STIMATORE

STATISTICA BAYESIANA

Uno stimatore Bayesiano (o regola di decisione B.) minimizza il valore atteso a posteriori della loss scelta

CHAIN RULE

$$\mathbb{E}[h(\bar{X}, \bar{Y})] = \mathbb{E}_{\bar{X}}[E[h(\bar{X}, \bar{Y}) \mid \bar{X}]]$$

MIGLIOR STIMATORE DI UNA VARIABILE ALEATORIA CON MSE

Data una V.A. $\bar{Y}$, vogliamo stimarla con un valore z.

Scegliamo come loss l'MSE

$$\ell(\bar{Y}, z) = (\bar{Y} - z)^2$$

troviamo z tale che

$$z^* = \underset{z \in Z}{\operatorname{argmin}} \; \mathbb{E}[(\bar{Y} - z)^2]$$

Dato che non ho gli appunti, troviamo il minimo di $\mathbb{E}[(\bar{Y} - z)^2]$

Si pone che $\frac{d}{dx}\mathbb{E}[f(x)] = \mathbb{E}\left[\frac{d}{dx}f(x)\right]$

quindi:

$$\frac{d}{dz}\mathbb{E}[(\bar{Y}-z)^2] = \mathbb{E}\left[\frac{d}{dz}(\bar{Y}-z)^2\right] =$$

$$\mathbb{E}[-2(\bar{Y}-z)] = 2\,\mathbb{E}[z-\bar{Y}]$$

per trovare il minimo, troviamo il punto stazionario dove la derivata è zero →

Sapendo che alla fine z è un numero fissato, lo portiamo fuori dal $\mathbb{E}$ e otteniamo →

$$\cancel{2}\,\mathbb{E}[z-\bar{Y}] = 0 \rightarrow \mathbb{E}[z] - \mathbb{E}[\bar{Y}] = 0$$

$$z - \mathbb{E}[\bar{Y}] = 0 \rightarrow z = \mathbb{E}[\bar{Y}]$$

Per essere sicuri che sia un punto di minimo, controlliamo che $\frac{d^2}{dx^2} > 0$ →

$$\frac{d}{dz}\mathbb{E}[2(z-\bar{Y})] = 2\,\mathbb{E}\left[\frac{d}{dz}(z-\bar{Y})\right] = 2 > 0 \checkmark$$

Quindi abbiamo dimostrato che la z che minimizza $\mathbb{E}[(\bar{Y}-z)^2]$ è proprio $\mathbb{E}[\bar{Y}]$, cioè

$$\boxed{z^* = \arg\min_{z \in Z} \mathbb{E}\left[(\bar{Y}-z)^2\right] = \mathbb{E}[\bar{Y}]}$$

MIGLIOR STIMATORE DI UNA FUNZIONE DI V.A O MIGLIORE FUNZIONE CHE STIMA UNA V.A.

Nel caso in cui ci sia

$\bar{Y} \sim g(\bar{X})$ vogliamo trovare

la g migliore tale che

$$g^* = \arg\min_{g \in G} \mathbb{E}\left[(\bar{Y} - g(\bar{X}))^2\right]$$

~~Analogamente a prima dobbiamo trovare il minimo di $\mathbb{E}$, quindi facciamo la derivata $\rightarrow$~~

Scansionato con CamScanner

Vogliamo applicare più o meno la regola vista prima, cioè

$$\arg\min_{t \in \mathbb{Z}} \mathbb{E}[(\bar{Y}-t)^2] = \mathbb{E}[\bar{Y}]$$

Per farlo però dobbiamo riscrivere l'espressione per rendere $g(\bar{X})$ costante in qualche modo.

Dalla chain rule:

$$\mathbb{E}[h(\bar{X},\bar{Y})] = \mathbb{E}_{\bar{X}}\left[\mathbb{E}[h(\bar{X},\bar{Y}) \mid \bar{X}]\right]$$

Se poniamo $h(\bar{X},\bar{Y}) = (\bar{Y} - g(\bar{X}))^2$

Possiamo riscrivere:

$$\mathbb{E}\left[(\bar{Y}-g(\bar{X}))^2\right] = \mathbb{E}_{\bar{X}}\left[\mathbb{E}[(\bar{Y}-g(\bar{X}))^2 \mid \bar{X}]\right]$$

Poiché stiamo sempre cercando l'argmin di $E_{\bar{x}}[E[(\bar{Y}-g(\bar{x}))^2|\bar{x}]]$, riflettiamo su come minimizzare il E esterno: se minimizziamo il valore all'interno minimizziamo anche quello all'esterno, cioè

$$\underset{g\in\mathcal{G}}{\text{argmin}}\, E_{\bar{x}}[E[(\bar{Y}-g(\bar{x}))^2|\bar{x}]] = \underset{g\in\mathcal{G}}{\text{argmin}}\, E[(\bar{Y}-g(\bar{x}))^2]$$

Quindi troviamo il minimo di $E[(\bar{Y}-g(\bar{x}))^2]$.

Poiché in questo contesto $\bar{x}$ è fissato ($|\bar{x}$), lo è anche $g(\bar{x})$, quindi $g(\bar{x}) = z \rightarrow$

$$\underset{g(\bar{x})\in\mathcal{G}}{\text{argmin}}\, E[(\bar{Y}-g(\bar{x}))^2] = E[\bar{Y}]$$

Quindi

$$\underset{g\in\mathcal{G}}{\text{argmin}}\, E[(\bar{Y}-g(\bar{x}))^2] = \underset{g\in\mathcal{G}}{\text{argmin}}\, E_{\bar{x}}[E[(\bar{Y}-g(\bar{x}))^2|\bar{x}]]$$

$$= \underline{E[\bar{Y}|\bar{x}]} = g^*(\bar{x}) = \hat{\bar{Y}}_{MMSE}$$

MINIMUM MEAN SQUARE ERROR

STIMATORI A MASSIMA VEROSIMIGLIANZA

Invece di minimizzare la loss come gli stimatori Bayesiani, gli stimatori a massima verosimiglianza **massimizzano** la funzione di verosimiglianza (wiki)

In questo caso abbiamo
- i campioni
- il modello

e vogliamo stimare i parametri del modello. Per fare ciò usiamo il criterio di massima verosimiglianza:

i valori preferiti di una funzione di verosimiglianza sono quelli che rendono massima la probabilità di ottenere i dati osservati

In particolare la funzione di verosimiglianza è il modello in cui i campioni sono noti e non si conoscono i parametri (quindi la funzione è in funzione dei parametri)

FUNZIONE DI VEROSIMIGLIANZA

1) Data una pdf $\mathcal{L}$ con un parametro θ (che vogliamo stimare) e N osservazioni $\{x_i\}_{i=1}^{N}$

2) Calcoliamo la probabilità di osservare i dati ottenuti:

$$\mathbb{P}\left(\{x_i\}_{i=1}^{N} \mid \theta\right) = \mathcal{L}\left(\theta \mid \{x_i\}_{i=1}^{N}\right)$$

Come enunciato nel criterio di massima verosimiglianza, l'obiettivo è trovare

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \, \mathcal{L}\left(\theta \mid \{x_i\}_{i=1}^{N}\right)$$

STIMA DELLA μ DI UNA GAUSSIANA

In questo caso la funzione di verosimiglianza è:

$$\ell(x,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}}$$

nel caso in cui abbiamo $\bar{X} = \{\bar{x}_1, \dots, \bar{x}_n\}$ indipendenti e identicamente distribuite (normali $\bar{x}_i \sim N(\theta, \sigma^2)$ la L sarà

$$\ell(X,\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\theta)^2}{2\sigma^2}}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \cdot e^{\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2}$$

$$\rightarrow \arg\max_{\theta} \ell(x,\theta) = \arg\min_{\theta} \sum (x_i-\theta)^2$$

$$\Rightarrow \frac{d}{d\theta} \sum (x_i-\theta)^2 = -2\sum_{i=1}^{n}(\theta - x_i)$$

$$= 2\sum_{i=1}^{n}(x_i-\theta) \rightarrow 2\sum_{i=1}^{n}(x_i-\theta) = 0$$

$$\cancel{2}\sum_{i=1}^{n} x_i - \cancel{2}n\theta = 0 \Rightarrow \hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Scansionato con CamScanner

ESERCIZIO - STIMATORE MLE

Data

- $\bar{Y} \sim N(0, \sigma_Y^2)$
- $W_i \sim N(0, \sigma_w^2)$

$\bar{X}_i = \bar{Y} - W_i$

Trovare lo stimatore bayesiamo di $\bar{Y}$

$\bar{Y} = \bar{X}_i - W_i$

~~Troviamo la distribuz~~

Sappiamo che il miglior stimatore

é $\hat{\bar{Y}} = \mathbb{E}[\bar{Y} \mid \bar{X}]$

Quindi

$\mathbb{E}[\bar{Y} \mid \bar{X}] = \mathbb{E}_{\bar{X}}\left[\mathbb{E}[(\bar{Y} - g(\bar{X}))^2 \mid \bar{X}]\right]$

dove $g(\bar{X}) = \bar{X} - W$

Oppure facciamo tutto da capo:

Vogliamo trovare uno stimatore

z di $\bar{Y}$: $z^* = \arg\min_{z \in ?} \mathbb{E}[(\bar{Y} - z)^2]$

Sappiamo che la migliore z é

$z^* = \mathbb{E}[\bar{Y}]$

in questo caso però, non abbiamo i valori di $\bar{Y}$, ma di $\bar{X}_i$

quindi sappiamo che la soluzione migliore è

$$g^* = \underset{g \in G}{\arg\min}\, \mathbb{E}\left[\left(\bar{Y} - (\bar{X}_i - W_i)\right)^2\right]$$

$$= \mathbb{E}[\bar{Y} \mid \bar{X}_i] = \mathbb{E}_{\bar{X}_i}\left[\mathbb{E}\left[(\bar{Y} - \bar{X}_i + W_i)^2 \mid X_i\right]\right]$$

Calcoliamo prima il valore atteso più interno:

$$\mathbb{E}\left[(\bar{Y} - \bar{X}_i + W_i)^2 \mid \bar{X}_i = x_i\right] = \quad !$$

attenzione: $\mathbb{E}[\bar{Y} \mid \bar{X}] \neq \mathbb{E}_{\bar{X}}\left[\mathbb{E}\left[(\bar{Y} - g(\bar{X}))^2 \mid \bar{X}\right]\right]$

ma è vero che

$$\mathbb{E}[\bar{Y} \mid \bar{X}] = \underset{g \in G}{\arg\min}\, \mathbb{E}_{\bar{X}}\left[\mathbb{E}\left[(\bar{Y} - g(\bar{X}))^2 \mid \bar{X}\right]\right]$$

Quindi

$$\mathbb{E}[\bar{Y} \mid \bar{X}_i] = f(\bar{x}_i)$$

Per calcolare questo valore atteso ci serve la distribuzione di $\bar{Y}$ →

$$\bar{Y} = \bar{X}_i - W_i =$$

$\bar{X}_i$ e w_i sono indipendenti e gaussiane quindi il valore atteso è la somma dei v.a. e la varianza è la somma delle varianze ⇒ quindi

$$\bar{X}_i - W_i \sim N(Y, 2\sigma_w^2)$$

??

Quindi

$$f_{\bar{Y}/\bar{X}_i}(Y \mid x_i) = \frac{1}{\sqrt{4\pi\sigma_w^2}} \cdot e^{\frac{(y-Y)^2}{8\sigma_w^2}}$$

Non riesco a trovarla in modo normale. Quindi facciamo un altro ~~ragiona~~ ragionamento:

Sappiamo che:

$$\bar{Y} \sim \frac{1}{2\pi\sigma_y^2} \cdot e^{-\frac{y^2}{2\sigma_y^2}}$$

$$W_i \sim \frac{1}{\sqrt{2\pi\sigma_w^2}} \cdot e^{-\frac{1}{2\sigma_w^2}}$$

$$\bar{X}_i \sim \frac{1}{\sqrt{2\pi\sigma_{x_i}^2}} \cdot e^{-\frac{(x_i-\mu_{x_i})^2}{2\sigma_w^2}}$$

NB ci ha chiesto ~~...~~ $E[\bar{Y} | \bar{X}]$ non $E[\bar{Y} | \bar{x}_i]$!!

Sappiamo che $E[\bar{X}_i] = y_i$ casuale

Per la legge dei grandi numeri se avessimo infinite $\bar{x}_i$!

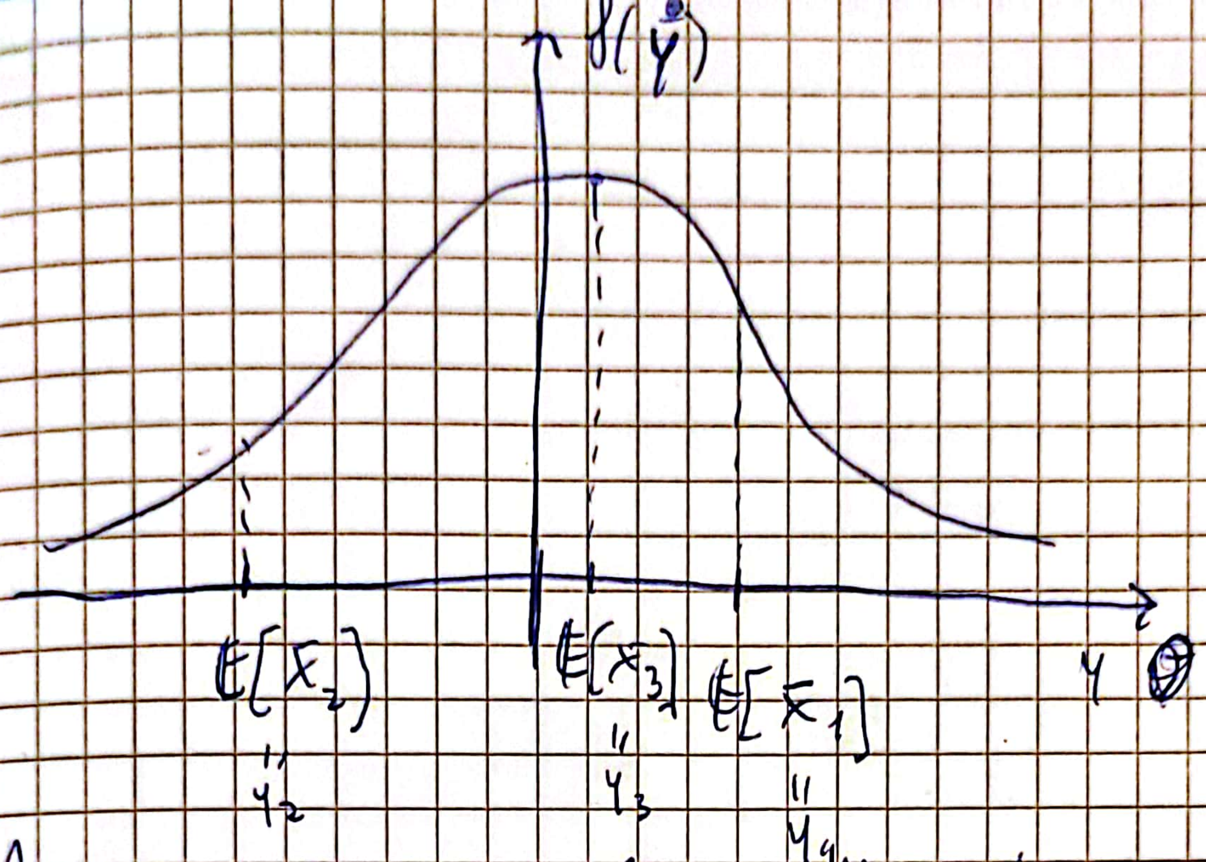
$$E[\bar{Y}] = E\left[\underbrace{E[\bar{X}_i]}_{\text{praticamente questa è una v.a.}}\right]$$

tuttavia per ipotesi $E[\bar{Y}] = 0$

quindi viene qualcosa del tipo:

~~$f(Y|E(\bar{x}_i)) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \cdot e^{-\frac{y - E[\bar{x}_i]}{...}}$~~

$$f(Y = E(\bar{x}_i)) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \cdot e^{-\frac{E[\bar{x}_i]}{2\sigma_y^2}}$$



Abbiamo quindi ottenuto

$f(y \mid \mathbb{E}[\bar{X}_i])$ (se così si può dire dato che $y = \mathbb{E}[\bar{X}_i]$)

Per la legge dei grandi numeri

$$\mathbb{E}(\mathbb{E}(\bar{X}_i)) \sim \frac{\sum \mathbb{E}[\bar{X}_i]}{n} \quad \text{per } i \to \infty$$

Infine se approssimiamo $\mathbb{E}[\bar{X}_i] = x_i$

Otteniamo che

$$\mathbb{E}[Y \mid \bar{X}] = \overbrace{\mathbb{E}[\mathbb{E}[\bar{X}_i] \mid \bar{X}_i]}^{?\ \text{SUi}} = \mathbb{E}[\bar{X}_i]$$

$$= \frac{\sum_i \bar{X}_i}{n} \text{ se } m = 1 \qquad \frac{\sum_j \sum_i x_{ij}}{m \cdot n} \text{ se } m > 1$$

Scansionato con CamScanner

SIMULAZIONI MONTE CARLO

STIMATORE BAYESIANO

- TEORIA

$$\hat{\theta}_B = \frac{\sigma_y^2}{N\sigma_y^2 + \sigma_w^2} \sum_{i=1}^{N} x_i$$

$$= \hat{\theta}_B = \frac{\sigma_y^2}{\sigma_y^2 + \frac{\sigma_w^2}{N}} \cdot \frac{\sum_{i=1}^{N} x_i}{N}$$

l'errore teorico sarà uguale a →

$$MSE = a^2 \frac{\sigma_w^2}{N} + (a-1)^2 \sigma_y^2 =$$

dove

$$a = \frac{\sigma_y^2}{\sigma_y^2 + \frac{\sigma_w^2}{N}} \begin{cases} = 1 \text{ per } \sigma_y^2 \to \infty \\ = 0 \text{ per } \sigma_w^2 = \infty \end{cases}$$

quindi

$$\lim_{\sigma_y^2 \to \infty} MSE = \cdot \frac{\sigma_w^2}{N} + (\quad -1)^2 \cdot \infty = \infty$$

$$\lim_{\sigma_w^2 \to \infty} MSE =$$

$$MSE = \frac{\sigma_y^4 \cdot \frac{\sigma_w^2}{N}}{\left(\sigma_y^2 + \frac{\sigma_w^2}{N}\right)^2} + \frac{\left(\frac{\sigma_w^2}{N}\right)^2 \cdot \sigma_y^2}{\left(\sigma_y^2 + \frac{\sigma_w^2}{N}\right)^2}$$

$$= \frac{\sigma_y^4 \cdot \frac{\sigma_w^2}{N} + \frac{\sigma_w^4}{N^2} \cdot \sigma_y^2}{\left(\sigma_y^2 + \frac{\sigma_w^2}{N}\right)^2}$$

$$= \frac{\left(\sigma_y^2 \cdot \frac{\sigma_w^2}{N}\right) \cancel{\left(\sigma_y^2 + \frac{\sigma_w^2}{N}\right)}}{\left(\sigma_y^2 + \frac{\sigma_w^2}{N}\right)^{\cancel{2}}}$$

$$= \boxed{\frac{\sigma_y^2 \cdot \frac{\sigma_w^2}{N}}{\sigma_y^2 + \frac{\sigma_w^2}{N}} = MSE}$$

QUINDI

dato l'errore MSE teorico

$$MSE = \frac{\sigma_y^2 \cdot \frac{\sigma_w^2}{N}}{\sigma_y^2 + \frac{\sigma_w^2}{N}}$$

ci aspettiamo che:

- per $\sigma_y^2 \to \infty \to \frac{\sigma_w^2}{N}$ ✓
- per $\sigma_w^2 \to \infty \to \sigma_y^2$ ✓
- per $N \to \infty \to 0$ ✓

Quindi mancano da plottare i grafici sui valori teorici