Zadatak 1: Algoritam maksimizacije očekivanja

(a) Napišite funkciju (nepotpune) log-izglednosti za općenit model miješane gustoće. Što je problem s tom funkcijom?

Nepotpuna log-izglednost:

$$\ln \mathcal{L}(\boldsymbol{\theta}|\mathbf{X}) = \ln p(\mathbf{X}|\boldsymbol{\theta}) = \ln \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$$

Problem:

imamo logaritam zbroja, pa ne možemo naći rješenje u zatvorenoj formi nego moramo koristiti iterativne metode.

Naravno pod rješenje mislim na rješenje problema maksimizacije log-izglednosti.

(b) Definirajte E-korak i M-korak općenitog EM-algoritma. Zašto algoritam provodimo iterativno?

E-korak:

Računamo očekivanje potpune log-izglednosti uz fiksirane trenutne vrijednosti parametra $\theta^{(t)}$. To očekivanje označavamo sa $\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ i računamo kao:

$$\begin{aligned}\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &\equiv \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(t)}}[\ln \mathcal{L}(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Z})] \\ &= \mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{(t)}}[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})] \\ &= \sum_{\mathbf{Z}} P(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{(t)}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})\end{aligned}$$

M-korak:

Odabiremo nove parametre $\theta^{(t+1)}$ koji maksimiziraju gornje očekivanje:

$$\boldsymbol{\theta}^{(t+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, \mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$$

Zašto iterativno:

Zato što ne znamo vrijednosti latentnih varijabli Z pa ne možemo raditi sa potpunom log-izglednošću, pa umjesto toga radimo sa očekivanjem potpune log-izglednosti. Osnovna ideja EM-algoritma jest iterativno ugađati parametre θ kako bi se maksimiziralo to očekivanje. [Maksimizacija očekivanja potpune log-izglednosti ujedno dovodi do povećanja nepotpune log-izglednosti, što je zapravo naš cilj.]

(c) Krenuvši od potpune log-izglednosti $\ln \mathcal{L}(\boldsymbol{\theta}|\mathcal{D},\mathcal{Z})$ za općenit model miješane gustoće, izvedite izraze za procjenu parametara π_k, $\boldsymbol{\mu}_k$ i $\boldsymbol{\Sigma}_k$ modela Gaussove mješavine.

Potpuna log-izglednost, općeniti model miješane gustoće:

$$\ln \mathcal{L}(\boldsymbol{\theta}|\mathcal{D},\mathcal{Z}) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_k^{(i)}\Big(\ln \pi_k + \ln p(\mathbf{x}^{(i)}|\boldsymbol{\theta}_k)\Big)$$

Očekivanje [ono se koristi u EM-algoritmu]: (jedino $z_k^{(i)}$ je slučajna varijabla)

$$\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathbb{E}_{\mathcal{Z}|\mathcal{D},\boldsymbol{\theta}^{(t)}}[\ln \mathcal{L}(\boldsymbol{\theta}|\mathcal{D},\mathcal{Z})] = \sum_{i=1}^{N}\sum_{k=1}^{K}\mathbb{E}[z_k^{(i)}|\mathcal{D},\boldsymbol{\theta}^{(t)}]\big(\ln \pi_k + \ln p(\mathbf{x}^{(i)}|\boldsymbol{\theta}_k)\big)$$

Uvodimo novu veličinu, odgovornost $h_k^{(i)}$, koja odgovara vjerojatnosti da je primjer $x^{(i)}$ generirala komponenta k:

$$\mathbb{E}[z_k^{(i)}|\mathcal{D},\boldsymbol{\theta}^{(t)}] = \mathbb{E}[z_k^{(i)}|\mathbf{x}^{(i)},\boldsymbol{\theta}^{(t)}] = P(z_k^{(i)}=1|\mathbf{x}^{(i)},\boldsymbol{\theta}^{(t)})$$

$$P(z_k^{(i)}|\mathbf{x}^{(i)},\boldsymbol{\theta}^{(t)}) = \frac{p(\mathbf{x}^{(i)}|z_k^{(i)},\boldsymbol{\theta}^{(t)})\pi_k^t}{\sum_{j=1}^{N} p(\mathbf{x}^{(i)}|z_j^{(i)},\boldsymbol{\theta}^{(t)})\pi_j^t} \equiv h_k^{(i)}$$

Uvrstimo odgovornost u očekivanje i razdvojimo pribrojnike:

$$\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \sum_{i=1}^{N}\sum_{k=1}^{K} h_k^{(i)}\ln \pi_k + \sum_{i=1}^{N}\sum_{k=1}^{K} h_k^{(i)}\ln p(\mathbf{x}^{(i)}|\boldsymbol{\theta}_k)$$

Sad idemo derivirati očekivanje po komponentama i izjednačavati sa nulom:

Iz $\nabla_{\pi_k}\mathcal{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = 0$ i uvjeta $\sum_k \pi_k = 1$ primjenom Lagrangeovih multiplikatora dobivamo:

$$\nabla_{\pi_k}\Big(\sum_{i=1}^{N}\sum_{k=1}^{K} h_k^{(i)}\ln \pi_k + \lambda\big(\sum_k \pi_k - 1\big)\Big) = 0$$

$\boxed{\pi_k}$

$$\frac{1}{\pi_k}\sum_{i=1}^{N} h_k^{(i)} + \lambda = 0$$

Trikovima i crnom magijom dobivamo da je Lagrangeov multiplikator λ = - N :

Ako obje strane jednadžbe pomnožimo s π_k i zatim lijevu stranu sumiramo po svim k i izjednačimo s nulom (što vrijedi budući da, ako su svi pojedinačni pribrojnici jednaki nuli, onda je i njihov zbroj jednak nuli), dobivamo

$$\sum_{k=1}^{K}\Big(\pi_k\frac{1}{\pi_k}\sum_{n=1}^{N} h_k^{(i)} + \pi_k\lambda\Big) = \sum_{i=1}^{N}\sum_{k=1}^{K} h_k^{(i)} + \sum_{k=1}^{N}\pi_k\lambda = N + \lambda = 0$$

pri čemu smo iskoristili $\sum_k h_k^{(i)} = \sum_k P(z_k^{(i)}|\mathbf{x}^{(i)},\boldsymbol{\theta}) = 1$ (marginalizacija) i $\sum_k \pi_k = 1$.

Konačno, uvršavanjem λ u prethodnu formulu, dobijemo:

$$\boxed{\pi_k^{(t+1)} = \frac{1}{N}\sum_{i=1}^{N} h_k^{(i)}}$$

Za preostale komponente gledamo drugi pribrojnik u onom zbroju za očekivanje:

$$\nabla_{\boldsymbol{\theta}_k} \sum_{i=1}^{N} \sum_{k=1}^{K} h_k^{(i)} \ln p(\mathbf{x}^{(i)}|\boldsymbol{\theta}_k) = \nabla_{\boldsymbol{\theta}_k} \sum_{i=1}^{N} h_k^{(i)} \ln p(\mathbf{x}^{(i)}|\boldsymbol{\theta}_k) = 0$$

Multivarijatna Gaussova razdioba:

$$p(\mathbf{X} = \mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

Dakle....

$$\nabla_{\boldsymbol{\theta}_k} \sum_{i=1}^{N} h_k^{(i)} \left(-\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{x}^{(i)}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}^{(i)}-\boldsymbol{\mu})\right) = 0$$

gdje $\boldsymbol{\theta}_k = (\mu_k, \sigma_k^2)$. Deriviranjem i rješavanjem po $\boldsymbol{\mu}_k$ odnosno $\boldsymbol{\Sigma}_k$ dobivamo:

$\boxed{\boldsymbol{\mu}_k}$

$$\boldsymbol{\mu}_k^{(t+1)} = \frac{\sum_i h_k^{(i)}\mathbf{x}^{(i)}}{\sum_i h_k^{(i)}}$$

$\boxed{\boldsymbol{\Sigma}_k}$

$$\boldsymbol{\Sigma}_k^{(t+1)} = \frac{\sum_i h_k^{(i)}(\mathbf{x}^{(i)}-\boldsymbol{\mu}_k^{(t+1)})(\mathbf{x}^{(i)}-\boldsymbol{\mu}_k^{(t+1)})^{\mathrm{T}}}{\sum_i h_k^{(i)}}$$

(d) Na predavanjima smo spomenuli da je algoritam k-srednjih vrijednosti poseban slučaj EM-algoritma uz pretpostavku dijeljene izotropne kovarijacijske matrice. Uz tu pretpostavku, napišite izraz za potpunu log-izglednost $\ln \mathcal{L}(\theta|\mathcal{D}, \mathcal{Z})$. Usporedite dobiveni izraz s funkcijom pogreške J algoritma k-srednjih vrijednosti.

Potpuna log-izglednost:

Uz pretpostavku dijeljene izotropne kovarijacijske matrice i zanemarenih istih članova zbroja

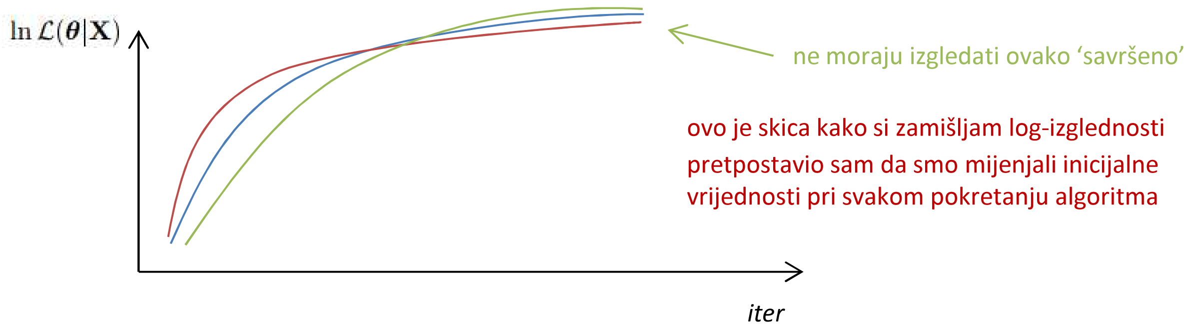
$$\ln \mathcal{L}(\boldsymbol{\theta}|\mathcal{D}, \mathcal{Z}) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_k^{(i)} \left(\ln \pi_k - \frac{1}{2\sigma^2}\|\mathbf{x}^{(i)} - \boldsymbol{\mu}_k\|^2\right)$$

Funkcija pogreške algoritma k-srednjih vrijednosti:

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} b_k^{(i)} \|\mathbf{x}^{(i)} - \boldsymbol{\mu}_k\|^2$$

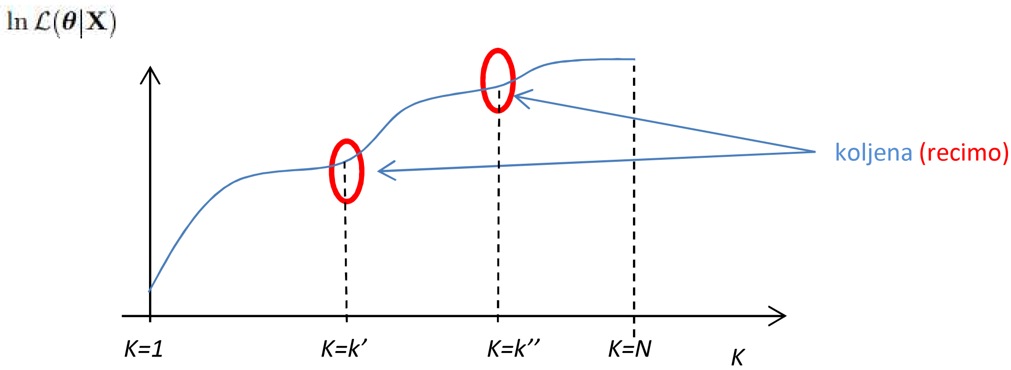
Što bih trebao usporediti, da je drugi pribrojnik sume u potpunoj log-izglednosti jednak izrazu u funkciji pogreške algoritma k-means? Ako bismo još vrijednosti $h_k^{(i)}$ zaokruživali na 0 ili 1, onda smo EM algoritam sveli na k-means algoritam.

(e) Skicirajte krivulju log-izglednosti kao funkciju od broja iteracija EM-algoritma. Obrazložite izgled krivulje. Pretpostavite da smo algoritam pokretali još tri puta i na istome grafu ucrtajte odgovarajuće krivulje. Jesu li krivulje identične? Obrazložite odgovor.



U čemu je razlika?
Razlika je u inicijalizaciji početnih parametara. Ovisno kako odaberemo početne parametre, završit ćemo u nekom lokalnom optimumu za log-izglednost, koji ne mora biti i globalni optimum. Znamo da će algoritam završiti u nekom optimumu jer uvijek konvergira. Konvergencija je zajamčena budući da algoritam u svakoj iteraciji povećava očekivanje izglednosti.

(f) Skicirajte krivulju log-izglednosti kao funkciju broja grupa K. Obrazložite izgled krivulje. Možete li temeljem ove krivulje odrediti optimalan broj grupa? Kako?



Obrazloženje skice:
Za neke vrijednosti (k', k'') primijetit ćemo 'skok' u vrijednosti od log-izglednosti (koljena na grafu). Možemo pretpostaviti da su te vrijednosti brojevi prirodnih grupa u našim uzorcima. Također valja primijetiti da će log-izglednost biti maksimalna ako za svaki primjer kažemo da pripada svojoj vlastitoj grupi, ali onda naše grupiranje gubi smisao [grupiranje u N grupa od po jednog elementa nema smisla]. Za optimalan K trebali bi koristiti neku od metoda provjere (validacije) grupa.
Ovo gore navedeno bi trebao biti primjer grafičke validacije, ali pošto osobno nisam iskusan sa tom metodom, radije bih se uzdao u nešto poput:

Minimiziramo kriterij koji kombinira kriterijsku funkciju i složenost modela te na neki način kažnjavamo modele s prevelikim brojem grupa. Općenit oblik tog kriterija je

$$K^* = \operatorname*{argmin}_K \left(J(K) + \lambda K\right)$$

gdje je $J(K)$ vrijednost kriterijske funkcije za model s K grupa, a λ težinski faktor. Veće vrijednosti faktora λ favoriziraju rješenja s manjim brojem grupa. Za $\lambda = 0$ povećanje broja grupa se ne kažnjava i optimalan broj grupa tada je $K^* = N$.