GMMs by Maximam Likelihood Estimation

Jiajun He

We are trying to find a set of parameters $\boldsymbol{\theta} = \{\boldsymbol{\pi}, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, ..., \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K\}$, to maximize the likelihood function:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{D}) &= \prod_n P(\boldsymbol{x}_n|\boldsymbol{\theta}) \\ &= \prod_n \sum_i \pi_i \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\end{aligned}$$

So our optimization task is:

$$\arg\min_{\boldsymbol{\theta}} -\ell(\boldsymbol{\theta}; \boldsymbol{D})$$

$$s.t. \sum_i \pi_i = 1, |\boldsymbol{\Sigma}_i| \geq 0, i = 1, 2, ..., K$$

To meet the constraints, set auxiliary variables: vector$\boldsymbol{\alpha}$ and square matrix $\boldsymbol{A}_i$, $i = 1, 2, ...K$. Let $\pi_i = \frac{\exp(\alpha_i)}{\sum_j \exp(\alpha_j)}$, $\boldsymbol{\Sigma}_i = \boldsymbol{A}_i \boldsymbol{A}_i^T$.

Then we can use standard gradient optimization algorithm to find$\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{A}}_1, \hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{A}}_2, ..., \hat{\boldsymbol{\mu}}_K, \hat{\boldsymbol{A}}_K$, then calculate$\hat{\boldsymbol{\theta}} = \{\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Sigma}}_1, \hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{\Sigma}}_2, ..., \hat{\boldsymbol{\mu}}_K, \hat{\boldsymbol{\Sigma}}_K\}$.

Predict z for new $\boldsymbol{x}$:

After finding the ML estimator, we can predict class labels for new input $\boldsymbol{x}$.

$$\begin{aligned}&P(z|\boldsymbol{x}) \\ \propto &P(\boldsymbol{x}|z)P(z) \\ = &\hat{\pi}_z \mathcal{N}(\boldsymbol{x}|\hat{\boldsymbol{\mu}}_z, \hat{\boldsymbol{\Sigma}}_z)\end{aligned}$$