più bassa quella risultante dall'utilizzo delle beta calcolate avvalendosi delle iterazioni Montecarlo, anche questo risultato è in linea con le aspettative. Mediando su più dataset si ottengono risultati meno affetti dalle specifiche caratteristiche che un certo dataset può avere e quindi migliori.

Indice delle figure

Figura 1: Calcolo pmf a posteriori 2
Figura 2: Grafico che mostra la posteriori nel caso in cui y=+1 3
Figura 3: Evoluzione delle beta per ogni iterazione senza tecnica Montecarlo 5
Figura 4: Evoluzione delle beta per ogni iterazione con tecnica Montecarlo 6
Figura 5: Scatterplot del modello ottenuto senza iterazioni Montecarlo 7
Figura 6: Scatterplot del modello ottenuto con iterazioni Montecarlo 7