GAUSSIAN MIXTURE MODEL (CONTINUED)

MACHINE LEARNING 1, SS21

Ceca Kraišniković Dipl.-Ing. Dipl. inž. elektr. i računar.
Institute of Theoretical Computer Science
Graz University of Technology

GAUSSIAN MIXTURE MODEL: ESTIMATION OF PARAMETERS Σ_k, π_k

Derivation for Σ_k

- Log-likelihood function

$$\ln p(X \mid \pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k),$$

$\Theta = \{\pi, \mu, \Sigma\}$

- Set the derivative to zero:

$$\frac{\partial \ln p(X \mid \Theta)}{\partial \Sigma_k} \overset{!}{=} 0$$

- Solution (similar steps as in the derivation for μ_k from the previous lecture):

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma_{nk} (x_n - \mu_k)(x_n - \mu_k)^T,$$

$$N_k = \sum_{n=1}^{N} \gamma_{nk}$$

Derivation for π_k

- Set the derivative of the log-likelihood function w.r.t. π_k to zero:

$$\frac{\partial \ln p(X \mid \Theta)}{\partial \pi_k} \stackrel{!}{=} 0$$

- Optimization with a constraint:

$$\sum_{k=1}^{K} \pi_k = 1$$

- Use Lagrange multipliers, optimize the following objective function:

$$J(k) = \ln p(X \mid \Theta) + \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right)$$

and set $\frac{\partial J(k)}{\partial \pi_k}$ to zero, which yields:

$$\sum_{n=1}^{N} \frac{\mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_{m=1}^{K} \pi_m \mathcal{N}(x_n \mid \mu_m, \Sigma_m)} + \lambda \stackrel{!}{=} 0$$

Derivation for π_k

- Multiply everything with π_k:

$$\sum_{n=1}^{N} \frac{\pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_{m=1}^{K} \pi_m \mathcal{N}(x_n \mid \mu_m, \Sigma_m)} + \lambda \pi_k \stackrel{!}{=} 0$$

Posterior probabilities (responsibilities)

γ_{nk}

Derivation for π_k

- This give s:

$$\sum_{n=1}^{N} \gamma_{nk} + \lambda \pi_k \stackrel{!}{=} 0$$

- We set $\sum_{n=1}^{N} \gamma_{nk} = N_k$, and then sum over all k components:

$$\sum_{k=1}^{K} N_k + \sum_{k=1}^{K} \lambda \pi_k = 0$$

- By making use of the constraint $\sum_{k=1}^{K} \pi_k = 1$, i.e., $\lambda \sum_{k=1}^{K} \pi_k = \lambda$, and $\sum_{k=1}^{K} N_k = N$, we obtain $\lambda = -N$.

By substituting in the first equation (of this slide): $N_k - N\pi_k = 0$,

$$\pi_k = \frac{N_k}{N}.$$