Mixtures of Gaussians

Thursday, April 5, 2018 10:48 PM

Consider $\mathcal{D} = \{ x^{(i)} \mid 1 \le i \le m \,;\, x^{(i)} \in \mathbb{R}^d \}$

We will consider a generative model for this data.

We will assume a discrete latent state variable $z^{(i)} \in \{1, \dots, K\}$

with a prior distribution $P(z^{(i)} = k) = \pi_k$, $k = 1..K$

such that $\sum_k \pi_k = 1$

We further assume that

$P(x^{(i)} \mid z^{(i)} = k)$ is a base distribution

such that

$$P(x^{(i)}) = \sum_{k=1}^{K} P(z^{(i)} = k) P(x^{(i)} \mid z^{(i)} = k)$$

This is a mixture distribution where π_k are the mixture proportions.

For a Gaussian mixture,

$$z^{(i)} \sim \text{Multinomial}(\pi) \,;\, \pi_k > 0,\ \sum_k \pi_k = 1$$

$$x^{(i)} \Big|_{z^{(i)} = k} \sim N(\mu_k, \Sigma_k)$$

$P(x^{(i)})$ is a linear superposition of K multivariate normal distributions.

We can approximate any pdf in $\mathbb{R}^d$ with a sufficient number of components (Gaussian components)

Given π_k, $k = 1..K$ mixture proportions

μ_k, Σ_k, $k = 1..K$ Gaussian mixture components

we can generate an x from this Gaussian mixture model. (GMM)

- select $z^{(i)} = k$ based on $\{\pi_1, \dots, \pi_K\}$
- generate $x^{(i)}$ from $N(x^{(i)} \mid \mu_k, \Sigma_k)$

Mixture models are useful for clustering. If we know that $\mathcal{D}$ is a dataset that comes from a GMM with parameters $\pi, \mu, \Sigma = \theta$

We can infer $z^{(i)}$ for each $x^{(i)}$

$$P(z^{(i)} = k \mid x^{(i)} ; \theta) = \frac{P(z^{(i)} = k) P(x^{(i)} \mid z^{(i)} = k ; \theta)}{\sum_{k'} P(z^{(i)} = k') P(x^{(i)} \mid z^{(i)} = k' ; \theta)}$$

In hard clustering, we assign $z^{(i)}$ to be the k which maximizes $p(z^{(i)} = k \mid x^{(i)} ; \theta)$

$$\begin{aligned} z^{(i)*} &= \operatorname*{argmax}_k P(z^{(i)} = k \mid x^{(i)} ; \theta) \\ &= \operatorname*{argmax}_k P(z^{(i)} = k) P(x^{(i)} \mid z^{(i)} = k ; \theta) \\ &= \operatorname*{argmax}_k \pi_k \cdot N(x^{(i)} ; \mu_k, \Sigma_k) \\ &= \operatorname*{argmax}_k \log \pi_k + \log N(x^{(i)} ; \mu_k, \Sigma_k) \end{aligned}$$

In soft clustering, we compute a vector of size K denoting the probability that $x^{(i)}$ was generated by component k.

Estimating GMMs from data

Given $\mathcal{D} = \{ x^{(i)} \mid 1 \le i \le m \,;\, x^{(i)} \in \mathbb{R}^d \}$

find the GMM parameters $\pi_1 \cdots \pi_K$ mixture proportions

$\mu_1 \cdots \mu_K$, $\Sigma_1 \cdots \Sigma_K$ } multivariate normals

MLE approach to parameter estimation

$$\begin{aligned} \mathcal{L}(\mathcal{D} ; \pi, \mu, \Sigma) &= \prod_{i=1}^{m} P(x^{(i)} ; \pi, \mu, \Sigma) \\ & \prod_{i=1}^{m} \sum_{k=1}^{K} P(z^{(i)} = k ; \pi) P(x^{(i)} \mid z^{(i)} = k ; \mu, \Sigma) \end{aligned}$$

The log likelihood is

$$\ell(\mathcal{D} ; \pi, \mu, \Sigma) = \sum_{i=1}^{m} \log \left[\sum_{k=1}^{K} P(z^{(i)} = k ; \pi) P(x^{(i)} \mid z^{(i)} = k ; \mu, \Sigma) \right]$$

If we knew the $z^{(i)}$, we can rewrite the log likelihood as

$$\ell_z(\mathcal{D} ; \pi, \mu, \Sigma) = \sum_{i=1}^{m} \log \pi_{z^{(i)}} + \log N(x^{(i)} ; \mu_{z^{(i)}}, \Sigma_{z^{(i)}})$$

If we maximize this expression wrt π, μ, Σ by setting $\frac{\partial \ell_z}{\partial \pi} = 0$, $\frac{\partial \ell_z}{\partial \mu} = 0$, $\frac{\partial \ell_z}{\partial \Sigma} = 0$

then we obtain

$$\pi_k = \frac{1}{m} \sum_{i=1}^{m} I(z^{(i)} = k)$$

$$\mu_k = \frac{\sum_{i=1}^{m} I(z^{(i)} = k) x^{(i)}}{\sum_{i=1}^{m} I(z^{(i)} = k)}$$

But, we do not know the $z^{(i)}$! We just have the $x^{(i)}$!

If we knew $\pi, \mu, \Sigma \longrightarrow$ we can infer $z^{(i)}$

If we knew $z^{(i)} \underset{\text{(MLE)}}{\longrightarrow}$ we can infer π, μ, Σ.

The EM algorithm

An iterative algorithm that alternates between inferring the missing $z^{(i)}$ values assuming parameters are known (E-step) and optimizing the parameters given the $z^{(i)}$ (M-step)

1. Guess values of $\theta = [\pi, \mu, \Sigma]$
2. E-step: calculate the responsibility of each component toward generating $x^{(i)}$

$$r_k^{(i)} = P(z^{(i)} = k \mid x^{(i)} ; \theta) = \frac{\pi_k N(x^{(i)} ; \mu_k, \Sigma_k)}{\sum_{k'} \pi_{k'} N(x^{(i)} ; \mu_{k'}, \Sigma_{k'})}$$

3. M-step: Given $r_k^{(i)}$ and $x^{(i)}$ $1 \le i \le m$, $k = 1..K$

re-estimate π, μ, Σ

$$\mu_k = \frac{\sum_{i=1}^{m} r_k^{(i)} x^{(i)}}{\sum_{i=1}^{m} r_k^{(i)}}$$

$$\pi_k = \frac{1}{m} \sum_{i=1}^{m} r_k^{(i)}$$

$$\Sigma_k = \frac{\sum_{i=1}^{m} r_k^{(i)} (x^{(i)} - \mu_k)(x^{(i)} - \mu_k)^T}{\sum_{i=1}^{m} r_k^{(i)}}$$

Hard vs soft EM

Hard EM

1. Guess $\theta = \pi, \mu, \Sigma$
2. E step: $P(z^{(i)} = k \mid x^{(i)} ; \theta)$

$$z^{(i)} = \operatorname*{argmax}_k \frac{\pi_k N(x^{(i)} ; \mu_k, \Sigma_k)}{\sum_{k'} \pi_{k'} N(x^{(i)} ; \mu_{k'}, \Sigma_{k'})}$$

3. M-step: Restimate θ given z

$$\pi_k = \frac{1}{m} \sum_{i=1}^{m} I(z^{(i)} = k)$$

$$\mu_k = \frac{\sum_{i=1}^{m} I(z^{(i)} = k) x^{(i)}}{\sum_{i=1}^{m} I(z^{(i)} = k)}$$

$$\Sigma_k = \frac{\sum_{i=1}^{m} I(z^{(i)} = k)(x^{(i)} - \mu_k)(x^{(i)} - \mu_k)^T}{\sum_{i=1}^{m} I(z^{(i)} = k)}$$

Soft EM

1. Guess $\theta = \pi, \mu, \Sigma$
2. E step: $P(z^{(i)} = k \mid x^{(i)} ; \theta)$

$$r_k^{(i)} = \frac{\pi_k N(x^{(i)} ; \mu_k, \Sigma_k)}{\sum_{k'} \pi_{k'} N(x^{(i)} ; \mu_{k'}, \Sigma_{k'})}$$

3. Restimate θ given r

$$\pi_k = \frac{1}{m} \sum_{i=1}^{m} r_k^{(i)}$$

$$\mu_k = \frac{\sum_{i=1}^{m} r_k^{(i)} x^{(i)}}{\sum_{i=1}^{m} r_k^{(i)}}$$

$$\Sigma_k = \frac{\sum_{i=1}^{m} r_k^{(i)} (x^{(i)} - \mu_k)(x^{(i)} - \mu_k)^T}{\sum_{i=1}^{m} r_k^{(i)}}$$

Demos here