

These notes are intended to give a summary of relevant concepts from the lectures which are helpful to complete the exercises. It is not intended to cover the lectures thoroughly. Learning this content is not a replacement for working through the lecture material and the exercises.

EM algorithm — The expectation maximisation (EM) algorithm can be used to learn the parameters θ of a statistical model $p(\mathbf{v}, \mathbf{h}; \theta)$ with latent (unobserved) variables \mathbf{h} and visible (observed) variables \mathbf{v} for which we have data \mathcal{D} . It updates the parameters θ by iterating between the expectation (E) and the maximisation (M) step:

$$\text{E-step: compute } J(\theta) = \mathbb{E}_{p(\mathbf{h}|\mathcal{D};\theta_{\text{old}})}[\log p(\mathcal{D}, \mathbf{h}; \theta)] \quad \text{M-step: } \theta_{\text{new}} \leftarrow \underset{\theta}{\operatorname{argmax}} J(\theta) \quad (1)$$

The update rule produces a sequence of parameters for which the log-likelihood is guaranteed to never decrease, i.e. $\ell(\theta_{\text{new}}) \geq \ell(\theta_{\text{old}})$.