Discussion 6
EM Algorithm
EM in mixture model

田鹏超

tianpch@shanghaitech.edu.cn

EM Algorithm - Precisely

EM is a general procedure for learning from partly observed data

Given observed variables X, unobserved Z (X={F,A,H,N}, Z={S}) ✓

Define $Q(\theta'|\theta) = E_{P(Z|X,\theta)}[\log P(X, Z|\theta')]$

current

M step new

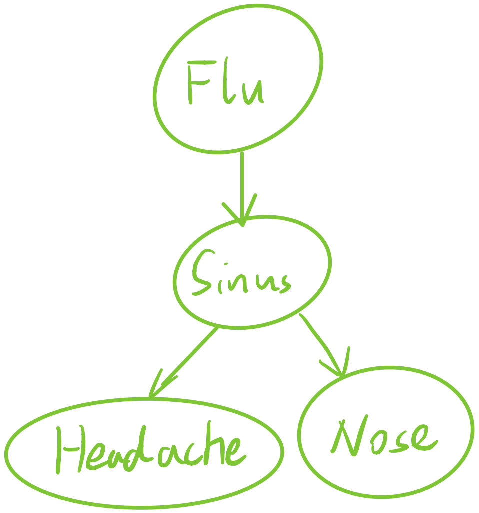
Iterate until convergence:

- E Step: Use X and current θ to calculate P(Z|X,θ)
- M Step: Replace current θ by

$$\theta \leftarrow \arg\max_{\theta'} Q(\theta'|\theta)$$

Guaranteed to find local maximum.
Each iteration increases $E_{P(Z|X,\theta)}[\log P(X, Z|\theta')]$

eg 1.



$X = \{F, N\}$ Observed variables

$Z = \{S, H\}$ Latent variables

$\{F, H, N\}$ 0/1 binary variables

$S \in \{0, 1, 2\}$

There are K training examples in total.

① Derive E step.

② Derive M step.

①

12

use EM to solve Gaussian mixture model

probability density function of one-dimensional Gaussian distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Joint probability density function for N-dimension variable X.

$$f(x) = \frac{1}{2\pi^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(X-u)^T \Sigma^{-1}(X-u)\right), X = (x_1, x_2, \ldots, x_n)$$

Gaussian Mixture Model (GMM) with K Gaussian model

$$p(x)=\sum_{k=1}^{K} p(k) p(x \mid k)=\sum_{k=1}^{K} \pi_{k} N\left(x \mid u_{k}, \Sigma_{k}\right)$$

How to use EM Algorithm to solve GMM?

To solve GMM, it's actually to figure out parameters $\theta = (\mu, \Sigma, \pi)$

First, assume the latent variables $Z = (z_1, \ldots, z_K)$ is a binary K-dimensional variable having only a single component equal to 1.
In fact, the latent variable describes the probability of selecting the k-th Gaussian model for each sample.

$$p\left(z_k = 1 | \theta\right) = \pi_k$$

$$p\left(y \mid z_k = 1, \theta\right) = N\left(y \mid \mu_k, \Sigma_k\right)$$

$$p(y) = \sum_z p(z) p(y \mid z) = \sum_{\mathrm{k}=1}^{K} \pi_k N\left(y \mid \mu_k, \Sigma_k\right)$$

For T training examples in total, $Y = (y_1, \ldots, y_T)$. If Z is known the well-informed data should be:

$$(y_t, z_{t,1}, z_{t,2} \ldots z_{t,K}), t = 1, 2 \ldots T$$

However, Z is unkown, we don't know which Gaussian model y is sampled from.

E-step

$$
\begin{aligned}
E\left(z_{t,k} \mid y_t, \mu^i, \Sigma^i, \pi^i\right) &= p(z_{t,k} = 1 \mid y_t, \mu^i, \Sigma^i, \Pi^i) \\
&= \frac{p\left(z_{t,k} = 1, y_t \mid \mu^i, \Sigma^i, \Pi^i\right)}{p\left(y_t\right)} \\
&= \frac{p\left(z_{t,k} = 1, y_t \mid \mu^i, \Sigma^i, \pi^i\right)}{\sum_{k=1}^{K} p\left(z_{t,k} = 1, y_t \mid \mu^i, \Sigma^i, \pi^i\right)} \\
&= \frac{p\left(y_t \mid Y_{t,k} = 1, \mu^i, \Sigma^i, \pi^i\right) p\left(z_{t,k} = 1 \mid \mu^i, \Sigma^i, \pi^i\right)}{\sum_{k=1}^{K} p\left(yt \mid z_{t,k} = 1, \mu^i, \Sigma^i, \pi^i\right) p\left(z_{t,k} = 1 \mid \mu^i, \Sigma^i, \pi^i\right)} \\
&= \frac{\pi_k^i N\left(y_t; \mu_k^i, \Sigma_k^i\right)}{\sum_{k=1}^{K} \pi_k^i N\left(y_t; \mu_k^i, \Sigma_k^i\right)}
\end{aligned}
$$

$$Q\left(\mu, \Sigma, \pi, \mu^{i}, \Sigma^{i}, \pi^{i}\right) = E_{Z}\left[\ln p(y, Z \mid \mu, \Sigma, \pi) \mid Y, \mu^{i}, \Sigma^{i}, \pi^{i}\right]$$

The likelihood functions is:

$$\begin{aligned}
L(\mu, \Sigma, \pi) &= p(y, Z \mid \mu, \Sigma, \pi) \\
&= \prod_{t=1}^{T} p\left(y_{t}, z_{t,1}, z_{t,2} \ldots z_{t,K} \mid \mu, \Sigma, \pi\right) \\
&= \prod_{t=1}^{T} \prod_{k=1}^{K} \left(\pi_{k} N\left(y_{t}; \mu_{k}, \Sigma_{k}\right)\right)^{z_{t,k}} \\
&= \prod_{k=1}^{K} \pi_{k}^{\sum_{t=1}^{T} z_{t,k}} \prod_{t=1}^{T} \left(N\left(y_{t}; \mu_{k}, \Sigma_{k}\right)\right)^{Y_{t,k}}
\end{aligned}$$

M-step

$$\mu^{i+1}, \Sigma^{i+1}, \pi^{i+1} = \arg\max Q\left(\mu, \Sigma, \pi, \mu^i, \Sigma^i, \pi^i\right)$$

Set the derivative with respect to μ_k, Σ_k, π_k seperately to 0.

$$\mu_k^{i+1} = \frac{\sum_{t=1}^{T} \frac{\pi_k^i N\left(y_t;\mu_k^i,\Sigma_k^i\right)}{\sum_{k=1}^{K} \pi_k^i N\left(y_t;\mu_k^i,\Sigma_k^i\right)} y_t}{E(\gamma_{t,k}|y_t,\mu^i,\Sigma^i,\pi^i)}, k = 1, 2 \ldots K$$

$$\Sigma_k^{i+1} = \frac{\sum_{t=1}^{T} \frac{\pi_k^i N\left(y_t;\mu_k^i,\Sigma_k^i\right)}{\sum_{k=1}^{K} \pi_k^i N\left(y_t;\mu_k^i,\Sigma_k^i\right)} \left(y_t - \mu_k^i\right)^2}{E(\gamma_{t,k}|y_t,\mu^i,\Sigma^i,\pi^i)}, k = 1, 2 \ldots K$$

$$\pi_k^{i+1} = \frac{E\left(\gamma_{t,k}|y_t,\mu^i,\Sigma^i,\Pi^i\right)}{T}, k = 1, 2 \ldots K$$