

## Tema 11 - laborator 13

### Modele mixte

Ex. 1.

1. Generați 500 de date dintr-o mixtură de trei distribuții Gaussiene. În fișierul alăturat aveți un astfel de exemplu.
2. Calibrați pe acest set de date un model de mixtură de distribuții Gaussiene cu 2, 3, respectiv 4 componente.
3. Comparați cele 3 modele folosind metodele WAIC și LOO. Care este concluzia?

Ex. 2. Un institut de cercetare în domeniul sănătății dorește să analizeze relația dintre numărul de ore de exerciții fizice pe săptămână și nivelul de colesterol la indivizi din mai multe grupuri demografice distincte. Acești indivizi provin din 3 până la 5 subpopulații diferite (ex: vârstă, stil de viață, genetică), iar proporțiile fiecărei subpopulații în eșantion nu sunt cunoscute.

Modelul presupune că fiecare individ aparține uneia dintre subpopulații, fiecare având propriul său model de regresie (polinomială). Alegerea subgrupului este determinată de următoarea mixtură

$$\text{Colesterol}_i = \sum_{k=1}^K w_k \cdot \mathcal{N}(\mu_{k,i}, \sigma_k^2),$$

unde  $K$  este numărul de subpopulații,  $w_k$  sunt ponderile subpopulațiilor,  $\mu_{k,i} = \alpha_k + \beta_k t_i + \gamma_k t_i^2$  (model polinomial), iar  $\sigma_k$  este deviația standard a fiecărui model. Observațiile privind numărul de ore de exerciții fizice pe săptămână și nivelul de colesterol sunt strânse în fișierul `date_colesterol.csv`.

1. Estimați ponderile și coeficienții de regresie pentru fiecare subpopulație, pentru fiecare  $K \in \{3, 4, 5\}$ .
2. Câte subpopulații reprezintă cel mai bine datele observate? Justificați acest lucru folosind criterii bayesiene, cum ar fi WAIC și/sau LOO.