Øvingsoppgaver
TEK5020/9020 – Mønstergjenkjenning
Del 5 – Ikke-ledet læring

Høsten 2023

Oppgave 1

La $p(\boldsymbol{x}|\boldsymbol{\theta})$ være en blandingstetthet med c multivariat normalfordelte komponenter

$$p(\boldsymbol{x}|\omega_i, \boldsymbol{\theta}_i) \sim N(\boldsymbol{\mu}_i, \sigma_i^2 I),$$

der $\boldsymbol{\mu}_i$, σ_i og $P(\omega_i)$ er ukjente for $i = 1, \ldots, c$. Vis at maksimum likelihood estimatet for σ_i^2 må tilfredsstille

$$\hat{\sigma}_i^2 = \frac{\frac{1}{d}\sum_{k=1}^{n} \hat{P}(\omega_i|\boldsymbol{x}_k, \hat{\boldsymbol{\theta}})\|\boldsymbol{x}_k - \hat{\boldsymbol{\mu}}_i\|^2}{\sum_{k=1}^{n} \hat{P}(\omega_i|\boldsymbol{x}_k, \hat{\boldsymbol{\theta}})},$$

der $\hat{\boldsymbol{\mu}}_i$ og $\hat{P}(\omega_i|\boldsymbol{x}_k, \hat{\boldsymbol{\theta}}_i)$ er gitt ved henholdsvis

$$\hat{\boldsymbol{\mu}}_i = \frac{\sum_{k=1}^{n} P(\omega_i|\boldsymbol{x}_k, \hat{\boldsymbol{\mu}})\boldsymbol{x}_k}{\sum_{k=1}^{n} P(\omega_i|\boldsymbol{x}_k, \hat{\boldsymbol{\mu}})}$$

og

$$\hat{P}(\omega_i|\boldsymbol{x}_k, \hat{\boldsymbol{\theta}}) = \frac{p(\boldsymbol{x}_k|\omega_i, \hat{\boldsymbol{\theta}}_i)\hat{P}(\omega_i)}{\sum_{j=1}^{c} p(\boldsymbol{x}_k|\omega_j, \hat{\boldsymbol{\theta}}_j)\hat{P}(\omega_j)}.$$

Oppgave 2

La $p(x|\mu_1, \mu_2)$ være en blandingstetthet av to univariate, normalfordelte komponenter, der apriorisannsynlighetene $P(\omega_1)$ og $P(\omega_2)$ og variansene σ_1 og σ_2 for de to underliggende klassene er kjente, mens kun forventningsverdiene μ_1 og μ_2 er ukjente. Anta videre at $P(\omega_1) = P(\omega_2) = 1/2$ og $\sigma_1 = \sigma_2 = \sigma = 2$, og la treningssettet være gitt ved

$$\mathscr{X} = \{1.0, 2.9, 4.2, 5.1, 7.3, 5.4, 7.9, 8.8, 9.8, 11.2\}.$$

a) Bruk uttrykket for forventningsestimatet i oppgave 1 til å vise at estimatene i dette tilfellet er gitt ved det implisitte uttrykket

$$\hat{\mu}_i = \frac{\sum_{k=1}^{n} P(\omega_i|x_k, \hat{\mu}_1, \hat{\mu}_2)x_k}{\sum_{k=1}^{n} P(\omega_i|x_k, \hat{\mu}_1, \hat{\mu}_2)}, \quad i = 1, 2,$$

der

$$\hat{P}(\omega_i|x_k, \hat{\mu}_1, \hat{\mu}_2) = \frac{p(x_k|\omega_i, \hat{\mu}_i)P(\omega_i)}{\sum_{j=1}^{2} p(x_k|\omega_j, \hat{\mu}_j)P(\omega_j)} = \frac{\exp\{-\frac{1}{2}(\frac{x_k - \hat{\mu}_i}{\sigma})^2\}}{\exp\{-\frac{1}{2}(\frac{x_k - \hat{\mu}_1}{\sigma})^2\} + \exp\{-\frac{1}{2}(\frac{x_k - \hat{\mu}_2}{\sigma})^2\}}.$$

Her er $x_{k,k=1\ldots n}$ de umerkede samplene i treningssettet.

b) Lag et program (f.eks. i Matlab) som finner estimatene av de to forventningsverdiene ved iterasjon. Bruk startverdiene $\mu_1 = 2$ og $\mu_2 = 10$. Prøv også med andre startverdier. Gir dette andre resultater?