Assignment 2
Statistical Methods in Applied Computer Science
DD2447

Jim Holmström 890503-7571
jimho@kth.se

November 23, 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 DGM which is a polytree and $D = \{x_1, ..., x_N\}$ (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

$$\mu_{kj} = \frac{\sum_i r_{ik} x_{ij}}{\sum_i r_{ik}} \tag{1}$$

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

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

Solution. The M step is to optimize the auxiliary function Q with respect to π, θ'. Q is the expected posterior log-likelihood[1] with respect to the last parameter θ and the observed data D. [2] The expression for this is

$$Q(\theta', \theta) = \mathrm{E}\left[\log \mathcal{L}_{MAP}(\theta')|D, \theta\right] = \mathrm{E}\left[\log\left(\mathcal{L}(\theta')p(\theta')\right)|D, \theta\right] = \tag{3}$$

$$= \mathrm{E}\left[\ell(\theta') + \log p(\theta')|D, \theta\right] = \mathrm{E}\left[\ell(\theta')|D, \theta\right] + \log p(\theta') \tag{4}$$

and to derive this expression we introducing the latent variable z_i which corresponds to the hidden or missing variables which basically is the *r.v.* for how x_i belongs to the class k [3]. We start with MAP and then derive ML by setting a uniform priori.

$$Q(\theta', \theta) = \mathrm{E}\left[\sum_i \log p(\, x_i, z_i | \theta'\,)\right] + p(\theta') = \tag{5}$$

since E is linear and we can factor on the different classes and the factors become given with the parameters (π_k, θ'_k) of class k:

$$= \sum_i \left[\mathrm{E}\, \log\left(\prod_k (\pi_k p(\, x_i | \theta'_k\,))^{\mathbb{I}(z_i = k)}\right)\right] + \log \prod_k p(\theta'_k) = \tag{6}$$

then log the inner parts and let E operate on the expression

$$= \sum_i \sum_k \left[\mathrm{E}\,[\mathbb{I}(z_i = k)] \log\left[p(\, x_i | \theta'_k\,)\right]\right] + \sum_k \log p(\theta'_k) = \tag{7}$$

then we have that the expected $\mathrm{E}\,[\mathbb{I}(z_i = k)]$ will be $p(\, z_i = k | x_i, \theta\,) = r_{ik}$ which is the expected class-belonging of x_i given the previous parameters θ. The log product rule gives us

$$= \sum_i \sum_k r_{ik} \log \pi_k + \sum_i \sum_k \{r_{ik} \log p(\, x_i | \theta'_k\,)\} + \sum_k \log p(\theta'_k) \tag{8}$$

Now since we have that $\sum_{i,k} r_{ik} \log \pi_k \perp \sum_{i,k} \{r_{ik} \log p(\, x_i | \theta'_k\,)\} + \sum_k \log p(\theta'_k)$ we can optimize the parameters π_k and θ'_k separately and only the θ'_k-term is asked for in the exercise which we denote $\ell(\theta'_k)$.

The model parameters in this case is denoted by μ_k which is a vector and in index-notation μ_{kj} [4].

[1] In the book it's actually just log-likelihood, but this will work in the same way instead of using $Q(\theta', \theta) + \log p(\theta')$ without derivation, we will use posterior-Q as Q instead.

[2] It can be shown that Q is so that the new parameters is always better or as good as the last one, but exclude the proof for this since it's not needed by the exercise.

[3] Responsibility $r_{ik} \triangleq p(\, z_i = k | x_i, \theta\,)$

[4] Not using Einstein notation for tensor product.

We start with the MAP

$$\hat{\mu}'_k = \underset{\mu'_k}{argmax} \left\{ \sum_i \{ r_{ik} \log p(\,x_i | \mu'_k\,) \} + \log p(\mu'_k) \right\} \tag{9}$$

which we find by $\frac{\partial \ell(\mu'_k)}{\partial \mu'_k} = 0$, where $p(\,x_i|\mu'_k\,) = \prod_j p(\,x_{ij} \middle| \mu'_{kj}\,)$ is the multivariate Bernoulli[5] distribution for class k. The priori in the same way is $p(\mu'_k) = \prod_j p(\mu'_{kj})$. In our case for MAP this is $\beta(\alpha, \beta) = \frac{{\mu'_{kj}}^{\alpha-1}(1-\mu'_{kj})^{\beta-1}}{B(\alpha,\beta)}$.

$$\frac{\partial(\ell(\mu'_k))}{\partial \mu'_{kj}} = \frac{\partial \sum_i \left\{ r_{ik} \log \prod_{j'} p(\,x_{ij'} \middle| \mu'_{kj'}\,) \right\} + \log \prod_{j'} p(\mu'_{kj'})}{\partial \mu'_{kj}} = \tag{10}$$

$$= \frac{\partial \sum_{i,j'} \left\{ r_{ik} \log p(\,x_{ij'} \middle| \mu'_{kj'}\,) \right\} + \sum_{j'} \log p(\mu'_{kj'})}{\partial \mu'_{kj}} = \tag{11}$$

$$= \left\{ \frac{\partial \sum_{j \neq j'} (\cdot)}{\partial \mu'_{kj}} = 0 \right\} = \frac{\partial \sum_i \left\{ r_{ik} \log \left({\mu'_{kj}}^{x_{ij}} (1 - \mu'_{kj})^{(1-x_{ij})} \right) \right\} + \log \frac{{\mu'_{kj}}^{\alpha-1}(1-\mu'_{kj})^{\beta-1}}{B(\alpha,\beta)}}{\partial \mu'_{kj}} = \tag{12}$$

$$= \frac{\partial \sum_i r_{ik} \left(x_{ij} \log \mu'_{kj} + (1 - x_{ij}) \log(1 - \mu'_{kj}) \right)}{\partial \mu'_k} + \tag{13}$$

$$+ \frac{\partial (\alpha - 1) \log \mu'_{kj} + (\beta - 1) \log(1 - \mu'_{kj}) - \log B}{\partial \mu'_k} = \tag{14}$$

$$= \frac{(\sum_i r_{ik} x_{ij}) + \alpha - 1}{\mu'_{kj}} - \frac{(\sum_i r_{ik}(1 - x_{ij})) + \beta - 1}{1 - \mu'_{kj}} = \tag{15}$$

$$= \frac{(1 - \mu'_{kj}) \left(\sum_i r_{ik} x_{ij}\right) - \mu'_{kj} \left(\sum_i r_{ik}(1 - x_{ij})\right)}{\mu'_{kj}(1 - \mu'_{kj})} + \tag{16}$$

$$+ \frac{(1 - \mu'_{kj}) (\alpha - 1) - \mu'_{kj} (\beta - 1)}{\mu'_{kj}(1 - \mu'_{kj})} = \tag{17}$$

$$= \frac{(\sum_i r_{ik} x_{ij}) - \mu'_{kj} \left(\sum_i r_{ik}\right) + (\alpha - 1) - \mu'_{kj}(\alpha + \beta - 2)}{\mu'_{kj}(1 - \mu'_{kj})} \tag{18}$$

and now find the zero of this expression

$$\frac{(\sum_i r_{ik} x_{ij}) - \mu'_{kj} \left(\sum_i r_{ik}\right) + (\alpha - 1) - \mu'_{kj}(\alpha + \beta - 2)}{\mu'_{kj}(1 - \mu'_{kj})} = 0 \tag{19}$$

[5] Often called multinoulli.

$$\left(\sum r_{ik}x_{ij}\right) - \mu'_{kj}\left(\sum r_{ik}\right) + (\alpha - 1) - \mu'_{kj}(\alpha + \beta - 2) = 0 \tag{20}$$

$$\mu'_{kj}\left[\left(\sum r_{ik}\right) + \alpha + \beta - 2\right] = \left(\sum r_{ik}x_{ij}\right) + \alpha - 1 \tag{21}$$

$$\mu'_{kj} = \frac{\left(\sum_i r_{ik}x_{ij}\right) + \alpha - 1}{\left(\sum_i r_{ik}\right) + \alpha + \beta - 2} \tag{22}$$

which is the same as the equation (2).

By just looking at the expression for $\beta(\cdot)$ we see that

$$\beta(1,1) = U(0,1) \tag{23}$$

and knowing that ML is a MAP with a uniform priori we can just set $\alpha, \beta = 1$ in the derived expression to get the ML.

$$\mu'_{kj} = \frac{\sum_i r_{ik}x_{ij} + 1 - 1}{\sum_i r_{ik} + 1 + 1 - 2} = \frac{\sum_i r_{ik}x_{ij}}{\sum_i r_{ik}} \tag{24}$$

which is the same as the equation (1).