

Assignment 2

Statistical Methods in Applied Computer Science

DD2447

Jim Holmström 890503-7571
jimho@kth.se

November 17, 2012

Exercise A1 Compute $p(D|T \in \text{Polytree})$ with Bernoulli CPD's
Show how to compute $p(D|T)$ where T is a GDM which is a polytree and $D = \{x_1, \dots, x_N\}$ (and x_i is an assignment of values to all variables of T). Assume that all variables are binary and all CPD's Bernoulli.

Solution.

Exercise A2 Marginalize over non-observed variables
Assume instead that each x_i is an assignment to a subset of the variables say O . Show how to marginalize over $V \setminus O$ (i.e., the non-observed variables).

Solution.

Exercise 11.3 EM for the mixtures of Bernoullis

- Show that the M step for ML estimation of a mixture of Bernoullis is given by

$$\mu_{kj} = \frac{\sum_i r_{ik} x_{ij}}{\sum_i r_{ik}} \quad (1)$$

- Show that the M step for MAP estimation of a mixture of Bernoullis with a $\beta(\alpha, \beta)$ prior is given by

$$\mu_{kj} = \frac{(\sum_i r_{ik} x_{ij}) + \alpha - 1}{(\sum_i r_{ik}) + \alpha + \beta - 1} \quad (2)$$

Solution.