ΘΕΜΑ 3 - 2019-2020

Δίνονται $D_1 = \{0, 1.8, 2, -2.6, 7, 8, 9, 11\}$, $D_2 = \{2.9, 4.7, 5.5, 6.9\}$

και $p(x|\omega_1) = w_{11} N(x|\mu_{11}, 1) + w_{12} N(x|\mu_{12}, 1)$, $p(x|\omega_2) = N(x|\mu_2, 1)$

(a) Για την ω_2, ο M.L. δίνει:

$$\mathcal{L}(D_2|\theta) = \mathcal{L}(D_2|\mu_2) = \prod_{i=1}^{|D_2|} N(x_i|\mu_2, 1) \Rightarrow \ell = \sum_{i=1}^{|D_2|} \ln N(x_i|\mu_2, 1) \Rightarrow$$

$$\Rightarrow \ell = -\frac{|D_2|}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{|D_2|}(x_i-\mu_2)^2 = -2\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{4}(x_i-\mu_2)^2$$

$$\partial\ell/\partial\mu_2 = 0 \Rightarrow \sum_i (x_i - \mu_2) = 0 \Rightarrow \sum_i x_i - 4\mu_2 = 0 \Rightarrow \mu_2 = \frac{1}{4}\sum_i x_i \Rightarrow$$

$$\Rightarrow \mu_2 = \frac{1}{4}(2.9 + 4.7 + 5.5 + 6.9) = \frac{1}{4}\cdot 20 \Rightarrow \mu_2 = 5$$

Η ω_1 αφορά δείγματα που έχουν ληφθεί από ένα μίγμα 2 κανονικών κατανομών. Η πληροφορία ως προς το από ποια κανονική έχει προκύψει κάθε x του D_1 δεν είναι διαθέσιμη. Δηλαδή, υπάρχει για κάθε $x_i \in D_1$ μια latent variable z_i τέτοια, ώστε

$$z_i = \begin{cases} 0, & \text{εάν το } x_i \text{ προέρχεται από την } N(\mu_{11}, 1) \\ 1, & \text{εάν το } x_i \text{ προέρχεται από την } N(\mu_{12}, 1) \end{cases}$$

Σε ό,τι αφορά την posterior κατανομή των z_i, ισχύει

$$p(z_i = k|x_i) = \frac{p(x_i|z_i = k)\cdot p(z_i = k)}{p(x_i)}, \quad k = 1, 2$$

Όμως $p(z_i = k) = w_{1k}$, $p(x_i|z_i = k) = N(x|\mu_{1k}, 1)$ και
$p(x_i) = p(x_i|\omega_1)$

Έτσι, προκύπτει

$$p(z_i = k \mid x_i) = \frac{w_{1k} N(x_i \mid \mu_{1k}, 1)}{\sum_k w_{1k} N(x_i \mid \mu_{1k}, 1)} =: \gamma_{z_i}(k)$$

Η συνάρτηση πιθανοφάνειας για την $p(x \mid \omega_1)$ είναι:

$$\mathcal{L} = \prod_{i=1}^{|D_1|} p(x_i \mid \omega_1) \Rightarrow \ell = \sum_{i=1}^{|D_1|} \ln p(x_i \mid \omega_1) \Rightarrow$$

$$\Rightarrow \ell = \sum_i \ln \sum_k w_{1k} N(x_i \mid \mu_{1k}, 1)$$, με παραμέτρους $\vec{\theta} = (\mu_{11}, \mu_{12}, w_{11}, w_{12})$

$$\partial \ell / \partial \mu_{1q} = 0 \Rightarrow \sum_i \frac{1}{\sum_k w_{1k} N(x_i \mid \mu_{1k}, 1)} \cdot w_{1q} \cdot N(x_i \mid \mu_{1q}, 1) \cdot (-x_i + \mu_{1q}) = 0$$

$$\Rightarrow \sum_i \frac{w_{1q} N(x_i \mid \mu_{1q}, 1)}{\sum_k w_{1k} N(x_i \mid \mu_{1k}, 1)} \cdot (x_i - \mu_{1q}) = 0 \Rightarrow \sum_i \gamma_{z_i}(q)(x_i - \mu_{1q}) = 0 \quad (I)$$

Σε ό,τι αφορά τις w_{11}, w_{12}, δεν αρκεί να θέσει κανείς $\partial \ell / \partial w_{1q} = 0$, καθώς αυτές υπόκεινται στον επιπλέον περιορισμό $\sum_k w_{1k} = 1$. Ως εκ τούτου, θα πρέπει $\partial \ell' / \partial w_{1q} = 0$, όπου

$$\ell' = \ell - \lambda \left(\sum_k w_{1k} - 1 \right)$$, με $\partial \ell' / \partial \lambda = 0$. Προκύπτει:

$$\frac{\partial \ell'}{\partial w_{1q}} = 0 \Rightarrow \sum_i \frac{N(x_i \mid \mu_{1q}, 1)}{\sum_k w_{1k} N(x_i \mid \mu_{1k}, 1)} - \lambda = 0 \Rightarrow$$

$$\frac{1}{w_{1q}} \sum_i \frac{w_{1q} N(x_i \mid \mu_{1q}, 1)}{\sum_k w_{1k} N(x_i \mid \mu_{1k}, 1)} = \lambda \Rightarrow w_{1q} = \frac{1}{\lambda} \sum_i \gamma_{z_i}(q)$$

Όμως, $\sum_k w_{1k} = 1 \Rightarrow \frac{1}{\lambda}\sum_k \sum_i \gamma_{z_i}(k) = 1 \Rightarrow$

$$\sum_i \sum_k p(z_i = k | x_i) = \lambda \Rightarrow \sum_i 1 = \lambda \Rightarrow |D_1| = \lambda$$, άρα τελικά

$$w_{1q} = \frac{1}{|D_1|}\sum_i \gamma_{z_i}(q) \quad (II)$$

Οι λύσεις των (I) και (II) αποτελούν τα M.L. estimates. Το πρόβλημα που δημιουργείται είναι πως οι εξ. (I) και (II) δεν είναι σε κλειστή μορφή, αφού οι παράμετροι γ εξαρτώνται από τις $\vec{\theta}$. Αυτό σημαίνει πως ο μόνος τρόπος να λυθεί το πρόβλημα είναι μέσω E.M. Παρ' όλα αυτά, μπορούμε να κάνουμε ένα iteration και να το βαφτίσουμε «απάντηση» στο ερώτημα (a). Παρατηρώντας ότι τα στοιχεία του D_1 μπορούν να χωριστούν σε

$$\{0, 1.8, 2, -2.6\} \rightsquigarrow k=1 \qquad \{7, 8, 9, 11\} \rightarrow k=2,$$

στην ουσία εκτιμούμε τις τιμές των z_i ως

$$z_i = \begin{cases} 0, & i=1,\dots,4 \\ 1, & i=5,\dots,8 \end{cases}$$, οπότε

$$\gamma_{z_i}(1) = \begin{cases} 1, & i=1,\dots,4 \\ 0, & \text{αλλιώς} \end{cases} \quad \text{και} \quad \gamma_{z_i}(2) = \begin{cases} 0, & i=1,\dots,4 \\ 1, & \text{αλλιώς} \end{cases}$$

Βάσει αυτών των εκτιμήσεων, η (II) δίνει:

$w_{11} = \frac{1}{8}\cdot[4\cdot 0 + 4\cdot 1] = 1/2$, $w_{12} = \frac{1}{8}\cdot[4\cdot 0 + 4\cdot 1] = 1/2$, ενώ η (I) δίνει:

$$\sum_i \gamma_{z_i}(q)\, x_i = \sum_i \gamma_{z_i}(q)\, \mu_{1q} \Rightarrow \mu_{1q} = \frac{\sum_i \gamma_{z_i}(q)\, x_i}{\sum_i \gamma_{z_i}(q)},$$

οπότε $\mu_{11} = \frac{x_1 + x_2 + x_3 + x_4}{4} \Rightarrow \mu_{11} = 0.3$ και

$\mu_{12} = \frac{x_5 + x_6 + x_7 + x_8}{4} \Rightarrow \mu_{12} = 8.75$

Τελικά, οι εκτιμήσεις είναι

$p(x|\omega_2) = N(x|5,1)$ και

$p(x|\omega_1) = \frac{1}{2}\left[N(x|0.3,1) + N(x|8.75,1)\right]$.

Αυτά τα οδηγα και είδαμε μόλις στο ερώτημα (α).

(β) Ισχύει $p(\omega_1) = \frac{|D_1|}{|D_1|+|D_2|} = \frac{8}{12} \Rightarrow p(\omega_1) = 2/3$

και $p(\omega_2) = \frac{|D_2|}{|D_1|+|D_2|} = 4/12 \Rightarrow p(\omega_2) = 1/3$

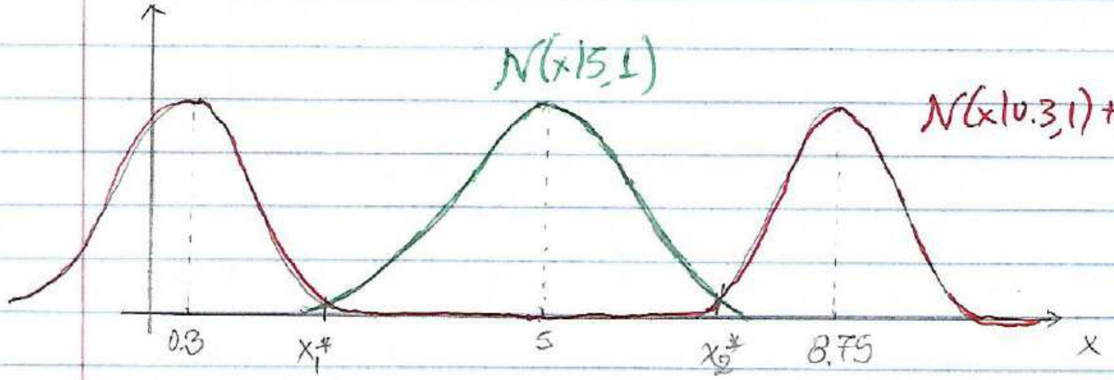
Ο κανόνας απόφασης Bayes προκύπτει ως λύση της

$p(\omega_1)\, p(x|\omega_1) = p(\omega_2)\, p(x|\omega_2) \Rightarrow$

$\frac{2}{3} \cdot \frac{1}{2}\left[N(x|0.3,1) + N(x|8.75,1)\right] = \frac{1}{3} \cdot N(x|5,1) \Rightarrow$

$$N(x|0.3,1) + N(x|8.75,1) = N(x|5,1) \quad (1)$$

Η εξίσωση (1) οδηγεί στον εντοπισμό των θέσεων x_1^* και x_2^* που φαίνονται στο ακόλουθο σχήμα:



Οι Γκαουσσιανές που συνιστούν το μίγμα είναι πολύ καλά διαχωρισμένες μεταξύ τους. Για παράδειγμα,

Για $x = 0.3 + \kappa \cdot \sigma = 0.3 + \kappa$, ισχύει

$$N(x|0.3,1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\kappa^2\right] \quad \text{και}$$

$$N(x|8.75,1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(\kappa - 8.45)^2\right] \quad \text{, οπότε}$$

$$\frac{N(x|0.3,1)}{N(x|8.75,1)} = \exp\left[-\frac{1}{2}\kappa^2 + \frac{1}{2}\kappa^2 + \frac{8.45^2}{2} - 8.45\kappa\right] =$$

$$= \exp\left[8.45(4.225 - \kappa)\right]$$

Το γεγονός αυτό υποδεικνύει πως ακόμα και για 4σ μακριά από την μ_1, η συνεισφορά της $N(x|0.3,1)$ στο άθροισμα είναι $\exp(8.45 \cdot 0.225) \simeq 6.7$ φορές υψηλότερη από την αντίστοιχη της $N(x|0.3,1)$. Αντίστοιχα αποτελέσματα προκύπτουν και για κ σ μακριά από την μ_2. Ως εκ τούτου, θα θεωρήσουμε κατά προσέγγιση πως

$$N(x_1^*|0.3,1) + N(x_1^*|8.75,1) = N(x_1^*|5,1) \Leftrightarrow N(x_1^*|0.3,1) = N(x_1^*|5,1)$$

και

$$N(x_2^*|0.3,1) + N(x_2^*|8.75,1) = N(x_2^*|5,1) \Leftrightarrow N(x_2^*|8.75,1) = N(x_2^*|5,1)$$

Εάν η x_1^* προκύψει πιο μακριά από την μ_{11} κατά $>4\sigma=4$, ή η x_2^* προκύψει πιο μακριά από την μ_{12} κατά $>4\sigma=4$, τότε η προσέγγισή μας θα είναι προβληματική.

• $\mathcal{N}(x_1^*|0.3,1)=\mathcal{N}(x_1^*|5,1) \Leftrightarrow (x_1^*-0.3)^2=(x_1^*-5)^2 \Leftrightarrow$

$|x_1^*-0.3|=|x_1^*-5| \Leftrightarrow x_1^*=2.65$

• $\mathcal{N}(x_2^*|8.75,1)=\mathcal{N}(x_2^*|5,1) \Leftrightarrow \ldots \Leftrightarrow |x_2^*-8.75|=|x_2^*-5| \Leftrightarrow$

$\Leftrightarrow x_2^*=6.875$

Το x_1^* απέχει μόλις 2σ + κάτι από τη μ_{11} και το ίδιο ισχύει και για την απόσταση του x_2^* από τη μ_{12}. Το γεγονός αυτό σημαίνει πως η προσέγγισή μας είναι παραπάνω από ικανοποιητική, αφού η επίλυση της ακριβούς εξίσωσης θα διόρθωνε τα παραπάνω αποτελέσματα στο 7ο δεκαδικό ($e^{8.45\cdot 2}\simeq 9\cdot 10^7$).

Τελικά, $x_1^*\simeq 2.65$ και $x_2^*\simeq 6.875$

(γ) $\omega_1 \to$ ○ $\quad \omega_2 \to$ ●

Οι επιλογές για ερωτήσεις είναι άπειρες. Όμως η ερώτηση $x>2.9$ είναι ισοδύναμη με όλες τις ερωτήσεις $x>\alpha$ για $\alpha\in(2.9, 4.7)$, οπότε στην πράξη μένουν 13 «ουσιώδεις» ερωτήσεις. Πρόσθετα, λόγω της συμμετρίας στην κατανομή των σημείων γύρω από την τιμή 5.1 (2 της ω_2 και 4 της ω_1 και από αριστερά και από δεξιά) έχει νόημα να διερευνηθούν μόνο 7 ερωτήσεις:

i) $x > -2.6-\varepsilon$; ii) $x > -2.6$; iii) $x > 0$;

iv) $x > 1.8$; v) $x > 2$; vi) $x > 2.9$; vii) $x > 4.7$;

Προκειμένου να φτιαχτεί ένας ταξινομητής που να γενικεύει καλύτερα, μπορούν οι αντίστοιχες ερωτήσεις να αφορούν τα μέσα των υπο-διαστημάτων που οριοθετούν τα σημεία και όχι τα άκρα τους.

Προφανώς, οι ερωτήσεις είναι: 1) $x > 6.95$ και
2) $x > 2.45$

αλλά οι πράξεις είναι πολλές για να τις κάνουμε αναλυτικά

(δ) Δεδομένου ενός NNR-3 με απλή πλειοψηφία (και όχι ομοφωνία) θα ισχύει: