Tutorial Week 8

COMP9418 – Advanced Topics in Statistical Machine Learning, 17s2, UNSW Sydney

Instructor: Edwin V. Bonilla

Last Update: Friday 8th September, 2017 at 11:53

1. When deriving the forward recursion for the filtering distribution in a HMM, we write:

$$p(z_t, x_{1:t}) = \sum_{z_{t-1}} p(z_t, z_{t-1}, x_{1:t-1}, x_t) \tag{1}$$

$$= p(x_t|z_t) \sum_{z_{t-1}} p(z_t|z_{t-1}) p(z_{t-1}, x_{1:t-1}) \tag{2}$$

$$\alpha(z_t) = p(x_t|z_t) \sum_{z_{t-1}} p(z_t|z_{t-1}) \alpha(z_{t-1}), \quad t > 1. \tag{3}$$

   (a) Explain what conditional independence assumptions were used in going from Equation 1 to Equation 2.

   (b) Show that these independence statements hold for an HMM using d-separation.

2. When deriving the backward recursion for the smoothing distribution in a HMM, we write:

$$p(x_{t:T}|z_{t-1}) = \sum_{z_t} p(x_t, x_{t+1:T}, z_t|z_{t-1}) \tag{4}$$

$$= \sum_{z_t} p(x_t|z_t) \underbrace{p(x_{t+1:T}|z_t)}_{\beta(z_t)} p(z_t|z_{t-1}) \tag{5}$$

$$\beta(z_{t-1}) = \sum_{z} p(x_t|z_t) p(z_t|z_{t-1}) \beta(z_t) \tag{6}$$

   (a) Explain what conditional independence assumptions were used in going from Equation 4 to Equation 5.

   (b) Show that these independence statements hold for an HMM using d-separation.

3. (Original version from Bishop, PRML, 2006) By using d-separation, show that the distribution $p(\mathbf{x}_1, ..., \mathbf{x}_T)$ of the observed data for the state space model (represented by the directed graph in slide 26 of the lecture slides) does not satisfy any conditional independence properties and hence does not exhibit the Markov property at any finite order.

4. (Original version from Bishop, PRML, 2006) Show that if any elements of the parameters $\boldsymbol{\pi}$ or $\mathbf{A}$ for a hidden Markov model are initially set to zero, then those elements will remain zero in all subsequent updates of the EM algorithm.

5. Consider a hidden Markov model with continuous observations and, as before, discrete latent states $z_t \in \{1, \ldots, K\}$. A possible choice for the emission densities is to consider a Gaussian distribution:

$$p(\mathbf{x}_t|z_t = k) = \mathcal{N}(\mathbf{x}_t|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \tag{7}$$

Derive the updates for the mean $\boldsymbol{\mu}_k$ and covariance $\boldsymbol{\Sigma}_k$ corresponding to the M-step of the EM algorithm.