CS-667 Advanced Machine Learning

Nazar Khan

PUCIT

Lecture 19
The EM Algorithm
May 4, 2016

The EM Algorithm I

- We have seen that K-means and GMMs are examples of latent variable models.
- Specifically for GMMs, we have seen an incremental algorithm for learning the parameters via ML.
- That algorithm is actually an instance of a powerful framework called *Expectation-Maximisation (EM)*.
- EM is used for *solving latent variable problems via ML*.
- We will now present a more general explanation of the EM algorithm.

- Maximum likelihood is equivalent to maximising the log-likelihood $\ln p(\mathbf{X}|\boldsymbol{\theta})$.
- Using the sum-rule

$$\ln p(\mathbf{X}|\boldsymbol{\theta}) = \ln \left(\sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) \right)$$

- Maximisation is no longer straight-forward since ln is 'blocked'by the summation.
- So we take another approach.

- We will denote $\{\mathbf{X}, \mathbf{Z}\}$ as the *complete* dataset.
- We will denote $\{\mathbf{X}\}$ as the *incomplete* dataset.
- The goal now is to maximise the complete-data log-likelihood function $p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$.
- But for that we need to know the values of $\mathbf{Z}$ which are unobserved. What *can* be computed about $\mathbf{Z}$, however, is the posterior $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta})$.
- So instead of the *uncomputable value* of log-likelihood, the *next best computable number* would be its *expected-value under the posterior* $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta})$.

- This yields the *E-step* of the EM algorithm.

$$\mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{\text{old}}}[\ln p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})] = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})$$

- Since we are eventually interested in optimal parameters $\boldsymbol{\theta}^*$ we treat this expectation as a function of $\boldsymbol{\theta}$ and denote it by $\mathcal{Q}(\boldsymbol{\theta},\boldsymbol{\theta}^{\text{old}})$.
- The *M-step* corresponds to maximising this expectation

$$\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta},\boldsymbol{\theta}^{\text{old}})$$

- In short, EM replaces the log-likelihood by the *expected log-likelihood* and maximises it.
- Each EM cycle either moves toward or stays at a local maximum of $\ln p(\mathbf{X}|\boldsymbol{\theta})$.

The General EM Algorithm

Goal is to maximise likelihood $p(\mathbf{X}|\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ by introducing joint distribution $p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$ involving latent variables $\mathbf{Z}$.

1. Choose initial $\boldsymbol{\theta}^{\text{old}}$
2. **E-step**: Evaluate $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}})$
3. **M-step**: Obtain new estimate $\boldsymbol{\theta}^{\text{new}}$ by maximising the expectation $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$

$$\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$$

   where $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$.
4. Check for convergence of either log-likelihood or parameters. If not converged, then

$$\boldsymbol{\theta}^{\text{old}} \leftarrow \boldsymbol{\theta}^{\text{new}} \tag{1}$$

   and return to step 2.

Extensions of EM

- EM for MAP estimation via prior $p(\boldsymbol{\theta})$ amounts to modifying the M-step only.

$$\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) + p(\boldsymbol{\theta})$$

- For problems with 'difficult' M-step, maximisation can be replaced by a step that just increases $\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$. This is known as the *Generalised EM* algorithm.

Proof of Convergence of EM