Esercizi

1) Prendo coppie di dati (x_1, y_1) (x_2, y_2) (x_3, y_3) (x_4, y_4)
$= (5, -2)$ $(-3, -2)$ $(-2, -1)$ $(0, 1)$

provengono con un rumore gaussiano ε :

$(\mu = 0)$

$$y_i = a x_i + b + \varepsilon_i \qquad \varepsilon_i \sim \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\varepsilon^2}{2\sigma^2}} = p_\theta(\varepsilon)$$

usando il principio di max verosomiglianza cerco il modello ideale

$$\rightarrow \varepsilon_i = y_i - a x_i - b$$

$$L = \prod_{i=1}^{4} p_\theta(\varepsilon_i)$$

ottenuto con $\exp^{-1}$ e somme di log

$$\ln L = \sum_{i=1}^{4} \left(-2\ln 2\pi - 2\ln \sigma^2 - \frac{1}{2\sigma^2} \varepsilon_i^2 \right) \rightarrow \sum_i \varepsilon_i^2$$

non dipendono da a e b

di ciò mi interessa

che viene derivato

((la costante prima sta fuori la derivata)

$$\frac{d}{da} \ln L = 0 \qquad \frac{d}{db} \ln L = 0$$

$$\frac{d}{da} \left(\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \varepsilon_4^2 \right) = 0$$

$$\frac{d}{da} \Big[(-2-5a-b)^2 + (-2+3a-b)^2 + (-1+2a-b)^2 + \underbrace{(1-b)^2}_{\text{non dipende da } a} \Big] = 0$$

$$\xrightarrow{da} 2(-2-5a-b)(-5) + 2(-2+3a-b)(3) + 2(-1+2a-b)\,2 = 0$$

$\rightarrow 20 + 50a + \cancel{10b} - 12 + 18a - 6b - 4 + \cancel{8a}$
$-\cancel{4b} = 0$ (se b non andasse via, allora confrontalo con l'eq ottenuta da db)

$a^* = -\frac{1}{18}$

qui non calcolata

$(a^*, b^*, \tilde{\sigma})$

sono: parametri che meglio descrivono i dati

se derivo rispetto b sapendo a ottengo $b^* = -1$

ottengo $f(x) = -\frac{1}{18}x - 1$

Ora posso utilizzare x diverse, es $f(13)$, $f(38)$ ...

Si può dire che le n coppie iniziali sono "il training test", da cui estrapolo la legge (modello) parametrica che meglio descrive i dati forniti. Ci si aspetta che l'output ottenuto segua la legge. Base del machine learning.

Il tipo rumore ha influenzato la costruzione del modello

2) Distribuzione esponenziale

$$f_\lambda(x) = \lambda e^{-\lambda x}$$

$D = \{x_i\}_{i=1}^{N}$
dati forniti

Vogliamo usare il p. di max verosomiglianza per fissare λ

$$L(\lambda) = \prod_{i=1}^{N} f_\lambda(x_i)$$

$$\ln L(\lambda) = \sum_{i=1}^{N} \ln f_\lambda(x_i)$$

$$= \sum_{i}^{N} (\ln \lambda - \lambda x)$$

$$\frac{d}{d\lambda} \ln L(\lambda) = \sum_{i=1}^{N} \left(\frac{1}{\lambda} - x \right) = \frac{N}{\lambda} - \sum_{i=1}^{N} x_i = 0$$

la derivata viene fatta sui termini della sommatoria

$$\rightarrow \quad \frac{1}{\lambda^*} = \frac{1}{N} \sum_{i} x_i$$ che posso invertire

corrisponde alla media

e' il valore di λ che nella distribuzione esponenziale descrive meglio i dati

3) $(x_1, \dots, x_n) \overset{\text{iid}}{\sim} p(X)$

estrazioni indipendenti

$E(X) = \mu$

$Var(X) = \sigma^2$

stimatore di σ:

$$\hat{\sigma}^2(X) = \frac{1}{n} \left(\sum_i X_i \right)^2 - n \cdot x_2 x_3$$

e' corretto o distorto?

Cioe' si osserva se stima bene σ

Calcolo il valore atteso:

$$\mathbb{E}_p \hat{\sigma}^2 = \frac{1}{n} \mathbb{E}_p \Big(\overbrace{\sum_i x_i}^{Y}\Big)^2 + \overbrace{\mathbb{E} x_2 \cdot \mathbb{E} x_3}^{\text{perché } x_2\ x_3 \text{ indip}}$$

preso dalla Var

$Var = E^2 - E \to E^2 = Var + E$ $\qquad \mu^2$

$$= \frac{1}{n} Var\Big(\sum x_i\Big) + \frac{1}{n} \Big(\mathbb{E}\Big[\sum_i x_i\Big]\Big)^2 - \mu^2$$

$$= \frac{1}{n} \sum_i Var(x_i) + \frac{1}{n}\Big(\sum_i \mathbb{E} x_i\Big)^2 - \mu^2$$

$$= \frac{1}{n} n \sigma^2 + \frac{1}{n}(n^2 \mu^2) - \mu^2 = \sigma^2 + \underbrace{(n-1)\mu^2}_{\text{e' distorto di } \bullet}$$

più semplice

$\to$ per correggere osservo $\bullet$, se fosse n otterrei 0 e quindi $= \sigma^2$ corretto