ΠΑΝΕΠΙΣΤΗΜΙΟΝ ΠΕΙΡΑΙΩΣ

ΕΠΩΝΥΜΟ: - ESTIMATION OF UNKNOWN

ΟΝΟΜΑ: PROBABILITY DENSITY

ΠΑΤΡΩΝΥΜΟ: FUNCTIONS

Α.Μ: ... ΕΞΑΜΗΝΟ: ...

ΜΑΘΗΜΑ: ΕΚΤΙΜΗΣΗ ΑΓΝΩΣΤΩΝ ΣΥΝΑΡΤΗΣΕΩΝ

ΗΜΕΡΟΜΗΝΙΑ: ΠΥΚΝΟΤΗΤΑΣ ΠΙΘΑΝΟΤΗΤΑΣ

#1

MAXIMUM LIKELIHOOD PARAMETER ESTIMATION:

Εκτίμηση Παραμέτρων μέσω Μέγιστης Πιθανοφάνειας

Θεωρούμε ένα M-ταξικό πρόβλημα ταξινόμησης όπου τα διανύσματα χαρακτηριστικών κατανέμονται σύμφωνα με τις εξής συναρτήσεις πυκνότητας πιθανότητας (class-conditional probability density functions):

$$\boxed{p(\underline{x}|\omega_i),\ \forall i \in [M]} \quad (*)$$

Υποθέτουμε πως οι παραπάνω συναρτήσεις πιθανοφάνειας δίνονται σε παραμετρική μορφή και οι αντίστοιχες παράμετροι για κάθε κλάση σχηματίζουν τα διανύσματα $\underline{\theta_i}$ τα οποία είναι άγνωστα.

Στόχος μας είναι η εκτίμηση αυτών των άγνωστων διανυσμάτων παραμέτρων κάνοντας χρήση ενός γνωστού συνόλου διανυσμάτων χαρακτηριστικών από κάθε κλάση. Επιπλέον, υποθέτοντας πως τα δεδομένα της μίας κλάσης δεν επηρεάζουν την διαδικασία εκτίμησης των παραμέτρων της άλλης, μπορούμε να διατυπώσουμε το πρόβλημα ανεξάρτητα από τις κλάσεις απλοποιώντας την μαθηματική σήμανση. Κατά συνέπεια το υποκείμενο μαθηματικό πρόβλημα δύναται να επιλυθεί για την κάθε κλάση ανεξάρτητα από την άλλη.

(*) Συνήθως, για να δηλώσουμε την εξάρτηση των συναρτήσεων πιθανοφάνειας από τα αντίστοιχα διανύσματα παραμέτρων $\underline{\theta_i}$ γράφουμε ότι:

$$\boxed{p(\underline{x}|\omega_i) = p(\underline{x}|\omega_i;\underline{\theta_i}),\ \forall i \in [M]}$$

Έστω $X = \{ \underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N \}$ ένα σύνολο τυχαίων δειγμάτων τα οποία έχουν ληφθεί από την συνάρτηση πυκνότητας πιθανότητας $p(\underline{x};\underline{\theta})$. Υποθέτοντας στατιστική ανεξαρτησία μεταξύ των δειγμάτων σχηματίζουμε την από κοινού συνάρτηση πυκνότητας πιθανότητας ως εξής:

$$\boxed{p(X;\underline{\theta}) \equiv p(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N ; \underline{\theta}) = \prod_{k=1}^{N} p(\underline{x}_k ; \underline{\theta})}$$

Η παραπάνω σχέση δηλώνει μια συνάρτηση του $\underline{\theta}$ που ονομάζεται συνάρτηση πιθανοφάνειας του $\underline{\theta}$ σε σχέση με το σύνολο X. Η μέθοδος της μέγιστης πιθανοφάνειας εκτιμά την τιμή του $\underline{\theta}$ έτσι ώστε να μεγιστοποιείται η τιμή της συνάρτησης πιθανοφάνειας.

$$\boxed{\hat{\underline{\theta}}_{ML} = \arg\max_{\underline{\theta}} \prod_{k=1}^{N} p(\underline{x}_k ; \underline{\theta})}$$

Μια αναγκαία συνθήκη που θα πρέπει να ικανοποιείται ώστε το $\hat{\underline{\theta}}_{ML}$ να μεγιστοποιεί την συνάρτηση πιθανοφάνειας είναι λεγόμενη συνθήκη πρώτης τάξης (First Order Conditions or FOCs):

$$\boxed{\frac{\partial \prod_{k=1}^{N} p(\underline{x}_k ; \underline{\theta})}{\partial \underline{\theta}} = \mathbf{0}}$$

Προκειμένου να εκμεταλλευθούμε την μονοτονία της λογαριθμικής συνάρτησης, ορίζουμε την λογαριθμική συνάρτηση πιθανοφάνειας ως εξής: [log-likelihood function]

$$\boxed{L(\underline{\theta}) \equiv \ln \prod_{k=1}^{N} p(\underline{x}_k ; \underline{\theta})}$$

Εναλλακτικά, η λογαριθμική συνάρτηση πιθανοφάνειας μπορεί να οριστεί ως εξής:

$$\boxed{L(\underline{\theta}) = \sum_{k=1}^{N} \ln(p(\underline{x}_k ; \theta))}$$

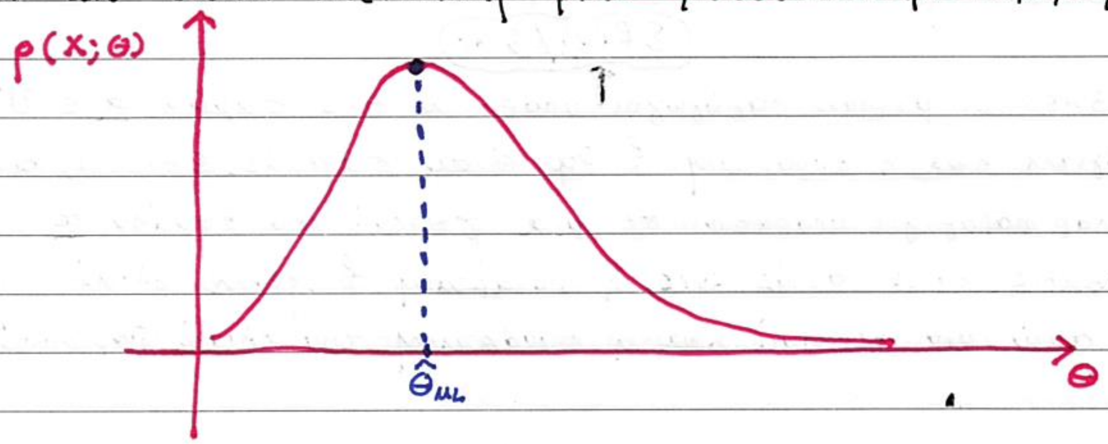
Συνεπώς, η συνθήκη πρώτης τάξης μπορεί να γραφεί ισοδύναμα ως εξής:

$$\frac{\partial L(\underline{\theta})}{\partial \underline{\theta}} = \frac{\partial}{\partial \underline{\theta}} \left\{ \sum_{k=1}^{N} \ln(p(\underline{x}_k ; \underline{\theta})) \right\} = \underline{0} \iff$$

$$\frac{\partial L(\underline{\theta})}{\partial \underline{\theta}} = \sum_{k=1}^{N} \frac{\partial \ln(p(\underline{x}_k ; \underline{\theta}))}{\partial \underline{\theta}} = \underline{0} \xLeftrightarrow{\ln(f(x))' = \frac{f'(x)}{f(x)}}$$

$$\boxed{\frac{\partial L(\underline{\theta})}{\partial \underline{\theta}} = \sum_{k=1}^{N} \frac{1}{p(\underline{x}_k ; \underline{\theta})} \cdot \frac{\partial}{\partial \underline{\theta}} p(\underline{x}_k ; \underline{\theta}) = \underline{0}}$$

Το παρακάτω σχήμα παρουσιάζει την μέθοδο της μέγιστης πιθανοφάνειας για την ειδική περίπτωση μιας μονοδιάστατης παραμέτρου θ.



Η εκτίμηση μέγιστης πιθανοφάνειας $\hat{\theta}_{ML}$ για την παράμετρο θ αντιστοιχεί στην τιμή που μεγιστοποιεί την αντίστοιχη συνάρτηση.

ΣΤΟΙΧΕΙΑ ΠΟΛΥΜΕΤΑΒΛΗΤΗΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ

Έστω $f: U \to \mathbb{R}$ όπου U είναι ένα ανοικτό υποσύνολο του διανυσματικού χώρου $\mathbb{R}^n$.

(I): [ΑΝΟΙΚΤΟ ΣΥΝΟΛΟ: OPEN SET]

Η έννοια του ανοικτού συνόλου στον Ευκλείδιο διανυσματικό χώρο ορίζεται ως εξής: Το $U \subseteq \mathbb{R}^n$ είναι ένα ανοικτό υποσύνολο του n-διάστατου ευκλείδιου διανυσματικού χώρου όταν:

$$\boxed{\forall \underline{x} \in U,\ \exists\, \varepsilon > 0 : \forall \underline{y} \in \mathbb{R}^n,\ \|\underline{x}-\underline{y}\| < \varepsilon,\ \underline{y} \in U.}$$

Ισοδύναμα, μπορούμε να γράψουμε ότι το U είναι ένα ανοικτό υποσύνολο του $\mathbb{R}^n$ όταν $\forall \underline{x} \in U,\ \exists\, \varepsilon > 0 : \Gamma_\varepsilon(x) \in U$.

[ΜΕΡΙΚΗ ΠΑΡΑΓΩΓΟΣ: PARTIAL DERIVATIVE]

(II): Η μερική παράγωγος μιας πολυμεταβλητής συνάρτησης $f: U \to \mathbb{R}$ σε ένα σημείο $\underline{a} = (a_1, \dots, a_n) \in U$ σε σχέση με την i-οστή μεταβλητή x_i ορίζεται σύμφωνα με το παρακάτω όριο:

$$\boxed{\frac{\partial f(\underline{a})}{\partial x_i} = \lim_{h \to 0} \frac{f(a_1, \dots, a_{i-1}, a_i + h, a_{i+1}, \dots, a_n) - f(a_1, \dots, a_n)}{h}}$$

$\partial f(\underline{a}) / \partial x_i$

Ακόμα και αν όλες οι μερικές παράγωγοι υπάρχουν στο σημείο $\underline{a} \in U$ αυτό δεν συνεπάγεται πως η συνάρτηση f θα είναι συνεχής. Ωστόσο, αν όλες οι μερικές παράγωγοι υπάρχουν σε μία γειτονιά του σημείου $\underline{a}$ και είναι συνεχείς εντός αυτής τότε η συνάρτηση f είναι ολικά διαφορίσιμη σε αυτή την γειτονιά και η συνάρτηση του ολικού διαφορικού είναι συνεχής.

Μάλιστα σύμφωνα με το Θεώρημα Clairaut ισχύει ότι:

$$\boxed{\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}}$$

ΠΑΝΕΠΙΣΤΗΜΙΟΝ ΠΕΙΡΑΙΩΣ

ΕΠΩΝΥΜΟ: ....
ΟΝΟΜΑ: ....
ΠΑΤΡΩΝΥΜΟ: ....
Α.Μ: .... ΕΞΑΜΗΝΟ: ....
ΜΑΘΗΜΑ: ....
ΗΜΕΡΟΜΗΝΙΑ: ....

[#5]

[ΟΛΙΚΗ ΠΑΡΑΓΩΓΟΣ : TOTAL DERIVATIVE]

(ΙΙΙ):

Αν $U \subseteq \mathbb{R}^n$ είναι ένα ανοικτό σύνολο και $f: U \to \mathbb{R}^m$, η συνάρτηση f είναι ολικά διαφορίσιμη σε ένα σημείο $\underline{a} \in U$ εάν υπάρχει ένας γραμμικός μετασχηματισμός $df_a: \mathbb{R}^n \to \mathbb{R}^m$ τέτοιος ώστε:

$$\boxed{\lim_{x \to a} \frac{\| f(x) - f(a) - df_a(x-a) \|}{\| x - a \|} = 0}$$

Η γραμμική απεικόνιση df_a καλείται ολική παράγωγος ή ολικό διαφορικό της f στο σημείο $\underline{a}$. Το ολικό διαφορικό της συνάρτησης f στο $\underline{a}$ συμβολίζεται και ως: $D_a f$ ή $Df(a)$.

Μια συνάρτηση f είναι ολικά διαφορίσιμη εάν το ολικό διαφορικό της υπάρχει σε κάθε σημείο του πεδίου ορισμού της.

Εννοιολογικά, ο ορισμός της ολικής παραγώγου εκφράζει την ιδέα ότι η ποσότητα df_a αποτελεί την καλύτερη δυνατή γραμμική προσέγγιση της f στο σημείο $\underline{a}$. Η προηγούμενη διατύπωση μπορεί να γίνει ακριβέστερη ποσοτικοποιώντας το σφάλμα της γραμμικής προσέγγισης της f στο $\underline{a}$ ως εξής:

$\boxed{f(a+h) = f(a) + df_a(h) + \varepsilon(h)}$ όπου $\varepsilon(h)$ το αντίστοιχο σφάλμα.

Συνεπώς, η διατύπωση πως η παράγωγος της f στο $\underline{a}$ είναι ίση με df_a είναι ισοδύναμη με την δήλωση: $\boxed{\varepsilon(h) = o(\|h\|)}$.

Αυτό σημαίνει πως $\boxed{\varepsilon(h) \ll \|h\|}$ καθώς $\boxed{h \to 0}$.

Έστω το πρόβλημα μεγιστοποίησης: (χωρίς περιορισμούς)

$\max_{\underline{x}} f(x)$: Το σημείο $\underline{x}^*$ είναι μέγιστο της συνάρτησης f όταν:

Οι συνθήκες πρώτης τάξης (FOCs) για το παραπάνω πρόβλημα εκφράζονται ως:

$$\boxed{\nabla_{\underline{x}} f(\underline{x}^*) = \underline{0}} \quad \Leftrightarrow \quad \boxed{\frac{\partial f(\underline{x}^*)}{\partial x_i} = 0, \ \forall i \in [n]}$$

Εναλλακτικά, οι συνθήκες πρώτης τάξης μπορούν να εκφραστούν μέσω του διανύσματος Βαθμίδας ως εξής: (gradient vector)

$$\boxed{Df(\underline{x}^*) = \left[\frac{\partial f(\underline{x}^*)}{\partial x_1}, \dots, \frac{\partial f(\underline{x}^*)}{\partial x_n}\right] = \underline{0}}$$

Οι συνθήκες δεύτερης τάξης (SOCs) για το παραπάνω πρόβλημα εκφράζονται μέσω της Χεσσιανής μήτρας (Hessian matrix)

$$\boxed{\underline{\underline{H}} \in M_{n \times n} \text{ με } H = [h_{ij}]} \ \text{και} \ \boxed{h_{ij}^{(\underline{x})} = \frac{\partial^2 f(\underline{x})}{\partial x_i \partial x_j}}$$

Το σημείο $\underline{x}^*$ είναι μέγιστο της συνάρτησης f όταν η Χεσσιανή μήτρα είναι αρνητικά ημιορισμένη, δηλαδή: (negative semi-definite)

για κάθε διάνυσμα $\boxed{\underline{h} \in \mathbb{R}^n : \ \underline{h}^T \underline{\underline{H}}\, \underline{h} \leq 0}$

⊛ Οι συνθήκες δεύτερης τάξης για την περίπτωση ενός προβλήματος ελαχιστοποίησης απαιτούν την Χεσσιανή μήτρα να είναι θετικά ημιορισμένη.

Ο εκτιμητής μέγιστης πιθανοφάνειας (ML estimate) διαθέτει ορισμένες εξαιρετικά επιθυμητές ιδιότητες. Αν $\hat{\underline{\theta}}_0$ είναι η πραγματική τιμή της άγνωστης παραμέτρου στην συνάρτηση πιθανοφάνειας $p(\underline{x};\underline{\theta})$ τότε ισχύουν τα παρακάτω:

(I): Ο εκτιμητής μέγιστης πιθανοφάνειας είναι ασυμπτωτικά αμερόληπτος (asymptotically unbiased) που εξ' ορισμού σημαίνει ότι:

$$\boxed{\lim_{N \to \infty} E[\hat{\theta}_{ML}] = \hat{\theta}_0}$$

Δηλαδή, η μέση τιμή του εκτιμητή συγκλίνει προς την πραγματική τιμή της παραμέτρου. Στην πραγματικότητα, η εκτίμηση μέγιστης πιθανοφάνειας $\hat{\theta}_{ML}$ αποτελεί και αυτή μια τυχαία μεταβλητή (τυχαίο διάνυσμα) καθώς εξαρτάται κάθε φορά από το δείγμα $\mathcal{X}$. Ένας εκτιμητής είναι αμερόληπτος όταν η μέση τιμή του ταυτίζεται με την πραγματική τιμή της άγνωστης παραμέτρου. Στην περίπτωση του εκτιμητή μέγιστης πιθανοφάνειας αυτό ισχύει ασυμπτωτικά καθώς $N \to \infty$.

(II): Ο εκτιμητής μέγιστης πιθανοφάνειας είναι ασυμπτωτικά συνεπής (asymptotically consistent) που σημαίνει πως:

$$\boxed{\lim_{N \to \infty} Pr\{\|\hat{\theta}_{ML} - \hat{\theta}_0\| < \varepsilon\} = 1}$$

όπου η τιμή του ε μπορεί να γίνει αυθαίρετα μικρή. Με άλλα λόγια, η τιμή του εκτιμητή μέγιστης πιθανοφάνειας συγκλίνει σε πιθανότητα προς την πραγματική τιμή της άγνωστης παραμέτρου. Εναλλακτικά, μπορούμε να πούμε πως για μεγάλες τιμές του N η εκτιμώμενη τιμή θα βρίσκεται αυθαίρετα κοντά στην πραγματική τιμή.

$$\boxed{\lim_{N \to \infty} E[\|\hat{\theta}_{ML} - \hat{\theta}_0\|^2] = 0}$$

Άσκηση 1η: Υποθέτουμε πως N σημεία δεδομένων έχουν παραχθεί από ανεξάρτητη, πανομοιότυπη δειγματοληψία (i.i.d) από μία (μονοδιάστατη) κανονική κατανομή με γνωστή την τιμή του μέσου (μ) και άγνωστης διακύμανσης (σ^2). Να ευρεθεί η εκτίμηση μέγιστης πιθανοφάνειας για την παράμετρο σ^2.

Η λογαριθμική συνάρτηση πιθανοφάνειας θα δίνεται από την σχέση:

$$L(\sigma^2) = \ln\left\{\prod_{k=1}^{N} p(x_k;\sigma^2)\right\} = \ln\left\{\prod_{k=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_k-\mu)^2}{2\sigma^2}\right\}\right\} \Leftrightarrow$$

$$L(\sigma^2) = \sum_{k=1}^{N} \ln\left\{\frac{1}{\sqrt{2\pi}\,\sigma} * \exp\left\{-\frac{(x_k-\mu)^2}{2\sigma^2}\right\}\right\} \Leftrightarrow$$

$$L(\sigma^2) = \sum_{k=1}^{N} \ln\frac{1}{\sqrt{2\pi}\,\sigma} + \ln\left\{\exp\left\{-\frac{(x_k-\mu)^2}{2\sigma^2}\right\}\right\} \Leftrightarrow$$

$$L(\sigma^2) = \sum_{k=1}^{N} \ln\frac{1}{\sqrt{2\pi}\sqrt{\sigma^2}} + \sum_{k=1}^{N} -\frac{(x_k-\mu)^2}{2\sigma^2} \Leftrightarrow$$

$$L(\sigma^2) = \sum_{k=1}^{N} -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=1}^{N}(x_k-\mu)^2 \Leftrightarrow$$

$$\boxed{L(\sigma^2) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=1}^{N}(x_k-\mu)^2}$$

Υπολογίζουμε την συνθήκη πρώτης τάξης (FOC):

$$\boxed{\frac{\partial L(\sigma^2)}{\partial \sigma^2} = 0}$$

ΠΑΝΕΠΙΣΤΗΜΙΟΝ ΠΕΙΡΑΙΩΣ

ΕΠΩΝΥΜΟ: ....

ΟΝΟΜΑ: ....

ΠΑΤΡΩΝΥΜΟ: ....

Α.Μ: .... ΕΞΑΜΗΝΟ: ....

ΜΑΘΗΜΑ: ....

ΗΜΕΡΟΜΗΝΙΑ: ....

#9

(*) $$\frac{\partial L(\sigma^2)}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\left[-\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=1}^{N}(x_k-\mu)^2\right] \Leftrightarrow$$

$$\frac{\partial L(\sigma^2)}{\partial \sigma^2} = -\frac{N}{2}\cdot\frac{\partial}{\partial \sigma^2}\left[\ln(2\pi\sigma^2)\right] - \frac{1}{2}\sum_{k=1}^{N}(x_k-\mu)^2 \cdot \frac{\partial}{\partial \sigma^2}\left[\frac{1}{\sigma^2}\right] \Leftrightarrow$$

Mind that we treat σ^2 as a single variable. That is, $\sigma^2 = a$.

$$\frac{\partial L(\sigma^2)}{\partial \sigma^2} = -\frac{N}{2}\cdot\frac{\partial/\partial\sigma^2(2\pi\sigma^2)}{2\pi\sigma^2} + \frac{1}{2}\sum_{k=1}^{N}(x_k-\mu)^2\cdot\frac{1}{\sigma^4} \Leftrightarrow$$

$$\frac{\partial L(\sigma^2)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{k=1}^{N}(x_k-\mu)^2 = 0 \Leftrightarrow$$

$$-N\sigma^2 + \sum_{k=1}^{N}(x_k-\mu)^2 = 0 \Leftrightarrow \boxed{\hat{\sigma}^2_{ML} = \frac{1}{N}\sum_{k=1}^{N}(x_k-\mu)^2}$$

#10

ΧΡΗΣΙΜΕΣ ΠΑΡΑΓΩΓΟΙ ΜΗΤΡΩΝ ΚΑΙ ΔΙΑΝΥΣΜΑΤΩΝ

DERIVATIVES OF DETERMINANTS:

(1): $$\frac{\partial |X|}{\partial X} = |X|(X^{-1})^T$$ ★ SOS

(2): $$\frac{\partial |AXB|}{\partial X} = |AXB|(X^{-1})^T = |AXB|(X^T)^{-1}$$

(3): $$\frac{\partial \ln|X|}{\partial X} = \frac{\partial |X| / \partial X}{|X|} = \frac{1}{|X|} \cdot \frac{\partial |X|}{\partial X} = \frac{1}{|X|} \cdot |X| \cdot (X^{-1})^T \Rightarrow$$

$$\frac{\partial \ln|X|}{\partial X} = (X^{-1})^T$$ SOS ★

DERIVATIVES OF INVERSES:

(4) $$\frac{\partial Y^{-1}}{\partial x} = -Y^{-1}\frac{\partial Y}{\partial x}Y^{-1}$$

$$\frac{\partial a^T X^{-1} b}{\partial X} = -X^{-T} a b^T X^{-T}$$ SOS ★ (5)

$$\frac{\partial |X^{-1}|}{\partial X} = -|X^{-1}|(X^{-1})^T$$ (6)

First Order:

(7) $$\frac{\partial x^T a}{\partial x} = \frac{\partial a^T x}{\partial x} = a$$

(8) $$\frac{\partial a^T X b}{\partial X} = a b^T$$ (9) $$\frac{\partial a^T X^T b}{\partial X} = b a^T$$

(10): $$\frac{\partial a^T X a}{\partial X} = \frac{\partial a^T X^T a}{\partial X} = a a^T$$

Second Order:

(11): $$\frac{\partial x^T B x}{\partial x} = (B + B^T)x$$

Άσκηση 24: Έστω $X = \{ \underline{x}_1, \dots, \underline{x}_N \}$ διανύσματα τα οποία έχουν προκύψει από δειγματοληψία μιας πολυδιάστατης κανονικής κατανομής με γνωστό μήτρα συνδιακύμανσης $\underline{\Sigma}$ και άγνωστο μέσο $\underline{\mu}$. Να ευρεθεί η εκτίμηση μέγιστης πιθανοφάνειας για την παράμετρο $\underline{\mu}$.

Η λογαριθμική συνάρτηση πιθανοφάνειας θα δίνεται από την σχέση:

$$L(\underline{\mu}) = \ln\left\{ \prod_{k=1}^{N} p(\underline{x}_k ; \underline{\mu}) \right\} = \ln\left\{ \prod_{k=1}^{N} \frac{1}{(2\pi)^{\ell/2} |\Sigma|^{1/2}} * \exp\left[-\frac{1}{2} (\underline{x}_k - \underline{\mu})^T \underline{\Sigma}^{-1} (\underline{x}_k - \underline{\mu}) \right] \right\} \Leftrightarrow$$

$$L(\underline{\mu}) = \sum_{k=1}^{N} \ln\left\{ \frac{1}{(2\pi)^{\ell/2} |\Sigma|^{1/2}} * \exp\left[-\frac{1}{2} (x_k - \mu)^T \Sigma^{-1} (x_k - \mu) \right] \right\} \Leftrightarrow$$

$$L(\underline{\mu}) = \sum_{k=1}^{N} \ln\left[\frac{1}{(2\pi)^{\ell/2} |\Sigma|^{1/2}} \right] - \frac{1}{2} \sum_{k=1}^{N} (x_k - \mu)^T \Sigma^{-1} (x_k - \mu) \Leftrightarrow$$

$$L(\underline{\mu}) = \sum_{k=1}^{N} -\frac{1}{2} \ln\left((2\pi)^{\ell} |\Sigma| \right) - \frac{1}{2} \sum_{k=1}^{N} (x_k - \mu)^T \Sigma^{-1} (x_k - \mu) \Leftrightarrow$$

$$\boxed{L(\underline{\mu}) = -\frac{N}{2} \ln\left((2\pi)^{\ell} |\Sigma| \right) - \frac{1}{2} \sum_{k=1}^{N} x_k^T \Sigma^{-1} x_k - 2\mu^T \Sigma^{-1} x_k + \mu^T \Sigma^{-1} \mu}$$

Υπολογίζοντας την συνθήκη πρώτης τάξης έχουμε: (FOC):

$$\boxed{\frac{\partial L(\underline{\mu})}{\partial \underline{\mu}} = \underline{0}}$$

$$\frac{\partial L(\underline{\mu})}{\partial \underline{\mu}} = \underbrace{\frac{\partial}{\partial \underline{\mu}} \left[-\frac{N}{2} \ln\left((2\pi)^{\ell} |\Sigma| \right) \right]}_{\to 0} - \frac{1}{2} \sum_{k=1}^{N} \frac{\partial}{\partial \underline{\mu}} \left[x_k^T \Sigma^{-1} x_k - 2\mu^T \Sigma^{-1} x_k + \mu^T \Sigma^{-1} \mu \right] \Leftrightarrow$$

$$\frac{\partial L(\underline{\mu})}{\partial \underline{\mu}} = -\frac{1}{2} \sum_{k=1}^{N} \underbrace{\frac{\partial}{\partial \mu} \left[x_k^T \Sigma^{-1} x_k \right]}_{\to 0} - 2 \frac{\partial}{\partial \mu} \left[\mu^T \Sigma^{-1} x_k \right] + \frac{\partial}{\partial \mu} \left[\mu^T \Sigma^{-1} \mu \right] \Leftrightarrow$$

$$\frac{\partial L(\mu)}{\partial \mu} = -\frac{1}{2} * \left\{ \sum_{k=1}^{N} \left[-2\Sigma^{-1} x_k + (\Sigma^{-1} + (\Sigma^{-1})^T)\mu \right] \right\} \Leftrightarrow$$

$$\frac{\partial L(\mu)}{\partial \mu} = -\frac{1}{2} * \sum_{k=1}^{N} -2\Sigma^{-1} x_k + 2\Sigma^{-1}\mu \Leftrightarrow$$

$$\frac{\partial L(\mu)}{\partial \mu} = -\frac{1}{2} \cdot 2 * \sum_{k=1}^{N} -\Sigma^{-1} x_k + \Sigma^{-1}\mu \Leftrightarrow$$

$$\frac{\partial L(\mu)}{\partial \mu} = + \sum_{k=1}^{N} \left\{ \Sigma^{-1} \cdot x_k - \Sigma^{-1} \cdot \mu \right\} = 0 \Leftrightarrow$$

$$\sum_{k=1}^{N} \Sigma^{-1} * (x_k - \mu) = 0 \Leftrightarrow$$

Here we multiply by Σ both sides of the equation.

$$\Sigma^{-1} * \sum_{k=1}^{N} (x_k - \mu) = 0 \Leftrightarrow$$

$$\sum_{k=1}^{N} (x_k - \mu) = 0 \Leftrightarrow \sum_{k=1}^{N} x_k - N\mu = 0 \Leftrightarrow$$

$$\boxed{\hat{\mu}_{ML} = \frac{1}{N} \sum_{k=1}^{N} x_k}$$

Άσκηση : #1

Έστω $\mathcal{X} = \{ \underline{x}_1, \underline{x}_2, \dots, x_N \}$ ένα σύνολο διανυσμάτων χαρακτηριστικών τα οποία έχουν προκύψει από δειγματοληψία μιας πολυδιάστατης κανονικής κατανομής με άγνωστο μέσο $\underline{\mu}$ και πίνακα συνδιακύμανσης $\underline{\underline{\Sigma}}$. Να ευρεθούν οι εκτιμήσεις μέγιστης πιθανοφάνειας για τις παραμέτρους, $\underline{\hat{\mu}}_{ML}$ και $\underline{\underline{\hat{\Sigma}}}_{ML}$.

Λαμβάνοντας υπόψιν το γεγονός πως η συνάρτηση πυκνότητας πιθανότητας που συνδέεται με την πιθανότητα παρατήρησης του κάθε δείγματος δίνεται από την σχέση:

$$\boxed{p(\underline{x}_k; \underline{\mu}, \underline{\underline{\Sigma}}) = \frac{1}{(2\pi)^{\ell/2}|\Sigma|^{1/2}} * \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}}$$

η αντίστοιχη λογαριθμική συνάρτηση πιθανοφάνειας θα δίνεται από την σχέση:

$$L(\underline{\mu}, \underline{\underline{\Sigma}}) = \ln\left\{\prod_{k=1}^{N} p(\underline{x}_k; \underline{\mu}, \underline{\underline{\Sigma}})\right\} = \sum_{k=1}^{N} \ln(p(\underline{x}_k; \underline{\mu}, \underline{\underline{\Sigma}})) \Leftrightarrow$$

$$L(\underline{\mu}, \underline{\underline{\Sigma}}) = \sum_{k=1}^{N} \ln\left\{\frac{1}{[(2\pi)^{\ell}|\Sigma|]^{1/2}} * \exp\left\{-\frac{1}{2}(x_k-\mu)^T\Sigma^{-1}(x_k-\mu)\right\}\right\} \Leftrightarrow$$

$$L(\underline{\mu}, \underline{\underline{\Sigma}}) = \sum_{k=1}^{N} -\frac{1}{2}\ln[(2\pi)^{\ell}|\Sigma|] + \sum_{k=1}^{N} -\frac{1}{2}(x_k-\mu)^T\Sigma^{-1}(x_k-\mu) \Leftrightarrow$$

$$\boxed{L(\underline{\mu}, \underline{\underline{\Sigma}}) = -\frac{N}{2}\ln[(2\pi)^{\ell}|\Sigma|] - \frac{1}{2}\sum_{k=1}^{N} x_k^T\Sigma^{-1}x_k - 2\mu^T\Sigma^{-1}x_k + \mu^T\Sigma^{-1}\mu \quad (A)}$$

$\underline{\underline{ή}}$

$$L(\underline{\mu}, \underline{\underline{\Sigma}}) = -\frac{N}{2}\{\ln(2\pi^{\ell}) + \ln(|\Sigma|)\} - \frac{1}{2}\sum_{k=1}^{N} x_k^T\Sigma^{-1}x_k - 2\mu^T\Sigma^{-1}x_k + \mu^T\Sigma^{-1}\mu \Leftrightarrow$$

$$\boxed{L(\underline{\mu}, \underline{\underline{\Sigma}}) = -\frac{\ell \cdot N}{2}\ln(2\pi) - \frac{N}{2}\ln(|\Sigma|) - \frac{1}{2}\sum_{k=1}^{N} x_k^T\Sigma^{-1}x_k - 2\mu^T\Sigma^{-1}x_k + \mu^T\Sigma^{-1}\mu}$$

$\underline{\underline{ή}}$

$$\boxed{L(\underline{\mu}, \underline{\underline{\Sigma}}) = -\frac{N\cdot\ell}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{k=1}^{N}(x_k-\mu)^T\Sigma^{-1}(x_k-\mu) \quad (B)}$$

#2 Διατυπώνοντας τις συνθήκες πρώτης τάξης (FOC$_s$) έχουμε ότι:

$$\left\{ \begin{array}{ll} \dfrac{\partial L(\underline{\mu}, \underline{\underline{\Sigma}})}{\partial \underline{\mu}} = \underline{0} & (1) \\[2ex] \dfrac{\partial L(\underline{\mu}, \underline{\underline{\Sigma}})}{\partial \underline{\underline{\Sigma}}} = \underline{\underline{0}} & (2) \end{array} \right\}$$

Η διατύπωση της συνθήκης (1) γίνεται καλύτερα από την σχέση (A):

$$\frac{\partial L}{\partial \underline{\mu}} = \frac{\partial}{\partial \underline{\mu}}\left[-\frac{N}{2}\ln\left((2\pi)^{\ell}|\Sigma|\right)\right] - \frac{1}{2}\frac{\partial}{\partial \underline{\mu}}\left[\sum_{k=1}^{N} x_k^T \Sigma^{-1} x_k - 2\mu^T \Sigma^{-1} x_k + \mu^T \Sigma^{-1}\mu\right] \Leftrightarrow$$

$$\frac{\partial L}{\partial \underline{\mu}} = -\frac{1}{2}\sum_{k=1}^{N} \frac{\partial}{\partial \underline{\mu}}\left[x_k^T \Sigma^{-1} x_k\right] - 2\frac{\partial}{\partial \underline{\mu}}\left[\mu^T \Sigma^{-1} x_k\right] + \frac{\partial}{\partial \underline{\mu}}\left[\mu^T \Sigma^{-1}\mu\right] \Leftrightarrow$$

$$\frac{\partial L}{\partial \underline{\mu}} = -\frac{1}{2}\sum_{k=1}^{N} -2\Sigma^{-1} x_k + \left(\Sigma^{-1} + (\Sigma^{-1})^T\right)\mu \Leftrightarrow \qquad \Sigma^{-1} = (\Sigma^{-1})^T \quad \underline{\text{ΣΥΜΜΕΤΡΙΚΟΣ}}$$

$$\frac{\partial L}{\partial \underline{\mu}} = -\frac{1}{2}\sum_{k=1}^{N} -2\Sigma^{-1} x_k + 2\Sigma^{-1}\mu \Leftrightarrow \frac{\partial L}{\partial \underline{\mu}} = \sum_{k=1}^{N} \Sigma^{-1} x_k - \Sigma^{-1}\mu \Leftrightarrow$$

$$\frac{\partial L}{\partial \underline{\mu}} = \Sigma^{-1} * \sum_{k=1}^{N}(x_k - \mu) = \underline{0} \quad \xLeftrightarrow{* \Sigma \text{ (both hand sides)}}$$

$$\Sigma \cdot \Sigma^{-1} \cdot \sum_{k=1}^{N}(x_k - \mu) = \underline{\underline{\Sigma}} \cdot \underline{0} \Leftrightarrow$$

$$\sum_{k=1}^{N}(x_k - \mu) = 0 \Leftrightarrow$$

$$\sum_{k=1}^{N} x_k - N\mu = 0 \Leftrightarrow \boxed{\hat{\mu}_{ML} = \frac{1}{N}\sum_{k=1}^{N} x_k}$$

#3 Η διατύπωση της συνθήκης (2) γίνεται καλύτερα από την σχέση (3):

$$\frac{\partial L}{\partial \underline{\underline{\Sigma}}} = \frac{\partial}{\partial \underline{\underline{\Sigma}}}\left[-\frac{N}{2}\ln|\Sigma|\right] - \frac{1}{2}\frac{\partial}{\partial \underline{\underline{\Sigma}}}\sum_{k=1}^{N}(x_k-\mu)^T\Sigma^{-1}(x_k-\mu) \Leftrightarrow$$

$$\frac{\partial L}{\partial \underline{\underline{\Sigma}}} = -\frac{N}{2}\cdot\frac{\partial \ln|\Sigma|}{\partial \underline{\underline{\Sigma}}} - \frac{1}{2}\sum_{k=1}^{N}\frac{\partial (x_k-\mu)^T\Sigma^{-1}(x_k-\mu)}{\partial \underline{\underline{\Sigma}}} \Leftrightarrow$$

$$\frac{\partial L}{\partial \underline{\underline{\Sigma}}} = -\frac{N}{2}\cdot\Sigma^{-1} - \frac{1}{2}\sum_{k=1}^{N} -\Sigma^{-1}(x_k-\mu)(x_k-\mu)^T\Sigma^{-1} \Leftrightarrow$$

$$\frac{\partial L}{\partial \underline{\underline{\Sigma}}} = -N\Sigma^{-1} + \Sigma^{-1} * \sum_{k=1}^{N}(x_k-\mu)(x_k-\mu)^T * \Sigma^{-1} = \underline{\underline{0}} \Leftrightarrow$$

$$N\Sigma^{-1} = \Sigma^{-1} * \sum_{k=1}^{N}(x_k-\mu)(x_k-\mu)^T \cdot \Sigma^{-1} \Rightarrow$$

$$N = \sum_{k=1}^{N}(x_k-\mu)(x_k-\mu)^T\Sigma^{-1} \Leftrightarrow$$

$$N\cdot\Sigma = \sum_{k=1}^{N}(x_k-\mu)(x_k-\mu)^T \Rightarrow \boxed{\hat{\Sigma}_{ML} = \frac{1}{N}\sum_{k=1}^{N}(x_k-\mu)(x_k-\mu)^T}$$

Άσκηση: Σε ένα πείραμα ρίψης νομίσματος η πιθανότητα εμφάνισης κεφαλής (head) (1) είναι q ενώ η πιθανότητα εμφάνισης γραμμάτων (tails) (0) είναι $1-q$. Έστω $x = \{x_1, x_2, \dots, x_N\}$ το σύνολο που περιγράφει την έκβαση N πειραμάτων με $x_i \in \{0,1\}, \forall i \in [N]$. #4

Να δείξετε ότι ο εκτιμητής μέγιστης πιθανοφάνειας για την πιθανότητα εμφάνισης κεφαλής δίνεται από την σχέση:

$$\boxed{\hat{q}_{ML} = \frac{1}{N} \sum_{k=1}^{N} x_i}$$

⊛ Υπόδειξη: Η συνάρτηση πιθανοφάνειας δίνεται από την σχέση:

$$\boxed{p(x;q) = \prod_{k=1}^{N} q^{x_k} \cdot (1-q)^{(1-x_k)}}$$

⊛ Σχηματίζουμε την λογαριθμική συνάρτηση πιθανοφάνειας:

$$L(q) = \ln\left[\prod_{k=1}^{N} q^{x_k}(1-q)^{(1-x_k)}\right] \Leftrightarrow$$

$$L(q) = \sum_{k=1}^{N} \ln\left[q^{x_k}(1-q)^{(1-x_k)}\right] \Leftrightarrow$$

$$L(q) = \sum_{k=1}^{N} \ln(q^{x_k}) + \ln(1-q)^{(1-x_k)} \Leftrightarrow$$

$$L(q) = \sum_{k=1}^{N} x_k \ln(q) + \sum_{k=1}^{N} (1-x_k)\ln(1-q) \Leftrightarrow$$

$$\boxed{L(q) = \ln(q)\sum_{k=1}^{N} x_k + \ln(1-q)\sum_{k=1}^{N}(1-x_k)}$$

$\boxed{\#5}$

$$\frac{\partial L}{\partial q} = \sum_{k=1}^{N} x_k \cdot \frac{\partial \ln q}{\partial q} + \sum_{k=1}^{N} (1-x_k) \cdot \frac{\partial \ln(1-q)}{\partial q} \iff$$

$$\frac{\partial L}{\partial q} = \sum_{k=1}^{N} x_k \cdot \frac{1}{q} + \sum_{k=1}^{N} (1-x_k) \cdot \frac{\partial(1-q)/\partial q}{1-q} \iff$$

$$\frac{\partial L}{\partial q} = \sum_{k=1}^{N} x_k \cdot \frac{1}{q} + \sum_{k=1}^{N} (1-x_k) \frac{-1}{1-q} = 0 \iff$$

$$\frac{1}{q} * \sum_{k=1}^{N} x_k - \frac{1}{1-q} \sum_{k=1}^{N} (1-x_k) = 0 \iff$$

$$\frac{1}{q} \sum_{k=1}^{N} x_k - \frac{1}{1-q} * \left\{ \sum_{k=1}^{N} 1 - \sum_{k=1}^{N} x_k \right\} = 0 \iff$$

$$\frac{1}{q} \sum_{k=1}^{N} x_k - \frac{1}{1-q} * \left\{ N - \sum_{k=1}^{N} x_k \right\} = 0$$

$$\sum_{k=1}^{N} x_k = S$$

$$\frac{1}{q} * S - \frac{1}{1-q} \left\{ N - S \right\} = 0 \iff$$

$$\frac{S}{q} - \frac{N}{1-q} + \frac{S}{1-q} = 0 \iff$$

$$\frac{S(1-q)}{q(1-q)} - \frac{N \cdot q}{(1-q) \cdot q} + \frac{S \cdot q}{q(1-q)} = 0 \iff$$

$$\frac{-Nq + S(1-q) + Sq}{q(1-q)} = 0 \iff$$

$$\frac{-Nq + S - \cancel{Sq} + \cancel{Sq}}{q(1-q)} = 0 \iff$$

$$Nq = S \Rightarrow \hat{q} = \frac{S}{N} \Rightarrow$$

$$\boxed{\hat{q} = \frac{1}{N} \sum_{k=1}^{N} x_k}$$