8 Machine Learning and Bayesian Inference (jt796)

Suppose a Bayesian network has the form of a *chain*: a sequence of Boolean random variables $X_1, \ldots X_n$ where $\mathsf{Parents}(X_i) = \{X_{i-1}\}$ for $i = 2, \ldots n$.

$$X_1 \to X_2 \to X_3 \to \cdots \to X_n \tag{1}$$

(*a*) Derive an expression for the probability $\mathsf{Pr}(X_1 = x_1 | X_n = \mathsf{True})$. You may neglect the normalising factor. [2 marks]

(*b*) Derive the time complexity for computing $\mathsf{Pr}(X_1 = x_1 | X_n = \mathsf{True})$ using variable elimination. Contrast against exact inference without memoization. [6 marks]

(*c*) State the **E** and **M** steps in the expectation-maximisation algorithm for parameter estimation in a problem involving latent variables Z and observed data X. [4 marks]

(*d*) Henceforth, let θ denote the parameters to be estimated and X denote data we observe. Justify why the EM algorithm locally maximises $p(X|\theta)$ with respect to θ. [3 marks]

(*e*) Consider the Bayesian network depicted in the figure below. $\{X_i\}_{i=1}^n$ denote observed variables while the $\{Z_i\}_{i=1}^n$ are unobserved. We place the following distributions on the random variables:

- Z_1 and $Z_j | Z_{j-1} = l$ are Bernoulli-distributed, where $l \in \{0, 1\}$.
- $X_j | Z_j = l$ is a univariate normal distribution, where $l \in \{0, 1\}$.

Collectively denote the parameters of the above distributions as θ. Give the factorisation of the joint probability indicated by the Bayesian network structure in terms of the given distributions. [5 marks]