4. **(Algoritmul EM pentru GMM, cazul uni-variat: demonstrația formulelor de „actualizare“ de la pasul M)**

Se consideră algoritmul EM pentru estimarea parametrilor unui model de mixtură de k distribuții gaussiene, despre care se presupune că au [toate] aceeași varianță σ^2, iar probabilitățile a priori de selecție sunt egale $(1/k)$. Prin urmare, se vor estima doar mediile celor două distribuții gaussiene, aplicându-se formulele următoare:

Pasul E:

$$E[z_{ij}] \stackrel{not.}{=} E[z_{ij}|X = x_i; \mu, \sigma^2] = P(z_{ij} = 1|X = x_i; \mu, \sigma^2) \stackrel{F.Bayes}{=} \ldots$$

$$= \frac{\exp\left(-\frac{1}{2\sigma^2}(x_i - \mu_j)^2\right)}{\sum_{p=1}^{k} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu_p)^2\right)} \quad \text{pentru } i = 1, \ldots, n \text{ și } j = 1, \ldots, k$$

Pasul M:

$$\mu_j = \frac{\sum_{i=1}^{n} E[z_{ij}]x_i}{\sum_{i=1}^{n} E[z_{ij}]}, \quad \text{pentru } j = 1, \ldots, k$$

unde

- **simbolul** exp **desemnează funcția exponențială (e^x),**
- n **este numărul de puncte/instanțe,**
- k **este numărul de clustere,**
- $\mu \stackrel{not.}{=} (\mu_1, \ldots, \mu_k)$,
- **iar z_{ij}, cu $i \in \{1, \ldots, n\}$ și $j \in \{1, \ldots, k\}$, sunt variabilele-indicator neobservabile (sau „ascunse“ / „latente“), luând valoarea $z_{ij} = 1$ dacă x_i a fost generat de gaussiana j și $z_{ij} = 0$ în caz contrar.**

a. Demonstrați formula de la pasul M.

Sugestie:
Veți scrie mai întâi (i) log-verosimilitatea unei instanțe „complete“ $y_i = (x_i, z_{i1}, \ldots, z_{ik})$, apoi (ii) funcția de log-verosimilitate a datelor complete $Y = \{y_1, \ldots, y_n\}$, după care veți calcula (iii) media acestei funcții și, în final, (iv) valorile μ_j pentru care se atinge valoarea optimă a mediei funcției de log-verosimilitate.

b. Demonstrați că algoritmul acesta buclează dacă se inițializează toate mediile μ_j cu o aceeași valoare.

a) ii) fct. de verosimilitate a datelor complete:

$P(Y|\mu) = P(y_1, \ldots, y_n|\mu) \stackrel{i.i.d.}{=} \prod_{i=1}^{n} P(y_i|\mu)$

$\Rightarrow \log P(Y|\mu) = \log \prod_{i=1}^{n} P(y_i|\mu) = \sum_{i=1}^{n} \ln P(y_i|\mu)$

i) $P(y_i) = P(x_i, z_{i1}, \ldots, z_{ik}) = P(x_i | z_{i1} \ldots z_{ik}) \cdot P(z_{i1}, \ldots, z_{ik}) =$ $\frac{1}{k}$

$= \frac{1}{k} \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$ (μ = media gaussienei care a generat pe instanța y_i)

$\Rightarrow \log P(y_i) = \log \frac{1}{k} + \log \frac{1}{\sqrt{2\pi}\sigma} - \frac{(x_i - \mu)^2}{2\sigma^2} =$

$= -\log k - \frac{1}{2} \log(2\pi\sigma) - \frac{(x_i-\mu)^2}{2\sigma^2}$ (*)

Obs. $\ln(P(y_i))$ e, în acest context, totuna cu $P(y_i|\mu)$. → notație simplificată.

11.

$\Rightarrow$ cont. înlocuim în ii)

$\ln P(Y|\mu) = \sum_{i=1}^{n} \ln P(y_i|\mu) \stackrel{(*)}{=} \sum_{i=1}^{n} \left(-\ln k - \frac{1}{2}\ln(2\pi\sigma) - \frac{(x_i-\mu)^2}{2\sigma^2}\right) =$

$= -\sum_{i=1}^{n} \ln k - \frac{1}{2}\sum_{i=1}^{n} \ln(2\pi\sigma) - \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2} =$

$= \text{——————} - \sum_{i=1}^{n}\sum_{j=1}^{k} z_{ij} \cdot \frac{(x_i-\mu_j)^2}{2\sigma^2}$

iii) $E[\ln P(Y|\mu)] = E\Big[-\sum_{i=1}^{n} \ln k - \frac{1}{2}\sum_{i=1}^{n} \ln(2\pi\sigma) - \sum_{i=1}^{n}\sum_{j=1}^{k} z_{ij} \cdot \frac{(x_i-\mu_j)^2}{2\sigma^2}\Big] =$

prop. de liniarit. a mediilor

$= \underbrace{-\sum_{i=1}^{n} \ln k - \frac{1}{2}\sum_{i=1}^{n} \ln(2\pi\sigma)}_{ct.} - \sum_{i=1}^{n}\sum_{j=1}^{k} E[z_{ij}] \cdot \frac{(x_i-\mu_j)^2}{2\sigma^2}$

iv) $\hat{\mu} = \arg\max_{\mu} E[\ln P(Y|\mu)] =$

$= \arg\min_{\mu} \sum_{i=1}^{n}\sum_{j=1}^{k} E[z_{ij}] \cdot \frac{(x_i-\mu_j)^2}{2\sigma^2}$; notez cu Expr.

Aflu minimul expresiei folosind derivarea:

$(Expr)' = \sum_{i=1}^{n}\sum_{j=1}^{k} E[z_{ij}] \cdot \frac{2(x_i-\mu_j) \cdot (-1)}{2\sigma^2}$

$= \arg\min_{\mu} \frac{1}{2\sigma^2}\sum_{i=1}^{n}\Big(\sum_{j=1}^{k} E[z_{ij}] \cdot (x_i^2 + \mu_j^2 - 2x_i\mu_j)\Big)$

funcția de mai sus are un minim în punctul $-\frac{b}{2a}$, adică

$$\frac{\sum_{i=1}^{n} 2x_i \cdot E[z_{ij}]}{2 \cdot \sum_{i=1}^{n} E[z_{ij}]} = \frac{\sum_{i=1}^{n} E[z_{ij}] \cdot x_i}{\sum_{i=1}^{n} E[z_{ij}]}.$$

12.

F.B.!

continuare 4:

e) Conform enunțului, avem: $\mu_1 = \mu_2 = \dots = \mu_k = \mu$ (**)

pasul E: $E[z_{ij}] \stackrel{def}{=} \dfrac{\exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_j)^2\right)}{\sum_{p=1}^{k} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu_p)^2\right)} \stackrel{(**)}{=} \dfrac{e^{\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right)}}{\sum_{p=1}^{k} e^{\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right)}}$ → Obs. că numitorul va fi = pt toate instanțele

dar și numitorul ! deci $E[z_{ij}] = \frac{1}{k}$, $\forall i, \forall j$

pasul M: Fie $\mu_j^{(j)}$ media unei gausiene oarecare din cele k.

$\mu_j \stackrel{def}{=} \dfrac{\sum_{i=1}^{m} E[z_{ij}] x_i}{\sum_{i=1}^{m} E[z_{ij}]} = \cancel{\dfrac{\sum_{i=1}^{m} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} x_i}{\sum_{i=1}^{m} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}}}$

$\Rightarrow \dfrac{\sum_{i=1}^{m} \dfrac{e^{\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right)}}{\sum_{p=1}^{k} e^{\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right)}} \cdot x_i}{\sum_{i=1}^{m} \dfrac{e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}}{\sum_{p=1}^{k} e^{\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right)}}} = \dfrac{\sum_{i=1}^{m} \dfrac{1}{\sum_{p=1}^{k} e^{\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right)}} \left(x_i \cdot e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}\right)}{\sum_{i=1}^{m} \dfrac{1}{\cdots} \left(e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}\right)} \neq 0$

Se obs. că mediile celor k gausiene au rămas egale între ele. ⇒ la următoarea iterație voi obține aceleași ~~valori~~ medii ⇒ algoritmul EM buclează.

continuare problema:

$\theta_{MLE} = \dfrac{n}{\sum_{i=1}^{n} x_i + n}$

⇒ pt. primele 5:

$\theta_{MLE} = \dfrac{5}{5+(0+21+23+8+9)} = \dfrac{5}{5+61} = \dfrac{5}{66} \approx 0{,}075$

→ pt primele 10:

$\theta_{MLE} = \dfrac{10}{10+0+21+23+8+9+2+9+0+7+8} = \dfrac{10}{10+87} = \dfrac{10}{97} \approx 0{,}103$

→ pt. toate:

$\theta_{MLE} = \dfrac{15}{15+0+21+23+8+9+2+9+0+7+8+20+9+7+4+17} = \dfrac{15}{159} \approx 0{,}094$