Notes

Sinan Shi

August 13, 2019

Contents

1 EM Algorithm

1.1 The General EM Algorithm

Given a joint distribution $p(\mathbf{X}, \mathbf{Z}|\theta)$ over observed variables $\mathbf{X}$ and latent variables $\mathbf{Z}$, governed by parameters θ, the goal is to maximize the likelihood function $p(\mathbf{X}|\theta)$.

1. Choose an initial setting for the parameters θ^{old}.
2. **E step** Evaluate $p(\mathbf{Z}|\mathbf{X}, \theta^{\mathbf{old}})$
3. **M step** Evaluate $\theta^{\mathbf{new}}$ given by

$$\theta^{\mathbf{new}} = \underset{x}{\operatorname{argmax}}\, Q(\theta, \theta^{\mathbf{old}})$$

$$Q(\theta, \theta^{old}) = \sum_{\mathbf{z}} p(\mathbf{Z}|\mathbf{X}, \theta^{old}) \log p(\mathbf{X}, \mathbf{Z}|\theta) \tag{1}$$

1.2 HMM

$$p(\mathbf{X}, \mathbf{Z}|\theta) = p(\mathbf{z}_1|\pi) \prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1}) \prod_{m=1}^{N} p(\mathbf{x}_m|\mathbf{z}_m, \phi) \tag{2}$$

$$\gamma(\mathbf{z}_n) = p(\mathbf{z}_n|\mathbf{X}, \theta^{old})$$

$$\xi(\mathbf{z}_{n-1}, \mathbf{z}_n) = p(\mathbf{z}_{n-1}, \mathbf{z}_n|\mathbf{X}, \theta^{old})$$

(Bilmes, 1998) (Bishop, 2006)

Reference

Bilmes,J. (1998) A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models.

Bishop,C.M. (2006) Pattern recognition and machine learning (information science and statistics) Springer-Verlag, Berlin, Heidelberg.