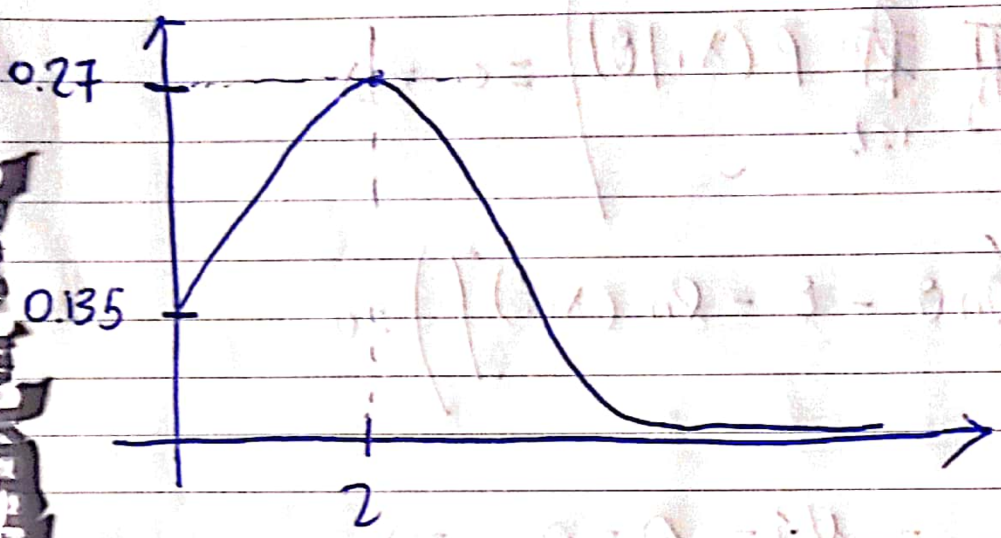
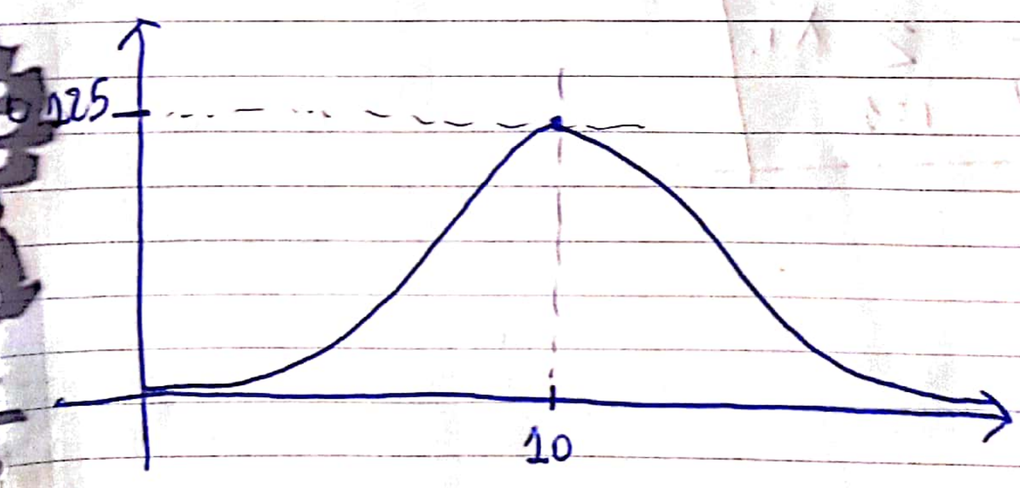
Εμμανουήλ-Αναστάσιος Σερλής/03118125/manosserlis@gmail.com

Άσκηση 1.1

α) Για $\theta=2$:



Για $\theta=10$:



(β) Ανεξαρτησία δειγμάτων: Για καθένα από τα δείγματα $x_1, x_2, \dots, x_N$, γνωρίζουμε ότι αυτά ακολουθούν κατανομή: $p(x_i|\theta) = \dfrac{\theta^{x_i}}{x_i!} e^{-\theta}$

Scanned with CamScanner

Έτσι, η maximum likelihood έκφραση για όλα τα δείγματα της κατανομής παίρνει την μορφή:

$$\frac{d}{d\theta}\left(\ln \prod_{i=1}^{\nu} p(x_i|\theta)\right) = 0 \Rightarrow$$

$$\Rightarrow \frac{d}{d\theta}\left(\sum_{i=1}^{\nu}\left[x_i \ln\theta - \theta - \ln(x_i!)\right]\right) = 0$$

$$\Rightarrow \frac{1}{\theta}\sum_{i=1}^{\nu} x_i - \nu \cdot 1 - 0 = 0 \Rightarrow$$

$$\Rightarrow \boxed{\hat{\theta} = \frac{1}{\nu}\sum_{i=1}^{\nu} x_i}$$

Scanned with CamScanner

Άσκηση 12

Αρχικά, υπολογίζουμε την παράγωγο του κριτηρίου κόστους ως:

$$\nabla J_s(a) = \nabla(\|Y \cdot a - b\|^2) = 2Y^T(Ya-b)$$

, η οποία ελαχιστοποιείται για $\begin{cases} a = Y^+ b \\ Y^+ = (Y^TY)^{-1}Y^T \end{cases}$

$$\Rightarrow a = (Y^TY)^{-1}Y^Tb \Rightarrow Y^TYa = Y^Tb \Rightarrow$$

$$\Rightarrow Y^T(Ya-b) = 0 \Rightarrow 2Y^T(Ya-b) = 0 \Rightarrow$$

$$\Rightarrow \boxed{\nabla J_s(a) = 0}$$

Scanned with CamScanner

Εφαρμόζουμε gradient descent για $n_k = \frac{1}{k}$ βήματα

$$\left\{\begin{aligned} a_{k+1} &= a_k - \frac{1}{k} Y^T(Ya_k - b) \\ a_k &= a_{k-1} - \frac{1}{k-1} Y^T(Ya_{k-1} - b) \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow a_{k+1} = \left[a_{k-1} - \frac{1}{k-1} Y^T(Ya_{k-1} - b)\right] -$$

$$- \frac{1}{k} Y^T \left(Y \cdot \left[a_{k-1} - \frac{1}{k-1} Y^T(Ya_{k-1} - b)\right] - b\right)$$

$$= a_{k-1}\left[1 - \frac{Y^TY}{k-1} - \frac{Y^TY}{k} + \frac{1}{k(k-1)}(Y^TY)^2\right]$$

$$+ b\left[\frac{Y^T}{k-1} + \frac{Y^T}{k} - \frac{Y^TY \cdot Y^T}{k(k-1)}\right]$$

Συνεπώς, για $\kappa \to +\infty$, παίρνουμε:

$$\lim_{\kappa \to +\infty} a_{n+1} = \left(\lim_{\kappa \to +\infty} a_{n-1} \right) \cdot 1 + 0 \cdot b$$

$$\Longrightarrow \quad a_{n+1} \Big|_{\kappa \to +\infty} = a_{n-1} \Big|_{\kappa \to +\infty} = a \ (?)$$

$$= a_n \Big|_{\kappa \to +\infty}$$

Scanned with CamScanner

Έτσι, αφού $\lim\limits_{k\to+\infty} a_{k+1} = \lim\limits_{k\to+\infty} a_k = 0 \Rightarrow$

$$\Rightarrow \lim_{k\to+\infty} Y^T(Ya_k - b) = \lim_{k\to+\infty} \left[k(a_k - a_{k+1}) \right]$$

$$= \lim_{k\to+\infty} \frac{k}{\dfrac{1}{a_k - a_{k+1}}} = \lim_{k\to+\infty} \frac{1}{\dfrac{+1}{(a_{k+1}-a_k)^2}(a'_{k+1} - a'_k)}$$

$=$

Scanned with CamScanner

$$\lim_{n\to+\infty} \frac{(a_{n+1}-a_n)^2}{\left[\frac{a_{n+2}-a_{n+1}}{1}-\frac{a_{n+1}-a_n}{1}\right]}$$

$$=\lim_{n\to+\infty} \frac{(a_{n+1}-a_n)^2}{(a_{n+2}-a_n)} \qquad \underline{\underline{a_{n+2}=a_{n+1}\ (n\to+\infty)}}$$

$$=\lim_{n\to+\infty} \frac{(a_{n+1}-a_n)^{\cancel{2}}}{\cancel{(a_{n+1}-a_n)}} = \lim_{n\to+\infty} (a_{n+1}-a_n)=0$$

, λόγω σχέσης (2).

<u>Άρα</u>: $\boxed{\lim_{n\to+\infty} Y^T(Ya-b)=0.}$

Scanned with CamScanner

Άσκηση 1.3

Αρχικά, οδηγηθήκαμε στην δημιουργία 15 δειγμάτων με την ζητούμενη αναλογία 2:1, τα οποία παρατίθενται παρακάτω:

[1.51, 1.44, 2.23, 1.4, 1.04, 2.58, 1.51, 1.0, 2.57, 1.52, 1.32, 2.02, 1.03, 1.01, 2.28]

Για μία παρατήρηση, έχουμε:

$$p(x,z|\theta) = p(x|z,\theta)\cdot p(z|\theta) =$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(x-\mu_0)}{2\sigma_0^2}} (1-p) \right)^{1-z} \left(\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)}{2\sigma_1^2}} \right)^{z}$$

Ορίζουμε την επιπμώμενη πιθανότητα $p(x,z|\theta)$ ως:

$$Q(\theta;\theta^+) = \mathbb{E}_{z|x,\theta^+}[\log p(x,z|\theta)]$$

$$= (1-\mathbb{E}[z])\left[-\log\sigma_0 - \frac{(x-\mu_0)^2}{2\sigma_0^2} + \log(1-p_0) + c_0\right]$$

$$+ \mathbb{E}[z]\left[-\log\sigma_1 - \frac{(x-\mu_1)^2}{2\sigma_1^2} + \log p + c_1\right]$$

Scanned with CamScanner

Για $n=15$ παρατηρήσεις, έχουμε:

$$Q(\theta;\theta^t) = \sum_{i=1}^{15} (1-\mathbb{E}[z_i])\left[-\log\sigma_0 - \frac{(x_i-\mu_0)^2}{2\sigma_0^2} + \log(1-p_0)\right] + \mathbb{E}[z_i]\left[-\log\sigma_1 - \frac{(x_i-\mu_1)^2}{2\sigma_1^2} + \log p_1\right]$$

<u>Expectation step:</u> $\mathbb{E}[z_i] = 0 \cdot P(z_i=0|x_i) + 1 \cdot P(z_i=1|x_i) = P(z_i=1|x_i) = \dfrac{P(x_i|z_i=1)p_1}{\sum_{j=0}^{1} P(x_i|z_i=j)\,p_j}$

$$\Rightarrow \mathbb{E}[z_i] = \frac{P(x_i|z_i=1)\cdot p_1}{P(x_i|z_i=1)\cdot p_1 + P(x_i|z_i=0)\cdot p_0}, \quad \mu\epsilon\ p_0+p_1=1$$

<u>Maximization step:</u>

$$\frac{dQ}{d\mu_1} = \sum \mathbb{E}[z_i]\left(\frac{x_i-\mu_1}{\sigma_1^2}\right) = 0 \Rightarrow \boxed{\mu_1 = \frac{\sum_1^{15}\mathbb{E}[z_i]x_i}{\sum_1^{15}\mathbb{E}[z_i]}}$$

$$\frac{dQ}{d\sigma_1} = \sum \mathbb{E}[z_i]\left(-\frac{1}{\sigma_1} + \frac{(x_i-\mu_1)^2}{\sigma_1^3}\right) = 0 \Rightarrow$$

$$\Rightarrow \boxed{\sigma_1^2 = \frac{\sum \mathbb{E}[z_i](x_i-\mu_1)^2}{\sum \mathbb{E}[z_i]}}$$

Scanned with CamScanner

$$\frac{dQ}{dp_1} = \sum_{i=1}^{15} \frac{\mathbb{E}[z_i]-1}{1-p_1} + \frac{\mathbb{E}[z_i]}{p_1} = 0 \Rightarrow$$

$$\Rightarrow \boxed{p_1 = \frac{\sum_{i=1}^{15} \mathbb{E}[z_i]}{15}} \quad \text{και} \quad \boxed{p_0 = 1 - p_1}$$

Ομοίως βγαίνει ότι:

$$\boxed{\mu_0 = \frac{\sum_{i}^{15} (1-\mathbb{E}[z_i])x_i}{\sum_{i=1}^{15} (1-\mathbb{E}[z_i])}}$$

$$\boxed{\sigma_0^2 = \frac{\sum (1-\mathbb{E}[z_i])(x_i-\mu_1)^2}{\sum (1-\mathbb{E}[z_i])}}$$

Scanned with CamScanner

<u>Αρχικοποίηση</u>: Αρχικά υπολογίζουμε τις εμπειρικές τιμές για mean και variance από τα δείγματα ως $\mu_{init} = 1,92$ $\sigma^2_{init} = 0,21$ και εν συνεχεία παίρνουμε τις κάτωθι περιπτώσεις αρχικοποίησης

<u>1η περίπτωση:</u>

	μ_0	μ_1	σ_0^2	σ_1^2	p_0	p_1
$i=0$	$\mu_{init}-0.01$	$\mu_{init}+0.01$	$\sigma^2_{init}-0.01$	$\sigma^2_{init}+0.01$	0.49	0.51
$i=1$	1.7	1.92	0.10	0.25	0.43	0.50
$i=2$	1.52	2.00	0.17	0.20	0.44	0.40
$i=3$	1.50	2.3	0.32	0.19	0.59	0.32

<u>2η περίπτωση:</u>

	μ_0	μ_1	σ_0^2	σ_1^2	p_0	p_1
$i=0$	$\mu_{init}+0.01$	$\mu_{init}-0.01$	$\sigma^2_{init}+0.01$	$\sigma^2_{init}-0.01$	0.51	0.49
$i=1$	1.92	1.7	0.25	0.15	0.50	0.43
$i=2$	2.059	1.52	0.30	0.04	0.57	0.43
$i=3$	2.031	1.53	0.5	0.004	0.59	0.41

Scanned with CamScanner

Παρατηρούμε ότι στην περίπτωση αρχικοποίησης
1 ..., η μέση τιμή συγκλίνει στις τιμές των
κατανομών, κάτι το οποίο δεν συμβαίνει στην 2η
περίπτωση. Από την άλλη, και στις 2 περιπτώσεις,
ο αλγόριθμος υπερτιμά την τιμή σ_0^2 οδηγώντας
το variance σ_1^2 σε χαμηλότερες τιμές. Τέλος,
όσον αφορά τις πιθανότητες $p(Z=0)=p_0$
και $p(Z=1)=p_1$, παρατηρούμε πως και στις 2
περιπτώσεις οδηγούμαστε σε αναλογία που
προσεγγίζει αυτή των δεδομένων εισόδου, δηλαδή
2 : 1

Άσκηση 1.4

α). Ανεξαρτησία δειγμάτων:

$$P(D|\theta) = P(x_1, \ldots, x_n | \theta) = \prod_{k=1}^{n} P(x_k|\theta)$$

$$P(x|\theta) = \prod_{i=1}^{d} \theta_i^{x_i} (1-\theta_i)^{1-x_i}$$

$$\Rightarrow P(D|\theta) = \prod_{k=1}^{n} \prod_{i=1}^{d} \theta_i^{x_{ki}} (1-\theta_i)^{x_{ki}} =$$

$$= \prod_{i=1}^{d} \theta_i^{(x_{1i}+x_{2i}+\ldots+x_{ni})} (1-\theta_i)^{\{(1-x_{1i})+(1-x_{2i})+\ldots+(1-x_{ni})\}} \xrightarrow{x_{1i}+x_{2i}+\ldots+x_{ni}=s_i}$$

$$\Rightarrow \boxed{P(D|\theta) = \prod_{i=1}^{d} \theta_i^{s_i} (1-\theta_i)^{n-s_i}} \quad (1)$$

β) Τύπος του Bayes: $P(\theta|D) = \dfrac{P(D|\theta) \cdot P(\theta)}{P(D)}$, με

$$P(D) = \int P(D|\theta) \cdot p(\theta) d\theta \quad (2) \qquad P(D|\theta) \to \text{από } 1^{\text{ο}} \text{ ερώτημα} \quad (4)$$

$$P(\theta_i) = \begin{cases} 0, \{\theta_i < 0\} \cap \{\theta_i > 1\} \\ 1, \quad 0 < \theta_i < 1 \end{cases} \quad (3)$$

Scanned with CamScanner

Άρα (3) $\xRightarrow{(4)}$ $P(D) = \int_0^1 \prod_{i=1}^{d} \theta_i^{s_i} (1-\theta_i)^{n-s_i} \underbrace{p(\theta)}_{1} d\theta$

$$= \prod_{i=1}^{d} \int_0^1 \theta_i^{s_i} (1-\theta_i)^{(n-s_i)} d\theta$$

Χρησιμοποιώντας την ιδιότητα της εκφώνησης:

$$\int_0^1 \theta^m (1-\theta)^m d\theta = \frac{m!\,n!}{(m+n+1)!}$$, έχουμε:

$$P(D) = \prod_{i=1}^{d} \frac{s_i!\,(n-s_i)!}{(n+1)!}$$

Συνολικά, λαμβάνουμε:

$$(1) \Rightarrow P(\theta|D) = \frac{\prod_{i=1}^{d} \theta_i^{s_i} (1-\theta_i)^{(n-s_i)}}{\prod_{i=1}^{d} \frac{s_i!\,(n-s_i)!}{(n+1)!}} \Longrightarrow$$

$$\Rightarrow \boxed{p(\theta|D) = \prod_{i=1}^{d} \frac{(n+1)!}{s_i!\,(n-s_i)!} \theta_i^{s_i} (1-\theta_i)^{(n-s_i)}}$$

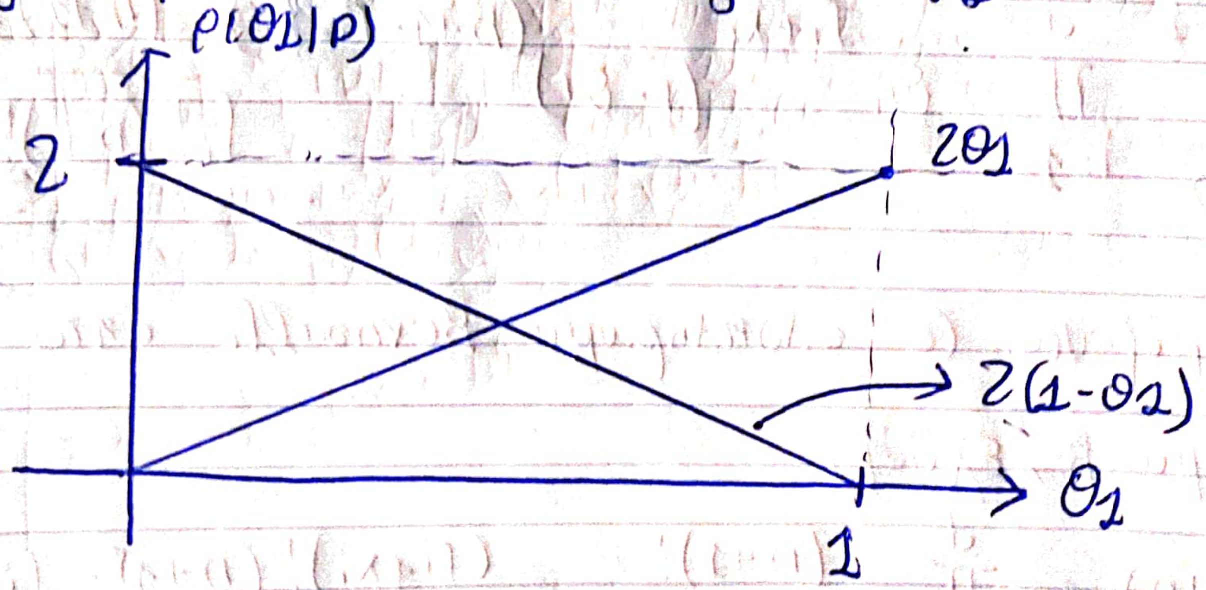
(γ) Για $d=1, n=1$, η εξίσωση του ερωτήματος β γίνεται:

$$P(\theta_1 | D) = \frac{2!}{s_1!(1-s_1)!} \theta_1^{s_1} (1-\theta_1)^{1-s_1}$$

ή

$$P(\theta_1 | D) = \frac{2}{s_1!(1-s_1)!} \theta_1^{s_1} (1-\theta_1)^{(1-s_1)}$$

Οι ζητούμενες πιθανότητες για το s_1:



σ). Έχουμε: $p(x|D) = \int p(x|\theta) \cdot p(\theta|D) d\theta$

$$= \int_0^1 \left(\prod_{i=1}^{d} \theta_i^{x_i} (1-\theta_i)^{(1-x_i)} \right) \cdot \left(\prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!} \cdot \theta_i^{s_i} (1-\theta_i)^{(n-s_i)} \right) d\theta =$$

$$= \int_0^1 \prod_{i=1}^{d} \theta_i^{s_i+x_i} (1-\theta_i)^{(n+1-s_i-x_i)} \frac{(n+1)!}{s_i!(n-s_i)!} d\theta$$

Εφαρμόζοντας το ολοκλήρωμα Bernoulli από το ερώτημα β, έχουμε:

$$p(x|D) = \prod_{i=1}^{d} \frac{(n+1)!}{s_i(n-s_i)!} \frac{(s_i+x_i)!(n+1-s_i-x_i)!}{(n+2)!}$$

$$= \prod_{i=1}^{d} \frac{(s_i+x_i)!(n+1-s_i-x_i)!}{s_i!(n-s_i)!(n+2)}$$

Scanned with CamScanner

ε) Από maximum likelihood, ζητούμε την εκτίμηση $\hat{\theta}$ η οποία μεγιστοποιεί την κατανομή $p(x|\theta) \Rightarrow \frac{d}{d\theta}\left(\ln \prod_{i=1}^{n} p(x_i|\theta)\right) = 0$

$$\Rightarrow \frac{d}{d\theta}\left(\ln p(D|\theta)\right) = \frac{d}{d\theta}\left(\ln \prod_{i=1}^{d} \theta_i^{s_i}(1-\theta_i)^{n-s_i}\right)$$

$$= 0 \Longrightarrow \frac{d}{d\theta}\left(\sum_{i=1}^{d} s_i \ln \theta_i + (n-s_i)\ln(1-\theta_i)\right) = 0$$

$$\Longrightarrow \sum_{i=1}^{d} \frac{s_i}{\theta_i} + \frac{n-s_i}{\theta_i - 1} = 0 \Rightarrow \sum_{i=1}^{d} s_i(\theta_i - 1) + (n-s_i)\theta_i =$$

$$= 0 \Rightarrow \sum_{i=1}^{d}(-s_i + n\theta_i) = 0$$

Άρα, για κάθε $i \in d$, πρέπει: $n\hat{\theta}_i - s_i = 0$ ή

$$\boxed{\hat{\theta}_i = \frac{s_i}{n}}$$

Άσκηση 1.5

(α) Ιδιότητες για trace: $tr[XY] = tr[YX]$

και $tr[YX^T] = X^T Y$

Έτσι, προκύπτει: $\boxed{a^T (Aa) = tr(Aa \cdot a^T) = tr(a^T \cdot A \cdot a)}$

(β) Ανεξαρτησία δειγμάτων: Χρησιμοποιούμε πολυδιάστατη γκαουσιανή

$$p(X|\Sigma) = \prod_{i=1}^{n} p(x_i|\Sigma) =$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left\{-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1} (x_i-\mu)\right\}$$

$$= \frac{1}{(2\pi)^{nd/2} |\Sigma|^{n/2}} \cdot \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (x_i-\mu)^T \Sigma^{-1} (x_i-\mu)\right\}$$

Από ερώτημα (α), έχουμε $a^T A a = tr(A \cdot a \cdot a^T)$

θέτουμε $a = (x_i - \mu)$ και $A = \Sigma^{-1}$ για να πάρουμε:

Scanned with CamScanner

$$p(X|\Sigma) = \frac{1}{(2\pi)^{nd/2} |\Sigma|^{n/2}} \cdot \exp\left\{-\frac{1}{2} tr\left(\sum_{i=1}^{n} \Sigma^{-1} (x_i-\mu) \cdot (x_i-\mu)^T\right)\right\}$$

ή

$$\boxed{p(X|\Sigma) = \frac{1}{(2\pi)^{nd/2} |\Sigma|^{n/2}} \cdot \exp\left\{-\frac{1}{2} tr\left(\Sigma^{-1} \sum_{i=1}^{n} (x_i-\mu)(x_i-\mu)^T\right)\right\}}$$

(γ) Από επιφώνηση, ο εκτιμητής μέγιστης πιθανοφάνειας είναι ο: $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (x_i-\mu)(x_i-\mu)^T$

$$\Longrightarrow \sum_{i=1}^{n} (x_i-\mu)(x_i-\mu)^T = n \cdot \hat{\Sigma}$$

Αντικαθιστώντας την άνωθι σχέση στο αποτέλεσμα του ερωτ. (Β) λαμβάνουμε:

$$p(X|\Sigma) = \frac{1}{(2\pi)^{nd/2} |\Sigma|^{n/2}} \exp\left(-\frac{1}{2} \text{trace}\left(\underbrace{\Sigma^{-1} \cdot \hat{\Sigma} \cdot n}_{A}\right)\right)$$

Scanned with CamScanner

$$= \frac{1}{(2\pi)^{nd/2} \cdot |\hat{\Sigma}|^{n/2}} |\hat{\Sigma}|^{n/2} \cdot \Sigma^{-n/2} \cdot exp\left(-\frac{n}{2} \cdot tr[A]\right)$$

$$= \frac{1}{(2\pi)^{nd/2} |\hat{\Sigma}|^{n/2}} \underbrace{(\Sigma^{-1} \cdot \hat{\Sigma})}_{A}^{n/2} \cdot exp\left(-\frac{n}{2} tr[A]\right)$$

, με: $|A| = \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_d$ και $tr[A] = \lambda_1 + \lambda_2 + \ldots + \lambda_d$
$= \sum_{i=1}^{d} \lambda_i$

Άρα: $$\boxed{p(X|\Sigma) = \frac{1}{(2\pi)^{nd/2} |\hat{\Sigma}|^{n/2}} (\lambda_1 \ldots \lambda_d)^{n/2} exp\left(-\frac{n}{2} \sum_{i=1}^{d} \lambda_i\right)}$$

(δ) Παραγωγίζοντας λαμβάνουμε ότι:

$$\frac{d\, p(X|\Sigma)}{d\lambda_i} = 0 \xRightarrow{i=1} \frac{d\, p(X|\Sigma)}{d\lambda_1} = 0$$

$$\Rightarrow \cancel{\frac{n}{2}} (\lambda_1 \ldots \lambda_d)^{n/2-1} \cdot (\lambda_2 \ldots \lambda_d) \cdot \cancel{exp\left(-\frac{n}{2} \sum_{i=1}^{d} \lambda_i\right)}$$

$$+ (\lambda_1 \ldots \lambda_d)^{n/2} \cdot \cancel{exp\left(-\frac{n}{2} \sum \lambda_i\right)} \cdot \left(-\cancel{\frac{n}{2}}\right) = 0$$

$$\Rightarrow \lambda_1^{n/2-1} \cdot \cancel{(\lambda_2 \dots \lambda_d)^{n/2}} - \lambda_1^{n/2} \cdot \cancel{(\lambda_2 \dots \lambda_d)^{n/2}} = 0$$

$$\Rightarrow \lambda_1 - 1 = 0 \Rightarrow \boxed{\lambda_1 = 0}$$

, Ομοίως αποδεικνύεται ότι για κάθε d_i, μεγιστοποίηση της πιθανοφάνειας πραγματοποιείται για:

$$\boxed{\lambda_1 = \lambda_2 = \dots = \lambda_d = 1}$$

Scanned with CamScanner