Φροντιστήριο 2

I) Τετραγωνικές Μορφές

Ως μορφή ορίζεται μια πολυωνυμική έκφραση της οποίας κάθε όρος έχει τον ίδιο βαθμό. Ο βαθμός ενός όρου είναι το άθροισμα των εκθετών των μεταβλητών που βρίσκονται στον όρο

Παράδειγμα Φ2.1

i) Το πολυώνυμο $9x + 5x^2 + 3x^3$ δεν είναι μορφή γιατί ο πρώτος όρος $(9x)$ έχει βαθμό 1, ο δεύτερος όρος έχει βαθμό 2 και ο τρίτος όρος $(3x^3)$ έχει βαθμό 3

ii) Το πολυώνυμο $9xy + 4x^2 + 8$ δεν είναι μορφή γιατί ο πρώτος όρος $(9xy)$ έχει βαθμό 2 (ο εκθέτης του x είναι 1, ο εκθέτης του y είναι 1, άρα το άθροισμά τους είναι 2), ο δεύτερος όρος $(4x^2)$ έχει βαθμό 2 και ο τρίτος όρος (8) έχει βαθμό 0

iii) Το πολυώνυμο $9xy - 5x^2 + 4y^2$ είναι μορφή γιατί ο πρώτος όρος $(9xy)$ έχει βαθμό 2, ο δεύτερος όρος $(-5x^2)$ έχει βαθμό 2 και ο τρίτος όρος έχει βαθμό 2.

Το παράδειγμα Φ.2.1. iii είναι ένα παράδειγμα τετραγωνικής μορφής, μιας μορφής που ο βαθμός κάθε όρου είναι 2.

Στα μαθηματικά της οικονομικής επιστήμης συναντάμε τις τετραγωνικές μορφές πολύ συχνά

Παράδειγμα Φ.2.2

i) Εστω μια συνάρτηση $f: X \times Y \to Z$, οπου $X, Y, Z \subset \mathbb{R}$. Αρα η συνάρτηση θα έχει την μορφή

$$z = f(x,y) = \{ z \in Z \mid x \in X,\ y \in Y \text{ and } (x,y,z) \in f \}$$

Το διαφορικό πρώτης τάξης (υποθέτουμε διαφορισιμότητα της f) θα είναι

$$dz = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = f_x dx + f_y dy$$

οπου $f_x \equiv \frac{\partial f}{\partial x}$ και $f_y \equiv \frac{\partial f}{\partial y}$

Το διαφορικό δευτερης ταξης $d(dz) = d^2 z$ θα είναι

$$d^2 z = d(dz) = d\left(f_x dx + f_y dy\right) \quad ***$$

$$= \frac{\partial}{\partial x}(f_x dx + f_y dy)\, dx + \frac{\partial}{\partial y}(f_x dx + f_y dy) dy$$

$$= (f_{xx} dx + f_{xy} dy) dx + (f_{yx} dx + f_{yy} dy) dy$$

$$= f_{xx} dx^2 + f_{xy} dy dx + f_{yx} dx dy + f_{yy} dy^2$$

Απο το θεώρημα του Young εφόσον f_{xy}, f_{yx} είναι συνεχείς θα έχουμε $f_{xy} = f_{yx}$

*** το διαφορικό $dz = f_x dx + f_y dy$ είναι συνάρτηση $dz: X \times Y \to Z$

Άρα το διαφορικό δεύτερης τάξης παίρνει την μορφή

$$d^2z = f_{xx}dx^2 + 2f_{xy}dx\,dy + f_{yy}dy^2$$

Αν θεωρήσουμε τις μερικές παραγώγους σαν σταθερές και τα διαφορικά σαν μεταβλητές τότε το d^2z είναι μια τετραγωνική μορφή

Το ενδιαφέρον είναι ότι μπορούμε να εκφράσουμε κάθε τετραγωνική μορφή q με μορφή πινάκων $\underset{1\times n}{z^T}\underset{n\times n}{A}\underset{n\times 1}{z} = \underset{1\times 1}{q}$

Για το Φ21 iii, αν ορίσουμε ως $z^T = (x\ y)$ το διάνυσμα μεταβλητών παρατηρούμε ότι για έναν πίνακα $A \in F_{2\times 2}$ ίσο με

$$A = \begin{pmatrix} 9 & 5 \\ -10 & 4 \end{pmatrix}$$

$$z^TAz = (x\ y)\begin{pmatrix} 9 & 5 \\ -10 & 4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = (x\ y)\begin{pmatrix} 9x+5y \\ -10x+4y \end{pmatrix}$$

$$= x(9x+5y) + y(-10x+4y)$$

$$= 9x^2 + 5xy - 10xy + 4y^2$$

$$= 9x^2 - 5xy + 4y^2$$

Για το Φ22 έχουμε ότι το d^2z μπορεί να γραφει ώς

$$dz = f_{xx}dx^2 + 2f_{xy}dxdy + f_{yy}dy^2 = (x\ y)\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

II) Θετικά και Αρνητικά Ορισμένες Μορφές

Μια τετραγωνική μορφή q καλείται

- Θετικά ορισμένη, αν $q > 0$
- Θετικά ημιορισμένη, αν $q \geq 0$
- αρνητικά ορισμένη, αν $q < 0$
- αρνητικά ημιορισμένη αν $q \leq 0$

Αν η τετραγωνική μορφή αλλάζει πρόσημο καθώς αλλάζουν οι τιμές των μεταβλητών τότε λέμε ότι η τετραγωνική μορφή είναι αόριστη.

!!! Σημείωση: Υπάρχουν κριτήρια ορίζουσας για τον πίνακα A στην τετραγωνική μορφή $q = z^T A z$ που μας επιτρέπουν να βρούμε αν η q είναι θετικά ορισμένη κ.λ.π.

!!! Σημείωση 2: Γενικά όταν βελτιστοποιούμε μια συνάρτηση $f: \mathbb{R}^n \to \mathbb{R}$, η οποία είναι C^2 θέλουμε η τετραγωνική μορφή που σχετίζεται με το διαφορικό δεύτερης τάξης $d^2 z = x^T \mathbb{H} x$ όπου

$$x^T = (x_1\ x_2 \ldots x_n)$$

$$\mathbb{H} = \begin{pmatrix} f_{11} & f_{12} & & f_{1n} \\ f_{21} & f_{22} & & f_{2n} \\ & & & \\ f_{n1} & f_{n2} & & f_{nn} \end{pmatrix}$$

να είναι

1) αρνητικά ορισμένη για μέγιστο της f
2) θετικά ορισμένη για ελάχιστο της f

III) Παραγώγηση Πινάκων

Εστω μια συνάρτηση $f: \mathbb{R}^n \to \mathbb{R}$, η οποία είναι δυο φορές συνεχώς διαφορίσιμη

Αυτή η συναρτηση δηλαδή μπορεί να γραφεί ως

$$y = f(x_1, x_2, \dots, x_n)$$

Συνήθως, στα οικονομικά, όταν έχουμε ενα διανυσμα $(x_1, x_2, \dots, x_n)$ το μεταχειριζόμαστε σαν έναν πίνακα $n \times 1$ (ή εναν πινακα στήλη) και το γράφουμε

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

και όταν θέλουμε να γράψουμε ένα διάνυσμα γραμμη $1 \times n$ τότε το γράφουμε $\mathbf{x}^T = (x_1\ x_2 \dots x_n)$

Αρα την συναρτηση f μπορουμε να την εκφράσουμε ως $f: \mathbb{R}^{n\times1} \to \mathbb{R}$ δηλαδη $y = f(\mathbf{x})$ (οπου $\mathbf{x}^T = (x_1\ x_2\ \ x_n)$) είναι συνάρτηση ενός διανύσματος στήλη $n \times 1$

Μπορούμε να ορίσουμε το διάνυσμα πρώτων μερικών παραγώγων της f ως

$$\frac{df(\mathbf{x})}{d\mathbf{x}} \equiv \nabla f(\mathbf{x}) = \begin{pmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \\ \vdots \\ \partial f / \partial x_n \end{pmatrix} = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \\ f_n(\mathbf{x}) \end{pmatrix} \quad \text{όπου } f_i(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_i}, \ \forall i = 1 \dots n$$

!!! Αρα παραγωγήσαμε εναν αριθμό $y = f(\mathbf{x}) \in \mathbb{R}$ ως προς ενα διάνυσμα στήλη $(n\times1)$ και πήραμε ενα διανυσμα στήλη $(n\times1)$

Ορίζουμε την μήτρα δευτέρων μερικών παραγώγων της f ως προς x, ως

$$\frac{d^2 f(x)}{dx^T dx} = \frac{d}{dx^T}\left(\nabla f(x)\right)$$

$$= \frac{d}{dx^T}\left[\begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{pmatrix}\right] = \begin{pmatrix} f_{11}(x) & f_{21}(x) & \dots & f_{n1}(x) \\ f_{12}(x) & f_{22}(x) & \dots & f_{n2}(x) \\ \vdots & & & \\ f_{1n}(x) & f_{2n}(x) & \dots & f_{nn}(x) \end{pmatrix}$$

όπου $f_{ij}(x) = \dfrac{\partial^2 f(x)}{\partial x_i \partial x_j}$

Ο πίνακας δευτέρων μερικών παραγώγων $\dfrac{d^2 f(x)}{dx^T dx}$ ονομάζεται Εσσιανός πίνακας και συμβολίζεται με H (ο ίδιος πίνακας με την παράδειγμα μορφή του παραδείγματος Φ.2.2. για 2 διαστάσεις)

<u>Σημείωση 3</u>: Γενικά παρατηρούμε ότι όταν παραγωγίζουμε έναν αριθμό με ένα διάνυσμα γραμμή (σειρά) καταλήγουμε με ένα διάνυσμα γραμμή (σειρά) μερικών παραγώγων

Σαν παράδειγμα στον Εσσιανό πίνακα (που παραγωγίζουμε το διάνυσμα df/dx ως προς ένα διάνυσμα γραμμή (το x^T), <u>κάθε</u> βαθμωτός στο $\nabla f(x)$, $f_1(x), f_2(x) \dots f_n(x)$ μετατρέπεται μετά την παραγώγηση σε ένα διάνυσμα γραμμή $(f_{11}(x) \dots f_{n1})$, $(f_{12}(x), \dots, f_{n2}(x)), \dots, (f_{1n}(x) \dots, f_{nn}(x))$

Όταν ισχύει το θεώρημα του Young (δηλαδή $f_{ij} = f_{ji}$)
τότε $H = H^T$ δηλαδή ο Εσσιανός είναι συμμετρικός

IV Κανόνες Παραγώγησης Πινάκων

Έστω $a^T = (a_1 \dots a_n)$, $x^T = (x_1\, x_2 \dots x_n)$ και

$$H = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{pmatrix} = H^T \quad (\text{H είναι συμμετρική})$$

Ισχύουν τα παρακάτω

1. $$\frac{\partial a^T x}{\partial x} = \underset{n\times 1}{a}$$

Παραγωγήσαμε έναν βαθμωτό $\underbrace{a^T x}_{1\times 1}$ ως προς ένα διάνυσμα στήλη $\underset{n\times 1}{x}$ και πήραμε ένα διάνυσμα στήλη $n\times 1$ το a

2. $$\frac{\partial x^T a}{\partial x^T} = \underset{1\times n}{a^T}$$

Παραγωγήσαμε έναν βαθμωτό $x^T a$ ως προς ένα διάνυσμα γραμμή $\underset{1\times n}{x^T}$ και πήραμε ένα διάνυσμα γραμμή $1\times n$

Ομοίως

3. $$\frac{\partial a^T x}{\partial x^T} = \underset{1\times n}{a^T}$$

4. $$\frac{\partial x^T a}{\partial x} = \underset{n\times 1}{a}$$

5. $\underset{n\times 1}{\dfrac{\partial}{\partial x}} (\underbrace{x^T H x}_{1\times 1}) = 2 \, \underbrace{H x}_{n\times 1}$

6. $\underset{n\times 1}{\dfrac{\partial}{\partial x^T}} (\underbrace{x^T H x}_{1\times 1}) = 2 \, \underbrace{x^T H}_{n\times 1}$

7. $\dfrac{\partial^2}{\partial x^T \partial x} \left(x^T H x \right) = \dfrac{\partial}{\partial x^T} \left(\dfrac{\partial (x^T H x)}{\partial x} \right) = \underset{1\times n}{\dfrac{\partial}{\partial x^T}} \left(2 \underset{n\times 1}{H x} \right)$

$= 2H$

8. $\underset{n\times n}{\dfrac{\partial}{\partial H}} \left(\underset{1\times 1}{x^T H x} \right) = x x^T$

Παρατηρείστε ότι στο 8 παραγωγίσαμε εναν βαθμωτό 1×1 ως προς μια μήτρα $n\times n$ και πήραμε μια μήτρα $n\times n$

Σημείωση 4: Αν ο πίνακας H δεν είναι συμμετρικός οι κανόνες 7 και 8 δεν ισχύουν και για να υπολογίσετε την παράγωγο μιας τετραγωνικής μορφής θα πρέπει να παραγωγίσετε το φ, 1×1 (αφού κάνετε τους πολλαπλασιασμούς) με το x, x^T

9. $\dfrac{\partial}{\partial A} (x^T A^{-1} x) = - A^{-1} x x^T A^{-1}$

10 $\dfrac{\partial}{\partial A} \ln |A| = (A^{-1})^T$ όπου $|A|$ η ορίζουσα του A

$\left(= A^{-1} \text{ αν } A^T = A \right)$

Σημείωση: Ισχύει ο κανόνας της αλυσίδας όπως τον ξέρουμε στην παραγώγηση πινάκων

Ασκηση 1

Εστω $A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 5 & 3 \\ -1 & 3 & 0 \end{pmatrix}$, $x^T = (x_1\ x_2\ x_3)$, $b^T = (b_1\ b_2\ b_3)$

Να υπολογίσετε μια τετραγωνική μορφη χρησιμοποιώντας τον πινακα A και το διανυσμα μεταβλητών x
Επειτα υπολογίσετε

i) $\dfrac{\partial b^T x}{\partial b}$ ii) $\dfrac{\partial x^T b}{\partial x^T}$ iii) $\dfrac{\partial x^T b}{\partial x}$ iv) $\dfrac{\partial q}{\partial x}$ v) $\dfrac{\partial^2 q}{\partial x^T \partial x}$

Λύση

Θα υπολογήσουμε την τετραγωνική μορφή q με βαση τον τύπο $q = x^T A x$. Εχουμε

$$q = x^T A x = (x_1\ x_2\ x_3) \begin{pmatrix} 1 & 2 & -1 \\ 2 & 5 & 3 \\ -1 & 3 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$= (x_1\ x_2\ x_3) \begin{pmatrix} x_1 + 2x_2 - x_3 \\ 2x_1 + 5x_2 + 3x_3 \\ -x_1 + 3x_2 \end{pmatrix}$$

$$= x_1(x_1 + 2x_2 - x_3) + x_2(2x_1 + 5x_2 + 3x_3) + x_3(-x_1 + 3x_2)$$

$$= x_1^2 + 2x_1x_2 - x_1x_3 + 2x_1x_2 + 5x_2^2 + 3x_2x_3 - x_1x_3 + 3x_2x_3$$

$$= x_1^2 + 4x_1x_2 - 2x_1x_3 + 6x_2x_3 + 5x_2^2$$

i) $\dfrac{\partial b^T x}{\partial b} = \dfrac{\partial}{\partial b}(b_1x_1 + b_2x_2 + b_3x_3) = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = b$

ii) $\dfrac{\partial x^T b}{\partial x^T} = \dfrac{\partial}{\partial x^T}(b_1x_1 + b_2x_2 + b_3x_3)$

$$= \left[\frac{\partial}{\partial x_1}(b_1x_1+b_2x_2+b_3x_3) \quad \frac{\partial}{\partial x_2}(b_1x_1+b_2x_2+b_3x_3) \quad \frac{\partial}{\partial x_3}(b_1x_1+b_2x_2+b_3x_3)\right]$$

$$= (b_1 \ \ b_2 \ \ b_3) = b^T$$

iii) $\dfrac{\partial x^T b}{\partial x} = \dfrac{\partial}{\partial x}(b_1x_1 + b_2x_2 + b_3x_3)$

$$= \begin{pmatrix} \frac{\partial}{\partial x_1}(b_1x_1+b_2x_2+b_3x_3) \\ \frac{\partial}{\partial x_2}(b_1x_1+b_2x_2+b_3x_3) \\ \frac{\partial}{\partial x_3}(b_1x_1+b_2x_2+b_3x_3) \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = b$$

iv) $\dfrac{\partial q}{\partial x} = \dfrac{\partial}{\partial x}(x^T A x) = 2Ax = 2\begin{pmatrix} 1 & 2 & -1 \\ 2 & 5 & 3 \\ -1 & 3 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$

$$= \begin{pmatrix} 2 & 4 & -2 \\ 4 & 10 & 6 \\ -2 & 6 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 + 4x_2 - 2x_3 \\ 4x_1 + 10x_2 + 6x_3 \\ -2x_1 + 6x_2 \end{pmatrix}$$

αφου A είναι συμμετρικός πίνακας

Εναλλακτικά μπορούμε να υπολογίσουμε το iv ακολούθως

$$\underset{3\times1}{\frac{\partial q}{\partial x}} = \frac{\partial}{\partial x}\left(x_1^2 + 4x_1x_2 - 2x_1x_3 + 6x_2x_3 + 5x_2^2\right)$$

$$= \begin{pmatrix} \partial q/\partial x_1 \\ \partial q/\partial x_2 \\ \partial q/\partial x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 + 4x_2 - 2x_3 \\ 4x_1 + 10x_2 + 6x_3 \\ -2x_1 + 6x_2 \end{pmatrix}$$

v) $$\frac{\partial^2 q}{\partial x^T \partial x} = \frac{\partial}{\partial x^T}\left(\frac{\partial q}{\partial x}\right) = \frac{\partial}{\partial x^T}\begin{pmatrix} 2x_1 + 4x_2 - 2x_3 \\ 4x_1 + 10x_2 + 6x_3 \\ -2x_1 + 6x_2 \end{pmatrix}$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1}(2x_1 + 4x_2 - 2x_3) & \frac{\partial}{\partial x_2}(2x_1 + 4x_2 - 2x_3 & \frac{\partial}{\partial x_3}(2x_1 + 4x_2 - 2x_3) \\ \frac{\partial}{\partial x_1}(4x_1 + 10x_2 + 6x_3) & \frac{\partial}{\partial x_2}(4x_1 + 10x_2 + 6x_3) & \frac{\partial}{\partial x_3}(4x_1 + 10x_2 + 6x_3) \\ \frac{\partial}{\partial x_1}(-2x_1 + 6x_2) & \frac{\partial}{\partial x_2}(-2x_1 + 6x_2) & \frac{\partial}{\partial x_3}(-2x_1 + 6x_2) \end{bmatrix}$$

$$= \begin{pmatrix} 2 & 4 & -2 \\ 4 & 5 & 3 \\ -2 & 3 & 0 \end{pmatrix} = A$$

Ασκηση 2 Εκτιμητής Μεγίστης Πιθανοφανειας για την Πολυμεταβλητή Κανονική Κατανομή

Εστω ενα τυχαίο διάνυσμα $x^T = (x_1, x_2, \dots, x_n)$ που περιλαμβανει n τυχαίες μεταβλητές x_i, $i=1,\dots,n$ και $\mu^T = (\mu_1, \mu_2 \dots \mu_n)$ το διανυσμα που περιλαβμάνει τους πληθυσμιακούς μέσους (αναμενόμενες τιμες) της καθε μεταβλητής και Σ ο πινακας διακύμανσης - συνδιακύμανσης των μεταβλητών

Δηλαδή $E(x) = \mu$ και $E[(x-\mu)(x-\mu)^T] = \Sigma$

Επίσης δίνεται ότι $\underset{n\times 1}{x} \overset{iid}{\sim} \mathcal{N}(\underset{n\times 1}{\mu}, \underset{n\times n}{\Sigma})$

Υποθετουμε ότι τραβάμε ενα δείγμα μεγέθους m απο τον πληθυσμό που περιλαμβάνει ανάλογα τυχαία διανύσματα $x^{j^T} = (x_1^j\ x_2^j \dots x_n^j)$, $j = 1, \dots, m$

Εφόσον καθε x^j ακολουθεί την πολυμεταβλητή κανονική κατανομή, η πιθανότητα να έχουμε την παρατήρηση x^j είναι

$$f(x^j) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x^j-\mu)^T \Sigma^{-1} (x^j-\mu)\right\}$$

οπου $f(x^j)$ η συναρτηση πυκνότητας πιθανότητας της παρατήρησης x^j και $|\Sigma| = \det(\Sigma) > 0$ η ορίζουσα του Σ

Η συνάρτηση πιθανοφανειας (πρακτικα η πιθανότητα να λάβουμε το συγκεκριμενο δείγμα $(x^1, x^2, \dots, x^m) = \xi$) ~~δίνεται~~ εφόσον x^j είναι iid δίνεται από

$$L(\mu, \Sigma, x^1, x^2, \dots x^m) = \prod_{j=1}^{m} f(x^j)$$

$$= \prod_{j=1}^{m} (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x^j-\mu)^T \Sigma^{-1} (x^j-\mu)\right\}$$

$$= (2\pi)^{-\frac{mn}{2}} |\Sigma|^{-\frac{m}{2}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{m}\left[(x^j-\mu)^T \Sigma^{-1} (x^j-\mu)\right]\right\}$$

<u>προσοχή</u> Σ^{-1} είναι ο αντίστροφος του πίνακα διακύμανσης συνδιακύμανσης Σ ~~εκ~~ $\sum_{j=1}^{m}$ είναι το σύμβολο αθροίσματος όρων από $j=1,\dots,m$

Ο λογάριθμος της συνάρτησης μέγιστης πιθανοφάνειας είναι

$$\log L(\cdot) = -\frac{mn}{2}\ln(2\pi) - \frac{m}{2}\ln(|\Sigma|) - \frac{1}{2}\sum_{j=1}^{m}\left[(x^j-\mu)^T \Sigma^{-1} (x^j-\mu)\right]$$

Σκοπός της εκτιμητικής είναι να χρησιμοποιήσουμε τις πληροφορίες που μας παρέχει το δείγμα έτσι ώστε να βρούμε εκτιμήσεις για τις άγνωστες παραμέτρους του πληθυσμού

Εδώ σκοπός μας είναι να βρούμε εκτιμήσεις $\hat{\mu}, \hat{\Sigma}$ για το διάνυσμα μ και πίνακα Σ τέτοιες ώστε να μεγιστοποιήσουν την πιθανότητα να λάβουμε το συγκεκριμένο δείγμα

Άρα θέλουμε να λύσουμε το ακόλουθο πρόβλημα μεγιστοποίησης

$$\max_{\hat{\mu},\,\hat{\Sigma}} \ln L(\cdot) = -\frac{mn}{2}\ln(2\pi) - \frac{m}{2}\ln|\hat{\Sigma}| - \frac{1}{2}\sum_{j=1}^{m}\left[(x^j-\hat{\mu})^T\hat{\Sigma}^{-1}(x^j-\hat{\mu})\right]$$

Οι συνθήκες πρώτης τάξης είναι

$$\underset{n\times 1}{\frac{\partial \ln L(\cdot)}{\partial \hat{\mu}}} = \underset{n\times 1}{0} \iff -\frac{1}{2}\sum_{j=1}^{m}\frac{\partial}{\partial\hat{\mu}}\left[(x^j-\hat{\mu})^T\hat{\Sigma}^{-1}(x^j-\hat{\mu})\right] = \underset{n\times 1}{\mathbb{0}}$$

$$\text{or } (-2)\left(-\frac{1}{2}\right)\sum_{j=1}^{m}\left(\hat{\Sigma}^{-1}(x^j-\hat{\mu})\right) = \underset{n\times 1}{\mathbb{0}}$$

$$\text{or } \hat{\Sigma}^{-1}\sum_{j=1}^{m}(x^j-\hat{\mu}) = \underset{n\times 1}{\mathbb{0}}$$

$$\text{or } \hat{\Sigma}\hat{\Sigma}^{-1}\sum_{j=1}^{m}(x^j-\hat{\mu}) = \mathbb{0}$$

$$\text{or } \sum_{j=1}^{m}x^j - \sum_{j=1}^{m}\hat{\mu} = \mathbb{0}$$

$$\text{or } \sum_{j=1}^{m}x^j = m\hat{\mu} \iff$$

$$\hat{\mu} = \frac{1}{m}\sum_{j=1}^{m}x^j$$

$$\frac{\partial \ln L(\cdot)}{\partial \hat{\Sigma}} = 0 \iff -\frac{m}{2}\hat{\Sigma}^{-1} - \frac{1}{2}\sum_{j=1}^{m}\frac{\partial}{\partial\hat{\Sigma}}\left[(x^j-\hat{\mu})^T\hat{\Sigma}^{-1}(x^j-\hat{\mu})\right] = \underset{n\times n}{\mathbb{0}}$$

$$\text{or } -\frac{m}{2}\hat{\Sigma}^{-1} + \frac{1}{2}\sum_{j=1}^{m}\left[\hat{\Sigma}^{-1}(x^j-\hat{\mu})(x^j-\hat{\mu})^T\hat{\Sigma}^{-1}\right] = \mathbb{0}$$

$$\text{or } -m\hat{\Sigma}^{-1} + \hat{\Sigma}^{-1}\left(\sum_{j=1}^{m}(x^j-\hat{\mu})(x^j-\hat{\mu})^T\right)\hat{\Sigma}^{-1} = \mathbb{0}$$

$$\text{or } m\hat{\Sigma}^{-1} = \hat{\Sigma}^{-1}\left(\sum_{j=1}^{m}(x^j-\hat{\mu})(x^j-\hat{\mu})^T\right)\hat{\Sigma}^{-1}$$

$$\text{or } m = \hat{\Sigma}^{-1}\left(\sum_{j=1}^{m}(x^j-\hat{\mu})(x^j-\hat{\mu})^T\right)$$

$$\text{or } m\hat{\Sigma} = \sum_{j=1}^{m}(x^j-\hat{\mu})(x^j-\hat{\mu})^T \iff \hat{\Sigma} = \frac{1}{m}\left[\sum_{j=1}^{m}(x^j-\hat{\mu})(x^j-\hat{\mu})^T\right]$$