These notes summarise selected lecture concepts and are not a substitute for working through the lecture slides, tutorials, and self-study exercises. Feel free to personalise and develop them into your own summary sheet.

Learning problem — Model: Stationary HMM with visibles $v_i \in \{1, \ldots, M\}$, latents $h_i \in \{1, \ldots, K\}$ and parametrisation

$$p(h_1 = k; \mathbf{a}) = a_k \qquad p(h_i = k | h_{i-1} = k'; \mathbf{A}) = A_{k,k'} \qquad p(v_i = m | h_i = k; \mathbf{B}) = B_{m,k}$$

Data: $\mathcal{D} = \{\mathcal{D}_1, \ldots, \mathcal{D}_n\}$, where each $\mathcal{D}_j$ is a sequence of visibles of length d_j. Task: Determine MLE for $\boldsymbol{\theta} = (\mathbf{a}, \mathbf{A}, \mathbf{B})$. Constraints: parameters are all non-negative; $\mathbf{a}$ and the values of each column of $\mathbf{A}$ and $\mathbf{B}$ sum to one.

Objective for EM — The objective for M-step of the EM algorithm is

$$\begin{aligned} J(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}}) = &\sum_{j=1}^{n} \sum_{k} p(h_1 = k | \mathcal{D}_j; \boldsymbol{\theta}_{\text{old}}) \log a_k + \\ &\sum_{j=1}^{n} \sum_{i=2}^{d_j} \sum_{k,k'} p(h_i = k, h_{i-1} = k' | \mathcal{D}_j; \boldsymbol{\theta}_{\text{old}}) \log A_{k,k'} + \\ &\sum_{j=1}^{n} \sum_{i=1}^{d_j} \sum_{m,k} \mathbb{1}(v_i^{(j)} = m) p(h_i = k | \mathcal{D}_j, \boldsymbol{\theta}_{\text{old}}) \log B_{m,k} \end{aligned}$$

subject to the constraints on the parameters.

EM (Baum-Welch) algorithm — Given parameters $\boldsymbol{\theta}_{\text{old}}$:

1. For each sequence $\mathcal{D}_j$ compute the posteriors

$$p(h_i, h_{i-1} \mid \mathcal{D}_j; \boldsymbol{\theta}_{\text{old}}) \qquad p(h_i \mid \mathcal{D}_j; \boldsymbol{\theta}_{\text{old}})$$

   using the alpha-beta recursion.

2. Update the parameters

$$\begin{aligned} a_k &= \frac{1}{n} \sum_{j=1}^{n} p(h_1 = k | \mathcal{D}_j; \boldsymbol{\theta}_{\text{old}}) \\ A_{k,k'} &= \frac{\sum_{j=1}^{n} \sum_{i=2}^{d_j} p(h_i = k, h_{i-1} = k' | \mathcal{D}_j; \boldsymbol{\theta}_{\text{old}})}{\sum_{k=1}^{K} \sum_{j=1}^{n} \sum_{i=2}^{d_j} p(h_i = k, h_{i-1} = k' | \mathcal{D}_j; \boldsymbol{\theta}_{\text{old}})} \\ B_{m,k} &= \frac{\sum_{j=1}^{n} \sum_{i=1}^{d_j} \mathbb{1}(v_i^{(j)} = m) p(h_i = k | \mathcal{D}_j; \boldsymbol{\theta}_{\text{old}})}{\sum_{j=1}^{n} \sum_{i=1}^{d_j} p(h_i = k | \mathcal{D}_j; \boldsymbol{\theta}_{\text{old}})} \end{aligned}$$

Repeat step 1 and 2 using the new parameters for $\boldsymbol{\theta}_{\text{old}}$. Stop if change in likelihood or parameters is less than a threshold.