Expectation maximization (EM) for Mixture of Gaussians Model (GMM)

Sergej Dogadov

May 11, 2016

1. Given data *iid* $x = \{x_1, ...x_N\}, \qquad x_i \in \mathbb{R}^D, \qquad i = \overline{1..N}$

1. Given data *iid* $x = \{x_1, ...x_N\}, \qquad x_i \in \mathbb{R}^D, \qquad i = \overline{1..N}$
2. Model parameters $\theta = \{\theta_1, ...\theta_M\}, \qquad$ e.g $\theta = \{\mu, \Sigma\}$

1. Log-likelihood

$$\ln p(x|\theta) = \sum_{i=1}^{N} \ln p(x_i|\theta)$$

2. Marginal log-likelihood (z - latent variable)

$$\ln p(x|\theta) = \ln \int p(x, z|\theta) dz = \ln \int p(x|z, \theta) p(z|\theta) dz$$

1. Log-$\mathbb{E}$xpectation

$$\ln p(x|\theta) = \ln \int p(x,z|\theta)dz = \ln \int q(z)\frac{p(x,z|\theta)}{q(z)}dz$$

$$\ln p(x|\theta) = \ln \mathbb{E}\left[\frac{p(x,z|\theta)}{q(z)}\right]$$

2. *Jensen's inequality* for concave function f

$$f(\mathbb{E}[X]) \geq \mathbb{E}[f(X)]$$

1. Evidence lower bound (*ELBO*)

$$\ln p(x|\theta) = \ln \mathbb{E}\left[\frac{p(x,z|\theta)}{q(z)}\right] \geq \mathbb{E}\left[\ln \frac{p(x,z|\theta)}{q(z)}\right] = \mathcal{L}(q(z),\theta)$$

2. Itegral form

$$\mathcal{L}(q(z),\theta) = \int q(z) \ln \frac{p(x,z|\theta)}{q(z)} dz$$

1. Maximize ELBO w.r.t. $q(z)$

$$\max_{q(z)} \mathcal{L}(q(z), \theta) \qquad s.t \int q(z)dz = 1$$

1. Maximize ELBO w.r.t. $q(z)$

$$\max_{q(z)} \mathcal{L}(q(z), \theta) \qquad s.t \int q(z)dz = 1$$

2. Constrained functional optimization

$$\mathcal{F}[q(z)] = \mathcal{L}(q(z), \theta) + \lambda(\int q(z)dz - 1)$$

$$\mathcal{L}(q(z), \theta) = \int q(z) \ln \frac{p(x, z|\theta)}{q(z)} dz$$

$$\mathcal{L}(q(z), \theta) = \int q(z) \ln p(x, z|\theta) - q(z) \ln q(z)) dz$$

1. Constraint functional

$$\mathcal{F}[q(z)] = \int q(z) \ln \frac{p(x, z|\theta)}{q(z)} dz + \lambda\left(\int q(z) dz - 1\right)$$

2. Functional derivarive with respect to $q(z)$

$$\frac{\delta \mathcal{F}[q(z)]}{\delta q(z)} = \ln p(x, z|\theta) - \ln q(z) - 1 + \lambda = 0$$

3. Stationary condition for $q(z)$

$$\ln q(z) = \ln p(x, z|\theta) + \lambda - 1 \Rightarrow q(z) = p(x, z|\theta) e^{\lambda - 1}$$

1. Constraint functional

$$\mathcal{F}[q(z)] = \int q(z) \ln \frac{p(x, z|\theta)}{q(z)} dz + \lambda(\int q(z) dz - 1)$$

2. Functional derivarive with respect to λ

$$\frac{\delta \mathcal{F}[q(z)]}{\delta \lambda} = \int q(z) dz - 1 = 0$$

3. Stationary condition for λ

$$\int q(z) dz = 1$$

$$q(z) = p(x, z|\theta)e^{\lambda-1}$$

$$\int p(x, z|\theta)e^{\lambda-1}dz = 1 \qquad \Rightarrow \qquad e^{\lambda-1} = \frac{1}{\int p(x, z|\theta)dz}$$

$$\boxed{q^*(z) = \frac{p(x, z|\theta)}{\int p(x, z|\theta)dz}} \qquad \textit{or} \qquad \boxed{q^*(z) \propto p(x, z|\theta)}$$

EM-Algorithm

1. E-step (maximize a distibution $q(z)$)

$$q^*(z) = \frac{p(x, z|\theta)}{\int p(x, z|\theta) dz} \qquad \textit{or} \qquad q^*(z) \propto p(x, z|\theta)$$

2. M-step (maximize model parameters θ)

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \int q^*(z) \ln \frac{p(x, z|\theta)}{q(z)} dz$$

EM for Mixture of Gaussians Model

1. Chain rule

$$\ln p(x, z|\theta) = \ln p(x|z, \theta)p(z|\theta)$$

2. Gaussian distibuted with hyperparameters μ, Σ

$$\ln p(x|z) = \ln \prod_{n=1}^{N} \prod_{k=1}^{K} \mathcal{N}(x_n|\mu_k, \Sigma_k)^{z_{nk}}$$

3. Multinomial distributed with hyperparamter $\pi \in \mathbb{R}^K$

$$\ln p(z) = \ln \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{nk}}$$

EM for Mixture of Gaussians Model

1. E-step

$$\ln \gamma_{nk} = \ln q^*(z) \propto \sum_{n=1}^{N} \sum_{k=1}^{N} z_{nk}(\ln \mathcal{N}(x_n|\mu_k, \Sigma_k) + \ln \pi_k)$$

$$\ln \gamma_{nk} \propto \sum_{n=1}^{N} \ln \mathcal{N}(x_i|\mu_k, \Sigma_k) + \ln \pi_k$$

$$\ln \mathcal{N}(x_n|\mu_k, \Sigma_k) = -\frac{1}{2}\left[\ln|\Sigma_k| + K\ln 2\pi + (x_i - \mu_k)^\top \Sigma_k^{-1}(x_i - \mu_k)\right]$$

1. Compute log γ

$$\boxed{\ln \gamma_{nk} \propto \sum_{i=1}^{N} \ln \pi_k - \frac{1}{2}\left[\ln|\Sigma_k| + (x_i - \mu_k)^\top \Sigma_k^{-1}(x_i - \mu_k)\right]}$$

2. Normalize γ By using log-sum-exp-trick

$$scipy.mics.logsumexp$$

EM for Mixture of Gaussians Model

M-step

1. Number of cluster point assignments

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

2. Cluster proportions

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

3. Cluster means

$$\mu_k := \frac{1}{N} \sum_{n=1}^{N} \gamma_{nk} x_n$$

EM for Mixture of Gaussians Model

M-step

1. Cluster covariance matrices

$$\Sigma_k := \frac{1}{N}\sum_{n=1}^{N}\gamma_{nk}(x_n - \mu_k)(x_n - \mu_k)^\top$$

2. Iterate E and M steps till convergence
   log − likelihood convergence