

GMMs by Maximam Likelihood Estimator

Jiajun He

We are trying to find a set of parameters $\theta = \{\pi, \mu_1, \Sigma_1, \mu_2, \Sigma_2, \dots, \mu_K, \Sigma_K\}$, to maximize the likelihood function:

$$\begin{aligned}\mathcal{L}(\theta; D) &= \prod_n P(\mathbf{x}_n | \theta) \\ &= \prod_n \sum_i \pi_i \mathcal{N}(\mathbf{x}_n | \mu_i, \Sigma_i)\end{aligned}$$

So our optimization task is:

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

$$s.t. ||\pi|| = 1, |\Sigma_i| \geq 0, i = 1, 2, \dots, K$$

To meet the constraints, set auxiliary variables: vector α and square matrix A_i , $i = 1, 2, \dots, K$.
Let $\pi_i = \frac{\exp(\alpha_i)}{\sum_j \exp(\alpha_j)}$, $\Sigma_i = A_i A_i^T$.

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

Predict z for new \mathbf{x} :

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

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