E9 205 Machine Learning for Signal Processing

MLE for Gaussian Mixture Model

11-09-2017

Instructor - Sriram Ganapathy (sriram@ee.iisc.ernet.in)

Teaching Assistant - Aravind Illa (aravindece77@gmail.com).

Expectation Maximization Algorithm

- Iterative procedure.
- Assume the existence of hidden variable $\mathbf{z}_i$ associated with each data sample $\mathbf{x}_i$
- Let the current estimate (at iteration n) be $\mathbf{\Theta}^n$ Define the Q function as

$$Q(\mathbf{\Theta}, \mathbf{\Theta}^n) = E_{\mathbf{z}|\mathbf{X},\mathbf{\Theta}^n}\left[\log(P(\mathbf{X}, \mathbf{z}|\mathbf{\Theta})\right]$$

$$= \sum_{\mathbf{z}} \log(P(\mathbf{X}, \mathbf{z}|\mathbf{\Theta}))P(\mathbf{z}|\mathbf{X}, \mathbf{\Theta}^n)$$

Established 1911

Expectation Maximization Algorithm

- It can be proven that if we choose

$$\mathbf{\Theta}^{n+1} = arg\max_{\mathbf{\Theta}} Q(\mathbf{\Theta}, \mathbf{\Theta}^{n})$$

  then $L(\mathbf{\Theta}^{n+1}) \geq L(\mathbf{\Theta}^{n})$

- In many cases, finding the maximum for the Q function may be easier than likelihood function w.r.t. the parameters.

- Solution is dependent on finding a good choice of the hidden variables which eases the computation

- Iteratively improve the log-likelihood function.

Established 1911

INDIAN INSTITUTE OF SCIENCE

भारतीय विज्ञान संस्थान

EM Algorithm Summary

- Initialize with a set of model parameters (n=1)
- Compute the conditional expectation (E-step)

$$E_{\mathbf{z}|\mathbf{X},\mathbf{\Theta}^n}\left[\log(P(\mathbf{X},\mathbf{z}|\mathbf{\Theta})\right]$$

- Maximize the conditional expectation w.r.t. parameter. (M-step) (n = n+1)
- Check for convergence
- Go back to E-step if model has not converged.

Established 1911

EM Algorithm for GMM

- The hidden variables $\mathbf{z}_i = l$ will be the index of the mixture component which generated $\mathbf{x}_i$

- Re-estimation formulae

$$\alpha_\ell^{new} = \frac{1}{N} \sum_{i=1}^{N} p(\ell | x_i, \Theta^g)$$

$$\mu_\ell^{new} = \frac{\sum_{i=1}^{N} x_i p(\ell | x_i, \Theta^g)}{\sum_{i=1}^{N} p(\ell | x_i, \Theta^g)}$$

$$\Sigma_\ell^{new} = \frac{\sum_{i=1}^{N} p(\ell | x_i, \Theta^g)(x_i - \mu_\ell^{new})(x_i - \mu_\ell^{new})^T}{\sum_{i=1}^{N} p(\ell | x_i, \Theta^g)}$$

Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान