<u>ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ</u>

ΕΡΓΑΣΙΑ 2Η ΑΚΑΔΗΜΑΙΚΟ ΕΤΟΣ : 2018-2019

ΦΟΙΤΗΤΗΣ: ΙΑΚΩΒΟΣ ΕΥΔΑΙΜΩΝ 3130059

Η εργασία αυτή επικεντρώνεται στην υλοποίηση του αλγορίθμου EM(expectation maximization) για την μεγιστοποίηση της πιθανοφάνειας σε μία μίξη Gaussian κατανομών της παρακάτω μορφής:

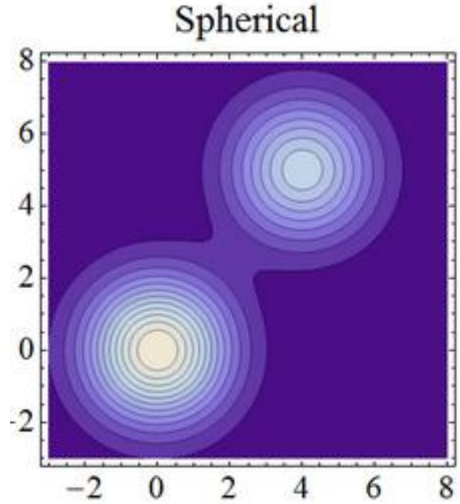
$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k \prod_{d=1}^{D} \frac{1}{\sqrt{2\,\pi\,\sigma_k^2}}\, e^{-\frac{1}{2\,\sigma_k^2}(x_d-\mu_{kd})^2} =$$

$$= \quad \sum_{k=1}^{K} \pi_k \prod_{d=1}^{D} N(x_d|\mu_{kd}, {\sigma_k}^2)$$

Όπου $N(x_d|\mu_{kd}, {\sigma_k}^2) = \frac{1}{\sqrt{2\,\pi\,\sigma_k^2}}\, e^{-\frac{1}{2\,\sigma_k^2}(x_d-\mu_{kd})^2}$ εδώ.

Σε αυτή την εργασία θα επεξεργαστούμε εικόνες όπου ισχύει ότι κάθε δεδομένο είναι ένα pixel και συγκεκριμένα ένα 3-διάστατο διάνυσμα $x_n = [x_r, x_g, x_b]$ όπου x_r, x_g, x_b είναι οι τρεις τιμές για το εκάστοτε κανάλι χρώματος .

Από τον τύπος της παραπάνω μίξης κατανομών διαπιστώνουμε ότι οι συνιστώσες κάθε x_n διανύσματος(δεδομένου/εδώ pixel) είναι ανεξάρτητες, δηλαδή για την εργασία μας τα κανάλια RGB κάθε pixel. Οι παραπάνω Gaussian συνιστώσες κατανομές έχουν πίνακα συμμεταβλητότητας/συνδιακύμανσης (covariance matrix) σφαιρικό, που σημαίνει ότι έχει όλες τις εκτός διαγωνίου καταχωρήσεις ίσες με μηδέν. Από κάτω παραθέτεται η εικόνα με την μορφή των σφαιρικών συνδιακυμάνσεων.



Έχουμε την υπόθεση ότι έχουμε μία τυχαία μεταβλητή z, η οποία επιλέγεται με βάση τις πιθανότητες π_k και ουσιαστικά ταυτοποιεί την ομάδα k της οποίας θα είναι μέλος κάθε δεδομένο x_n που θα παραχθεί. Έτσι έχουμε μία K-διαστάσεων(K = αριθμός ομάδων) δυαδική τυχαία μεταβλητή z και σε μία αναπαράσταση 1-του-K στην οποία ένα στοιχείο $z_k = 1$ και όλα τα άλλα ίσα με μηδέν. Οπότε, υπάρχουν K πιθανές καταστάσεις για το διάνυσμα z ώστε ένα στοιχείο να είναι μη-μηδενικό. Άρα $z_k \in \{0, 1\}$ και $\sum_k z_k = 1$. Ουσιαστικά η πιθανότητα ενός στοιχείου $z_k = 1$ είναι ίση με το π_k το οποίο συμβολίζει την εκ των προτέρων πιθανότητα της ομάδας k.

$$p(z_k = 1) = \pi_k$$

$$\text{και } \sum_{k=1}^{K} \pi_k = 1 \; \kappa\alpha\iota \; 0 \leq \pi_k \leq 1$$

Το πιθανοτικό μοντέλο ομαδοποίησης (k ομάδων) υποθέτει ότι κάθε δεδομένο παράγεται από τις εξής συνθήκες. Καθορίζεται η ομάδα από την οποία θα παραχθεί το δεδομένο δηλαδή καθορίζεται η τιμή της τυχαίας μεταβλητής z.

$$p(z) = \prod_{k=1}^{K} \pi_k^{z_k} \quad (1.1)$$

Και καθορίζεται η παραγωγή του δεδομένου δοθέντος του z (θεωρούμε ότι οι διαστάσεις του δεδομένου είναι υπό συνθήκη ανεξάρτητες):

$$p(x|z_k = 1) = \prod_{d=1}^{D} \mathcal{N}(x_d|\boldsymbol{\mu_{kd}}, \sigma_{\boldsymbol{k}}{}^2) \ (1.2)$$

Το οποίο μπορεί να γραφτεί και στην μορφή:

$$p(\mathrm{x}|\mathrm{z}) = \prod_{k=1}^{K} \prod_{d=1}^{D} \mathcal{N}(x_d|\boldsymbol{\mu_{kd}}, \sigma_{\boldsymbol{k}}{}^2)^{z_k}$$

Από τις εξισώσεις (1.1) και (1.2) ορίζουμε την ολική κατανομή του X/μίξη κατανομών η οποία υπολογίζεται εφαρμόζοντας τον κανόνα του αθροίσματος και περιθωριοποιώντας τον άγνωστο z και είναι ίδια με τον τύπο της εκφώνησης, επιβεβαιώνοντάς την.

$$p(\mathrm{x}) = \sum_{\mathrm{z}} p(\mathrm{z})p(\mathrm{x}|\mathrm{z}) = \sum_{k=1}^{K} \pi_k \prod_{d=1}^{D} N(x_d|\mu_{kd}, \sigma_k{}^2)$$

και $\sum_{k=1}^{K} \mathrm{\pi}_k = 1 \ \kappa\alpha\iota \ 0 \leq \pi_k \leq 1$

Όπου οι παράμετροι του μοντέλου που πρέπει να εκτιμήσουμε είναι οι : $\{\pi_k, \mu_{kd}, \sigma_k{}^2\}_{k=1}^{K}$

Έστω ότι υπολογίσαμε τις παραπάνω παραμέτρους θα πρέπει επίσης να εισάγουμε και ένα συμπέρασμα για την άγνωστη τυχαία/κρυμμένη μεταβλητή z για αυτό το λόγο πρέπει να υπολογίσουμε την υπό συνθήκη πιθανότητα του z δοθέντος του x χρησιμοποιώντας το θεώρημα του Bayes

$$\gamma(z_k) = p(z_k = 1|x) = \frac{p(x|z_k = 1)p(z_k=1)}{p(x)} = \frac{\mathrm{\pi_k} \prod_{d=1}^{D} \mathcal{N}\left(x_d \middle| \mu_{\boldsymbol{kd}}, \sigma_{\boldsymbol{k}}{}^2\right)}{\sum_{j=1}^{K} \pi_j \prod_{d=1}^{D} N\left(x_d \middle| \mu_{jd}, \sigma_j{}^2\right)}$$

και $\gamma(z_{nk}) = \dfrac{\mathrm{\pi_k} \prod_{d=1}^{D} \mathcal{N}\left(x_{nd} \middle| \mu_{\boldsymbol{kd}}, \sigma_{\boldsymbol{k}}{}^2\right)}{\sum_{j=1}^{K} \pi_j \prod_{d=1}^{D} N\left(x_{nd} \middle| \mu_{jd}, \sigma_j{}^2\right)}$

$X = \{x_1, x_2, x_3, \ldots, x_N\}$ δηλαδή αποτελεί όλα τα παραδείγματα x_n που έχουμε, δηλαδή το X είναι το σύνολο των παραδειγμάτων εκπαίδευσης, όπου κάθε $x_n = \{x_{n1}, x_{n2}, \ldots, x_{nK}\}$ έχει παραχθεί ανεξάρτητα από την μίξη.

Οπότε η από κοινού κατανομή που έχουμε είναι η εξής:

$$p(X|\pi, \mu, \sigma^2) = \prod_{n=1}^{N} \sum_{k=1}^{K} \pi_k \prod_{d=1}^{D} \mathcal{N}(x_{nd}|\mu_{kd}, {\sigma_k}^2)$$

και η λογαριθμική πιθανοφάνεια η εξής:

$$\mathcal{L}(\pi, \mu, \sigma^2) = \sum_{n=1}^{N} \log \sum_{k=1}^{K} \pi_k \prod_{d=1}^{D} \mathcal{N}(x_{nd}|\mu_{kd}, {\sigma_k}^2)$$

Για την εκτίμηση των παραμέτρων $\{\pi_k, \mu_{kd}, {\sigma_k}^2\}_{k=1}^{K}$ πρέπει να μεγιστοποιήσουμε την λογαριθμική πιθανοφάνεια ως προς κάθε μία από τις παραμέτρους που ψάχνουμε. Η μεγιστοποίηση γίνεται με την βοήθεια του αλγορίθμου EM.

Αρχικά μεγιστοποιούμε την $\mathcal{L}(\pi, \mu, \sigma^2)$ ως προς την παράμετρο μ_{kd}(μέση τιμή κάθε Γκαουσιανής κατανομής) και εργαζόμαστε ως εξής:

$$\frac{\partial \mathcal{L}(\pi, \mu, \sigma^2)}{\partial \mu_{kd}} =$$

$$= \sum_{n=1}^{N} \frac{1}{\sum_{j=1}^{K} \pi_j \prod_{d=1}^{D} \mathcal{N}\left(x_{nd}\middle|\mu_{jd}, {\sigma_j}^2\right)} \frac{\partial\left(\sum_{k=1}^{K} \pi_k \prod_{d=1}^{D} \mathcal{N}(x_{nd}|\mu_{kd}, {\sigma_k}^2)\right)}{\partial \mu_{kd}}$$

$$= \sum_{n=1}^{N} \frac{1}{\sum_{j=1}^{K} \pi_j \prod_{d=1}^{D} \mathcal{N}\left(x_{nd}\middle|\mu_{jd}, {\sigma_j}^2\right)} \pi_k \frac{\partial\left(\prod_{d=1}^{D} \mathcal{N}(x_{nd}|\mu_{kd}, {\sigma_k}^2)\right)}{\partial \mu_{kd}} \qquad (1)$$

Όπου, το:

$$\frac{\partial\left(\prod_{d=1}^{D}\mathcal{N}(\mathrm{x_{nd}}|\mathrm{\mu_{kd}},\mathrm{\sigma_k}^2)\right)}{\partial\mathrm{\mu_{kd}}} =$$

$$= \prod_{m=1}^{d-1}\mathcal{N}(\mathrm{x_{nm}}|\mathrm{\mu_{km}},\mathrm{\sigma_k}^2)\prod_{p=d+1}^{D}\mathcal{N}\left(\mathrm{x_{np}}\middle|\mathrm{\mu_{kp}},\mathrm{\sigma_k}^2\right)\frac{\partial\left(\mathcal{N}(\mathrm{x_{nd}}|\mathrm{\mu_{kd}},\mathrm{\sigma_k}^2)\right)}{\partial\mathrm{\mu_{kd}}}$$

$$= \prod_{m=1}^{d-1}\mathcal{N}(\mathrm{x_{nm}}|\mathrm{\mu_{km}},\mathrm{\sigma_k}^2)\prod_{p=d+1}^{D}\mathcal{N}\left(\mathrm{x_{np}}\middle|\mathrm{\mu_{kp}},\mathrm{\sigma_k}^2\right)\frac{\partial\left(\frac{1}{\sqrt{2\,\pi\,\sigma_k^2}}\;e^{-\frac{1}{2\,\sigma_k^2}(x_{nd}-\mu_{kd})^2}\right)}{\partial\mathrm{\mu_{kd}}}$$

$$= \prod_{m=1}^{d-1}\mathcal{N}(\mathrm{x_{nm}}|\mathrm{\mu_{km}},\mathrm{\sigma_k}^2)\prod_{p=d+1}^{D}\mathcal{N}\left(\mathrm{x_{np}}\middle|\mathrm{\mu_{kp}},\mathrm{\sigma_k}^2\right)\frac{1}{\sqrt{2\,\pi\,\sigma_k^2}}\;e^{-\frac{1}{2\,\sigma_k^2}(x_{nd}-\mu_{kd})^2}$$

$$\frac{\partial\left(-\frac{1}{2\,\sigma_k^2}(x_{nd}-\mu_{kd})^2\right)}{\partial\mathrm{\mu_{kd}}} = \mathcal{N}(\mathrm{x_{nd}}|\mathrm{\mu_{kd}},\mathrm{\sigma_k}^2)\prod_{m=1}^{d-1}\left(\mathcal{N}(\mathrm{x_{nm}}|\mathrm{\mu_{km}},\mathrm{\sigma_k}^2)\right)$$

$$\prod_{p=d+1}^{D}\left(\mathcal{N}\left(\mathrm{x_{np}}\middle|\mathrm{\mu_{kp}},\mathrm{\sigma_k}^2\right)\right)\;\frac{\partial(-\frac{1}{2\,\sigma_k^2}(x_{nd}-\mu_{kd})^2)}{\partial\mathrm{\mu_{kd}}} =$$

$$= \left(\prod_{d=1}^{D}\left(\mathcal{N}(\mathrm{x_{nd}}|\mathrm{\mu_{kd}},\mathrm{\sigma_k}^2)\right)\right)\frac{1}{\sigma_k^2}(x_{nd}-\mu_{kd})\ (2)$$

Εφαρμόζοντας την σχέση (2) *στην* (1) *και το γεγονός ότι*

$$\gamma(z_{nk}) = \frac{\pi_k\prod_{d=1}^{D}\mathcal{N}(x_{nd}|\boldsymbol{\mu_{kd}},\boldsymbol{\sigma_k}^2)}{\sum_{j=1}^{K}\pi_j\prod_{d=1}^{D}N\left(x_{nd}\middle|\mu_{jd},\sigma_j{}^2\right)}\ \textit{τότε καταλήγουμε ότι:}$$

$$(1)\underset{(2)}{\Rightarrow}\sum_{n=1}^{N}\frac{\pi_k\prod_{d=1}^{D}\mathcal{N}(\mathrm{x_{nd}}|\mathrm{\mu_{kd}},\mathrm{\sigma_k}^2)}{\sum_{j=1}^{K}\pi_j\prod_{d=1}^{D}\mathcal{N}\left(\mathrm{x_{nd}}\middle|\mathrm{\mu_{jd}},\mathrm{\sigma_j}^2\right)}\frac{1}{\sigma_k^2}(x_{nd}-\mu_{kd}) =$$

$$= \sum_{n=1}^{N}\gamma(z_{nk})\,\frac{1}{\sigma_k^2}(x_{nd}-\mu_{kd})$$

και αφού θέλουμε $\frac{\partial \mathcal{L}(\pi, \mu, \sigma^2)}{\partial \mu_{kd}} = 0 \Rightarrow$

$$\sum_{n=1}^{N} \gamma(z_{nk}) \frac{1}{\sigma_k^2}(x_{nd} - \mu_{kd}) = 0 \Rightarrow$$

$$\sum_{n=1}^{N} \gamma(z_{nk}) (x_{nd} - \mu_{kd}) = 0 \Rightarrow$$

$$\sum_{n=1}^{N} \gamma(z_{nk}) x_{nd} = \sum_{n=1}^{N} \gamma(z_{nk}) \mu_{kd} \Rightarrow$$

$$(\boldsymbol{\mu_{kd}})^{\boldsymbol{new}} = \frac{\sum_{\boldsymbol{n=1}}^{\boldsymbol{N}} \boldsymbol{\gamma}(\boldsymbol{z_{nk}}) \boldsymbol{x_{nd}}}{\sum_{\boldsymbol{n=1}}^{\boldsymbol{N}} \boldsymbol{\gamma}(\boldsymbol{z_{nk}})}$$

Και αυτός είναι ο τελικός μας τύπος για την ενημέρωση της παραμέτρου $\boldsymbol{\mu_{kd}}$.

Έπειτα, μεγιστοποιούμε την $\mathcal{L}(\pi, \mu, \sigma^2)$ ως προς την παράμετρο σ_k^2 (διακύμανση κάθε Γκαουσιανής κατανομής) και εργαζόμαστε ως εξής:

$$\frac{\partial \mathcal{L}(\pi, \mu, \sigma^2)}{\partial \sigma_k^2} =$$

$$= \sum_{n=1}^{N} \frac{1}{\sum_{j=1}^{K} \pi_j \prod_{d=1}^{D} \mathcal{N}\left(x_{nd} \middle| \mu_{jd}, {\sigma_j}^2\right)} \frac{\partial\left(\sum_{k=1}^{K} \pi_k \prod_{d=1}^{D} \mathcal{N}(x_{nd} | \mu_{kd}, {\sigma_k}^2)\right)}{\partial \sigma_k^2}$$

$$= \sum_{n=1}^{N} \frac{1}{\sum_{j=1}^{K} \pi_j \prod_{d=1}^{D} \mathcal{N}\left(x_{nd} \middle| \mu_{jd}, {\sigma_j}^2\right)} \pi_k \frac{\partial\left(\prod_{d=1}^{D} \mathcal{N}(x_{nd} | \mu_{kd}, {\sigma_k}^2)\right)}{\partial \sigma_k^2} \quad (3)$$

Όπου, το:

$$\frac{\partial\left(\prod_{d=1}^{D}\mathcal{N}(\mathrm{x_{nd}}|\mathrm{\mu_{kd}},\mathrm{\sigma_k}^2)\right)}{\partial\sigma_k^2}=$$

$$=\sum_{d=1}^{D}\left(\prod_{m=1}^{\mathrm{d-1}}\mathcal{N}(\mathrm{x_{nm}}|\mathrm{\mu_{km}},\mathrm{\sigma_k}^2)\prod_{\mathrm{p=d+1}}^{\mathrm{D}}\mathcal{N}\left(\mathrm{x_{np}}\middle|\mathrm{\mu_{kp}},\mathrm{\sigma_k}^2\right)\frac{\partial\left(\mathcal{N}(\mathrm{x_{nd}}|\mathrm{\mu_{kd}},\mathrm{\sigma_k}^2)\right)}{\partial\sigma_k^2}\right)$$

$$=\sum_{d=1}^{D}\left(\prod_{m=1}^{\mathrm{d-1}}\mathcal{N}(\mathrm{x_{nm}}|\mathrm{\mu_{km}},\mathrm{\sigma_k}^2)\prod_{\mathrm{p=d+1}}^{\mathrm{D}}\mathcal{N}\left(\mathrm{x_{np}}\middle|\mathrm{\mu_{kp}},\mathrm{\sigma_k}^2\right)\frac{\partial\left(\frac{1}{\sqrt{2\,\pi\,\sigma_k^2}}\;e^{-\frac{1}{2\,\sigma_k^2}(x_{nd}-\mu_{kd})^2}\right)}{\partial\sigma_k^2}\right)$$

Σχέση (4.1)

Και επίσης το :

$$\frac{\partial\left(\frac{1}{\sqrt{2\,\pi\,\sigma_k^2}}\;e^{-\frac{1}{2\,\sigma_k^2}(x_{nd}-\mu_{kd})^2}\right)}{\partial\sigma_k^2}=\frac{\partial\left(\frac{1}{\sqrt{2\,\pi\,\sigma_k^2}}\right)}{\partial\sigma_k^2}\;e^{-\frac{1}{2\,\sigma_k^2}(x_{nd}-\mu_{kd})^2}\quad+$$

$$\frac{\partial\left(e^{-\frac{1}{2\,\sigma_k^2}(x_{nd}-\mu_{kd})^2}\right)}{\partial\sigma_k^2}\frac{1}{\sqrt{2\,\pi\,\sigma_k^2}}=\left(-\frac{1}{2\sqrt{2\pi}\;(\sigma_k^2)^{\frac{3}{2}}}\right)e^{-\frac{1}{2\,\sigma_k^2}(x_{nd}-\mu_{kd})^2}+$$

$$\frac{1}{\sqrt{2\,\pi\,\sigma_k^2}}\;e^{-\frac{1}{2\,\sigma_k^2}(x_{nd}-\mu_{kd})^2}\;\left(\frac{1}{2}(x_{nd}-\mu_{kd})^2\;\;(\sigma_k^2)^{-2}\right)=$$

$$=\left(-\frac{1}{2\sqrt{2\pi\sigma_k^2}}\;\frac{1}{\sigma_k^2}\right)e^{-\frac{1}{2\,\sigma_k^2}(x_{nd}-\mu_{kd})^2}+$$

$$\frac{1}{\sqrt{2\,\pi\,\sigma_k^2}}\;e^{-\frac{1}{2\,\sigma_k^2}(x_{nd}-\mu_{kd})^2}\;\left(\frac{1}{2}(x_{nd}-\mu_{kd})^2\;\;(\sigma_k^2)^{-2}\right)=$$

$$=\left(\frac{1}{\sqrt{2\,\pi\,\sigma_k^2}}\;e^{-\frac{1}{2\,\sigma_k^2}(x_{nd}-\mu_{kd})^2}\right)\left(-\frac{1}{\sigma_k^2}+\frac{1}{2}(x_{nd}-\mu_{kd})^2\;\;(\sigma_k^2)^{-2}\right)=$$

$$= \mathcal{N}(\mathrm{x_{nd}}|\mathrm{\mu_{kd}}, \mathrm{\sigma_k}^2) \left(-\frac{1}{\sigma_k^2} + \frac{1}{2}(x_{nd} - \mu_{kd})^2 \ (\sigma_k^2)^{-2}\right) \quad (4.2)$$

Εφαρμόζοντας την σχέση (4.2) στην σχέση (4.1) τότε έχουμε ότι :

$$(4.1) \underset{(4.2)}{\Longrightarrow} \sum_{d=1}^{D} \mathcal{N}(\mathrm{x_{nd}}|\mathrm{\mu_{kd}}, \mathrm{\sigma_k}^2) \left(-\frac{1}{\sigma_k^2} + \frac{1}{2}(x_{nd} - \mu_{kd})^2 \ (\sigma_k^2)^{-2}\right) \prod_{m=1}^{\mathrm{d}-1} \mathcal{N}(\mathrm{x_{nm}}|\mathrm{\mu_{km}}, \mathrm{\sigma_k}^2) \prod_{\mathrm{p=d+1}}^{\mathrm{D}} \mathcal{N}\left(\mathrm{x_{np}}\middle|\mathrm{\mu_{kp}}, \mathrm{\sigma_k}^2\right) =$$

$$= \sum_{d=1}^{D} \prod_{\mathrm{d=1}}^{\mathrm{D}} \left(\mathcal{N}(\mathrm{x_{nd}}|\mathrm{\mu_{kd}}, \mathrm{\sigma_k}^2)\right) \left(-\frac{1}{\sigma_k^2} + \frac{1}{2}(x_{nd} - {\mu_{kd}}^{new})^2 \ (\sigma_k^2)^{-2}\right) (4.3)$$

Εφαρμόζοντας την σχέση (4.3) *στην* (3) *και το γεγονός ότι*

$$\gamma(z_{nk}) = \frac{\mathrm{\pi_k} \prod_{d=1}^{D} \mathcal{N}(x_{nd}|\boldsymbol{\mu_{kd}}, \boldsymbol{\sigma_k}^2)}{\sum_{j=1}^{K} \pi_j \prod_{d=1}^{D} N\left(x_{nd}\middle|\mu_{jd}, {\sigma_j}^2\right)} \ \textit{τότε καταλήγουμε ότι:}$$

$$(3) \underset{(4.3)}{\Longrightarrow} \sum_{n=1}^{N} \left(\sum_{d=1}^{D} \left(\frac{\mathrm{\pi_k} \prod_{\mathrm{d=1}}^{\mathrm{D}} \mathcal{N}(\mathrm{x_{nd}}|\mathrm{\mu_{kd}}, \mathrm{\sigma_k}^2)}{\sum_{\mathrm{j=1}}^{\mathrm{K}} \mathrm{\pi}_j \prod_{\mathrm{d=1}}^{\mathrm{D}} \mathcal{N}\left(\mathrm{x_{nd}}\middle|\mathrm{\mu_{jd}}, \mathrm{\sigma_j}^2\right)} \left(-\frac{1}{\sigma_k^2} + \frac{1}{2}(x_{nd} - {\mu_{kd}}^{new})^2 \ (\sigma_k^2)^{-1} \ (\sigma_k^2)^{-1}\right)\right)\right) =$$

$$= \sum_{n=1}^{N} \sum_{d=1}^{D} \gamma(z_{nk}) \left(-\frac{1}{\sigma_k^2} + \frac{1}{2}(x_{nd} - {\mu_{kd}}^{new})^2 \ (\sigma_k^2)^{-1} \ (\sigma_k^2)^{-1}\right)$$

και αφού θέλουμε $\dfrac{\partial \mathcal{L}(\mathrm{\pi}, \mathrm{\mu}, \mathrm{\sigma}^2)}{\partial \sigma_k^2} = 0 \Rightarrow$

$$\sum_{n=1}^{N} \sum_{d=1}^{D} \gamma(z_{nk}) \left(-\frac{1}{\sigma_k^2} + \frac{1}{2}(x_{nd} - {\mu_{kd}}^{new})^2 \ (\sigma_k^2)^{-1} \ (\sigma_k^2)^{-1}\right) = 0 \Rightarrow$$

$$\sum_{n=1}^{N}\sum_{d=1}^{D}\gamma(z_{nk})\left(1-\frac{1}{2}(x_{nd}-\mu_{kd}{}^{new})^2\,(\sigma_k^2)^{-1}\right)=0\Rightarrow$$

$$\sum_{n=1}^{N}\sum_{d=1}^{D}\gamma(z_{nk})\,\left(\sigma_k^2-\frac{1}{2}(x_{nd}-\mu_{kd}{}^{new})^2\right)=0\Rightarrow$$

$$-\frac{1}{2}\sum_{n=1}^{N}\sum_{d=1}^{D}\gamma(z_{nk})\,(x_{nd}-\mu_{kd}{}^{new})^2=-\sum_{n=1}^{N}\gamma(z_{nk})\,(\sigma_k^2)\Rightarrow$$

$$\sum_{n=1}^{N}\sum_{d=1}^{D}\gamma(z_{nk})\,(x_{nd}-\mu_{kd}{}^{new})^2=\sum_{n=1}^{N}\sum_{d=1}^{D}\gamma(z_{nk})\,(\sigma_k^2)\Rightarrow$$

$$\sum_{n=1}^{N}\sum_{d=1}^{D}\gamma(z_{nk})\,(x_{nd}-\mu_{kd}{}^{new})^2=D\sum_{n=1}^{N}\gamma(z_{nk})\,(\sigma_k^2)\Rightarrow$$

$$\boldsymbol{(\sigma_k^2)^{new}=\frac{\sum_{n=1}^{N}\sum_{d=1}^{D}\gamma(z_{nk})\,(x_{nd}-\mu_{kd}{}^{new})^2}{D\sum_{n=1}^{N}\gamma(z_{nk})}}$$

Όπου αυτός είναι ο τελικός μας τύπος για την ενημέρωση της παραμέτρου $\boldsymbol{\sigma_k^2}$.

Τέλος, μεγιστοποιούμε την $\mathcal{L}(\pi,\mu,\sigma^2)$ ως προς την παράμετρο π_k. Εδώ λαμβάνουμε υπόψιν τον περιορισμό: $\sum_{k=1}^{K}\pi_k\ =1$ που απαιτεί τους συντελεστής μίξης να αθροίζουν στο ένα. Αυτό ικανοποιείται χρησιμοποιώντας έναν Lagrange πολλαπλασιαστή λ και μεγιστοποιώντας την παρακάτω ποσότητα:

$$\mathcal{L}_{lagrange}(\pi,\mu,\sigma^2)=\mathcal{L}(\pi,\mu,\sigma^2)+\lambda\left(\sum_{k=1}^{K}\pi_k-1\right)=$$

$$=\sum_{n=1}^{N}\log\sum_{k=1}^{K}\pi_k\prod_{d=1}^{D}\mathcal{N}(x_{nd}|\mu_{kd},\sigma_k{}^2)\ +\lambda\left(\sum_{k=1}^{K}\pi_k-1\right)$$

Σχέση (6)

$$\frac{\partial \mathcal{L}_{lagrange}(\pi, \mu, \sigma^2)}{\partial \pi_k} =$$

$$= \left(\sum_{n=1}^{N} \frac{1}{\sum_{j=1}^{K} \pi_j \prod_{d=1}^{D} \mathcal{N}\left(x_{nd} \middle| \mu_{jd}, {\sigma_j}^2\right)} \frac{\partial\left(\sum_{k=1}^{K} \pi_k \prod_{d=1}^{D} \mathcal{N}(x_{nd}|\mu_{kd}, {\sigma_k}^2)\right)}{\partial \pi_k}\right)$$

$$+ \ \lambda \frac{\partial(\sum_{k=1}^{K} \pi_k - 1)}{\partial \pi_k} \ =$$

$$= \left(\sum_{n=1}^{N} \frac{\prod_{d=1}^{D} \mathcal{N}(x_{nd}|\mu_{kd}, {\sigma_k}^2)}{\sum_{j=1}^{K} \pi_j \prod_{d=1}^{D} \mathcal{N}\left(x_{nd} \middle| \mu_{jd}, {\sigma_j}^2\right)}\right) + \ \lambda$$

Και θέλουμε το $\frac{\partial \mathcal{L}_{lagrange}(\pi,\mu,\sigma^2)}{\partial \pi_k} = 0$ οπότε :

$$\left(\sum_{n=1}^{N} \frac{\prod_{d=1}^{D} \mathcal{N}(x_{nd}|\mu_{kd}, {\sigma_k}^2)}{\sum_{j=1}^{K} \pi_j \prod_{d=1}^{D} \mathcal{N}\left(x_{nd} \middle| \mu_{jd}, {\sigma_j}^2\right)}\right) + \ \lambda \ = \ 0 \Rightarrow$$

Πολλαπλασιάζοντας και τα 2 μέλη με το π_k και αθροίζοντας για κάθε k λαμβάνοντας υπόψιν τον περιορισμό: $\sum_{k=1}^{K} \pi_k = 1$ βρίσκουμε ότι $\lambda = -N$ καθώς $\sum_{k=1}^{K} \pi_k \lambda = -\sum_{k=1}^{K} \sum_{n=1}^{N} \gamma(z_{nk})$ όπου $\sum_{k=1}^{K} \sum_{n=1}^{N} \gamma(z_{nk}) = N$ άρα $\lambda = -N$.
Χρησιμοποιούμε αυτό για να εξαλείψουμε το λ και το γεγονός ότι $\gamma(z_{nk}) = \frac{\pi_k \prod_{d=1}^{D} \mathcal{N}\left(x_{nd} \middle| \boldsymbol{\mu_{kd}}, \boldsymbol{\sigma_k}^2\right)}{\sum_{j=1}^{K} \pi_j \prod_{d=1}^{D} N\left(x_{nd} \middle| \mu_{jd}, {\sigma_j}^2\right)}$ τότε καταλήγουμε ότι:

$$\left(\sum_{n=1}^{N} \frac{\pi_k \prod_{d=1}^{D} \mathcal{N}(x_{nd}|\mu_{kd}, {\sigma_k}^2)}{\sum_{j=1}^{K} \pi_j \prod_{d=1}^{D} \mathcal{N}\left(x_{nd} \middle| \mu_{jd}, {\sigma_j}^2\right)}\right) + \ \lambda \pi_k \ = \ 0 \Rightarrow$$

$$\left(\sum_{n=1}^{N} \gamma(z_{nk})\right) + \ \lambda \pi_k \ = \ 0 \Rightarrow$$

$$\left(\sum_{n=1}^{N} \gamma(z_{nk})\right) \ = \ -\lambda \pi_k \xRightarrow{\lambda=-N}$$

$$\frac{\sum_{\mathrm{n=1}}^{\mathrm{N}} \gamma(z_{nk})}{N} = \pi_k$$

Τα ίδια αποτελέσματα θα είχαμε άμα είχαμε πάρει την expected λογαριθμική complete πιθανοφάνεια για να αποδείξουμε τους παραπάνω τύπους, δηλαδή την :

$$Q\left(\theta, \theta^{(t)}\right) = \sum_{k=1}^{K} \left(\sum_{n=1}^{N} \gamma(z_{nk}) \right) \log \pi_k + \sum_{k=1}^{K} \left(\sum_{n=1}^{N} \gamma(z_{nk}) \sum_{\mathrm{d=1}}^{\mathrm{D}} \log \mathcal{N}\left(\mathbf{x_n} | \boldsymbol{\mu_k}, \Sigma_k\right) \right)$$

LOGSUMEXP TRICK

Για να έχουμε ευσταθείς αριθμητικούς υπολογισμούς στην εργασία μας χρησιμοποιήσαμε το logsumexp trick. Όπου το $\gamma(z_{nk})$ γίνεται ίσο με:

$$\gamma(z_{nk}) = \frac{e^{f_k - m}}{\sum_{j=1}^{K} e^{f_j - m}}$$

Όπου $f_k = e^{\log \pi_k + \sum_{d=1}^{D} log\mathcal{N}\left(x_{nd} | \mu_{\boldsymbol{kd}}, \sigma_{\boldsymbol{k}}{}^2\right)}$

Όπου $log\mathcal{N}(x_{nd} | \mu_{\boldsymbol{kd}}, \sigma_{\boldsymbol{k}}{}^2) = -\frac{1}{2} log(2\pi\sigma_{\boldsymbol{k}}{}^2) - \frac{1}{2\,\sigma_k^2}(x_d - \mu_{kd})^2$

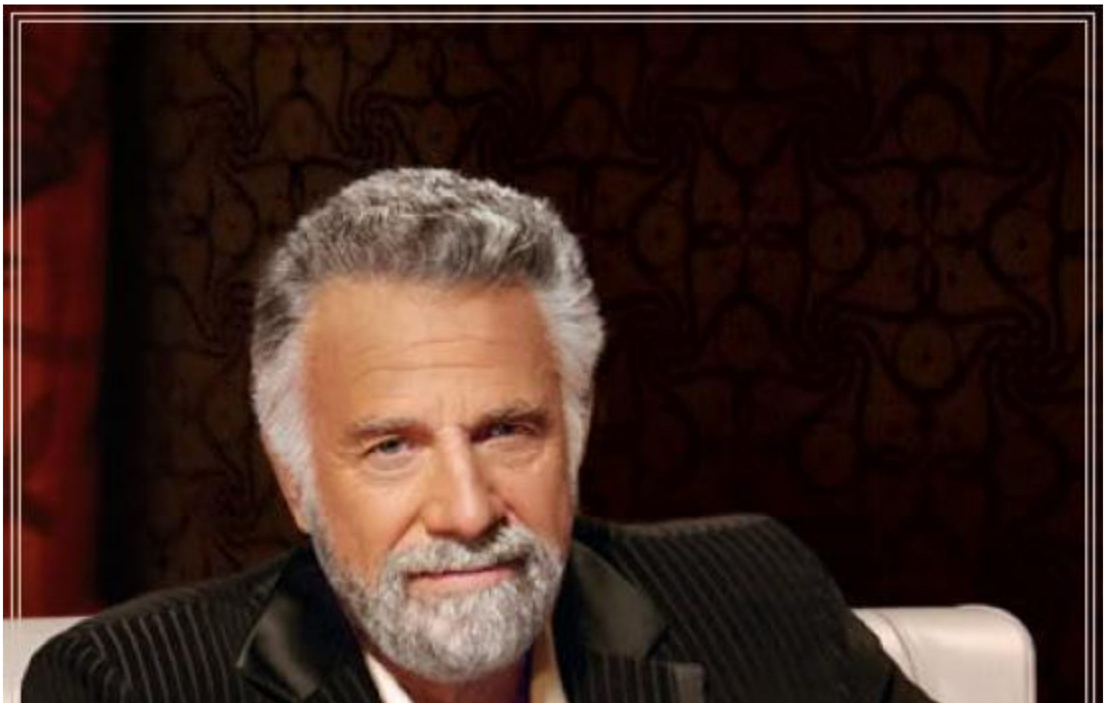
Και $m = \max(f_1, f_2, \ldots, f_K)$

Επίσης, χρησιμοποιούμε logsumexp trick και για τον υπολογισμό της λογαριθμικής πιθανοφάνειας η οποία γίνεται :

$$\mathcal{L}(\pi, \mu, \sigma^2) = \sum_{\mathrm{n=1}}^{\mathrm{N}} m + \log \sum_{\mathrm{k=1}}^{\mathrm{K}} e^{f_k - m}$$

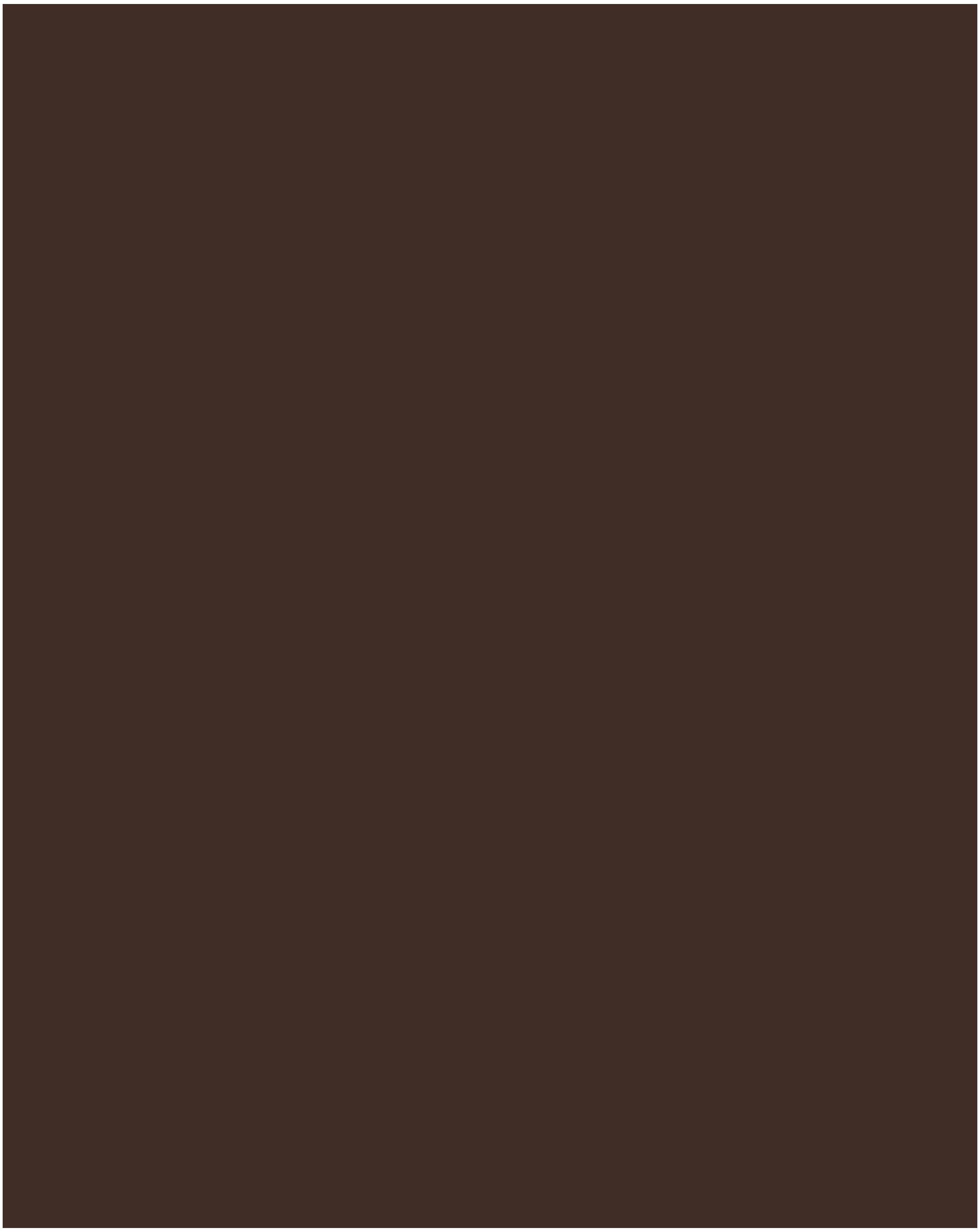
<u>ΠΑΡΑΔΕΙΓΜΑΤΑ ΥΛΟΠΟΙΗΣΗΣ ΕΜ</u>

Υλοποιούμε τον ΕΜ για K=[1,2,4,8,16,32] (όπου K είναι ο αριθμός των ομάδων) και πάνω στην εξής φωτογραφία που βρίσκεται στον φάκελο Image στο παραδοτέο project.



--Αρχική φωτογραφία—

Για αριθμό ομάδων(K) ίσο με 1 έχουμε το ακόλουθο αποτέλεσμα:

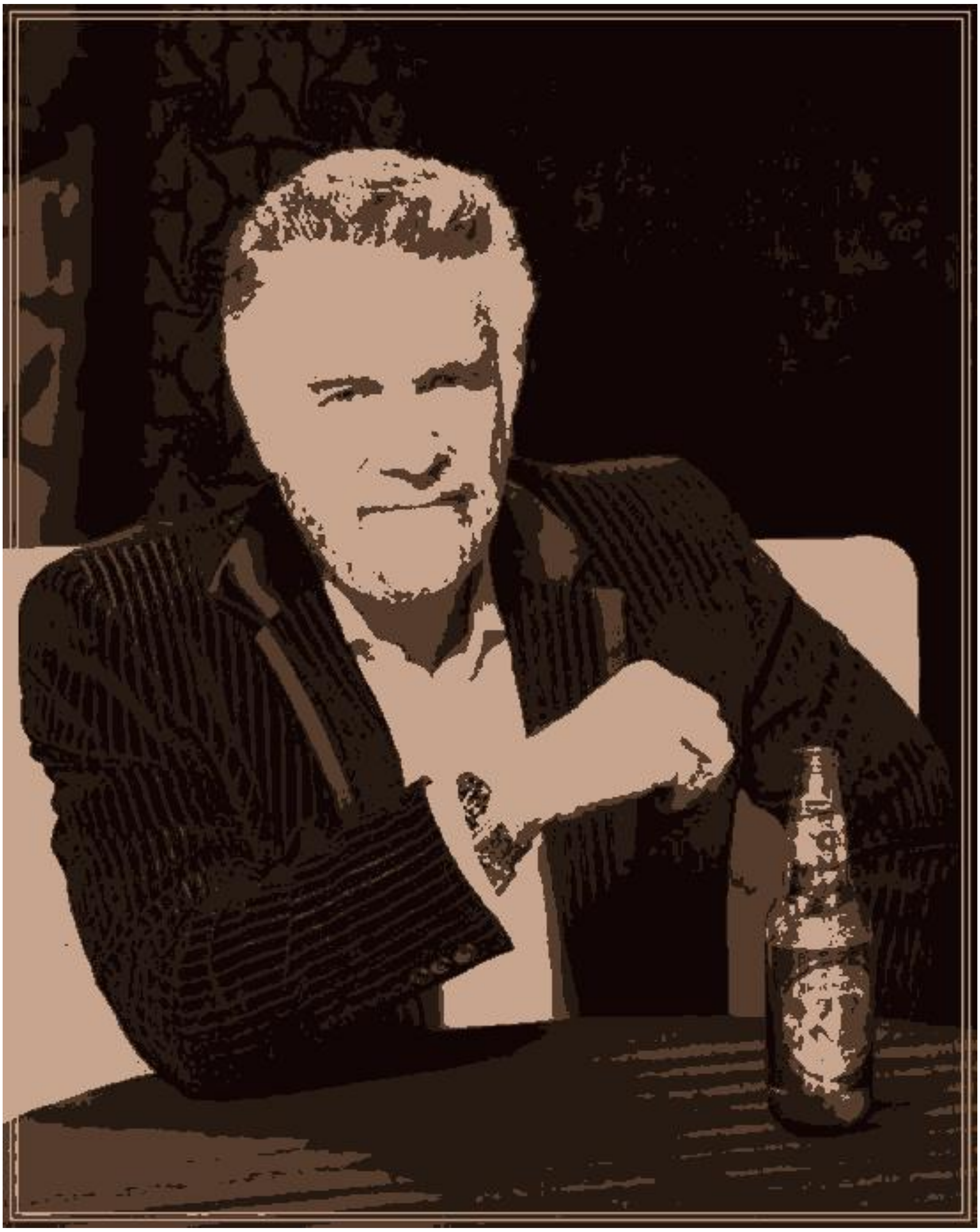


```
--Cluster: 1------
Iteration:  1
Iteration:  2
Convergence criterion is met
Total iterations:  2
Time of execution of EM for cluster(k) = 1  is 0.6706511974334717 seconds
Error of reconstruction: 0.029689080812792795
```

Για αριθμό ομάδων(K) ίσο με 2 έχουμε το ακόλουθο αποτέλεσμα:

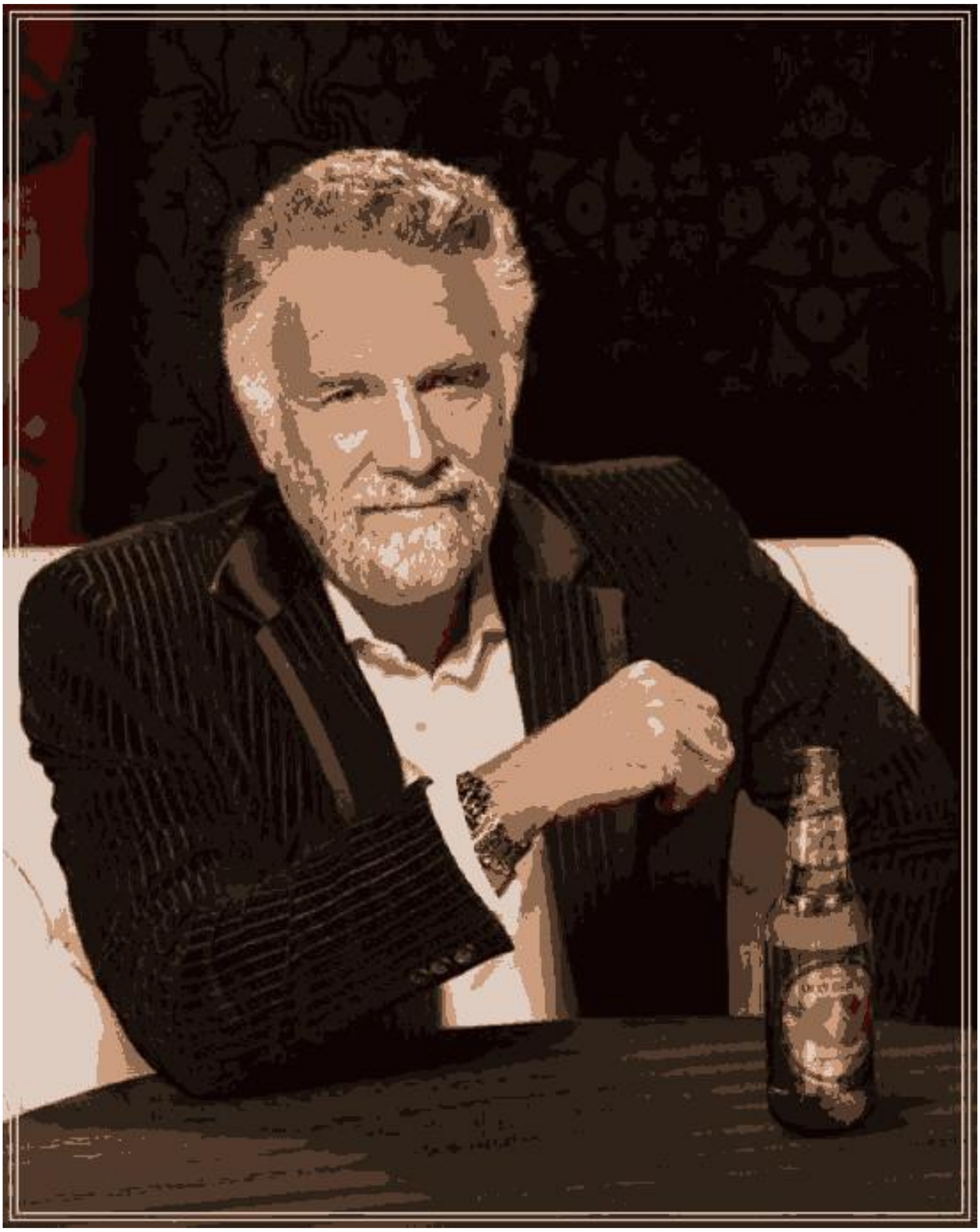
```
Convergence criterion is met
Total iterations:  30
Time of execution of EM for cluster(k) = 2  is 17.1512451171875 seconds
Error of reconstruction: 0.026437034988610133
```

Για αριθμό ομάδων(K) ίσο με 4 έχουμε το ακόλουθο αποτέλεσμα:



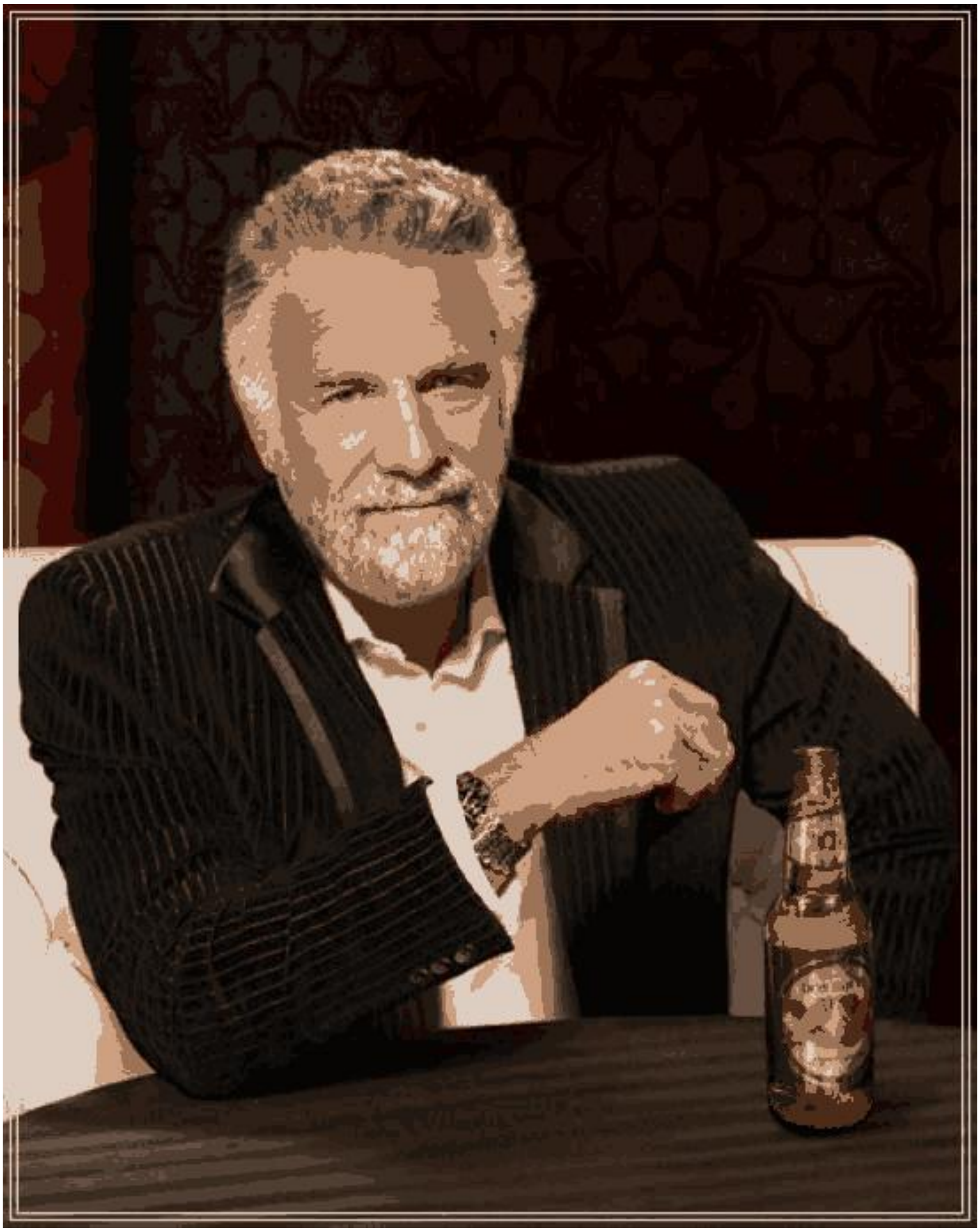
```
Convergence criterion is met
Total iterations:  210
Time of execution of EM for cluster(k) = 4  is 242.62264442443848 seconds
Error of reconstruction: 0.022027688591410143
```

Για αριθμό ομάδων(K) ίσο με 8 έχουμε το ακόλουθο αποτέλεσμα:



```
Convergence criterion is met
Total iterations:  251
Time of execution of EM for cluster(k) = 8  is 549.7296178340912 seconds
Error of reconstruction: 0.019438892701334174
```

Για αριθμό ομάδων(Κ) ίσο με 16 έχουμε το ακόλουθο αποτέλεσμα:



```
Time of execution of EM for cluster(k) = 16  is 2147.7413337230682 seconds
Error of reconstruction: 0.0159829307941458
```

Για αριθμό ομάδων(K) ίσο με 32 έχουμε το ακόλουθο αποτέλεσμα:

```
Time of execution of EM for cluster(k) = 32  is 4517.353576421738 seconds
Error of reconstruction: 0.013109559654380818
```

Για αριθμό ομάδων(Κ) ίσο με 64 έχουμε το ακόλουθο αποτέλεσμα:

```
Time of execution of EM for cluster(k) = 64  is 8567.164550304413 seconds
Error of reconstruction: 0.010639547133201652
```