# Assignment 2
# Statistical Methods in Applied Computer Science
# DD2447


Jim Holmström 890503-7571

jimho@kth.se


November 21, 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 $D = \{x_1, ..., x_N\}$ (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

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

*Solution.* The M step is to optimize the auxiliary function $Q$ with respect to $\pi, \theta'$. $Q$ is the expected log-likelihood with respect to the former parameter $\theta$ and the observed data $D$. [1] The expression for this is

$$Q(\theta', \theta) = \mathrm{E}\left[\ell(\theta')|D, \theta\right] \tag{3}$$

and to derive the expression we will use when optimizing with ML/MAP we start with introducing the latent variable $z_i$ which corresponds to the hidden or missing variables which basically is the *r.v.* for how $x_i$ is in the class $k$ [2].

$$Q(\theta', \theta) = \mathrm{E}\left[\sum_i \log p(\, x_i, z_i \big| \theta' \,)\right] \tag{4}$$

since E is linear and we can factor on the different classes and the factors becomes given only the parameters $(\pi_k, \theta'_k)$ of class $k$:

$$= \sum_i \mathrm{E}\left[\log\left[\prod_k (\pi_k p(\, x_i \big| \theta'_k \,))^{\mathbb{I}(z_i = k)}\right]\right] \tag{5}$$

then log the inner parts and let E operate on the expression like this:

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

then we put the expected $\mathrm{E}\left[\mathbb{I}(z_i = k)\right]$ to be $p(\, z_i = k | x_i, \theta \,) = r_{ik}$ that is the expected class belonging of $x_i$ given the previous parameters $\theta$. Now put in the denotation $r_{ik}$ and use the log on product rule to finally get

$$= \sum_i \sum_k r_{ik} \log \pi_k + \sum_i \sum_k r_{ik} \log p(\, x_i \big| \theta'_k \,) \tag{7}$$

Now since we have that $\sum\limits_{i,k} r_{ik} \log \pi_k \perp \sum\limits_{i,k} r_{ik} \log p(\, x_i | \theta'_k \,)$ we can optimize the parameters $\pi_k$ and $\theta'_k$ separately and only the $\theta'_k$-term is asked for in the exercise which will be denoted by $\ell(\theta'_k)$.

The model parameters in this case is denoted by $\mu_k$ which is a vector and in index-notation $\mu_{kj}$ [3].

We start with the ML

$$\hat{\mu}'_k = \underset{\mu'_k}{argmax}\left\{\sum_i r_{ik} \log p(\, x_i \big| \mu'_k \,)\right\} \tag{8}$$

[1] It can be shown that $Q$ is so that the new parameters is always better or as good as the last one, but exclude this proof.

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

[3] Not using Einstein notation for Tensor product.

which we find by $\frac{\partial \ell(\mu'_k)}{\partial \mu'_k} = 0$, where $\prod_j p(x_{ij} \Big| \mu'_{kj})$ is the multivariate Bernoulli for class $k$

$$\frac{\partial(\ell(\mu'_k))}{\partial \mu'_{kj}} = \frac{\partial \sum_i r_{ik} \log \prod_{j'} p(x_{ij'} \Big| \mu'_{kj'})}{\partial \mu'_{kj}} = \frac{\partial \sum_{i,j'} r_{ik} \log p(x_{ij'} \Big| \mu'_{kj'})}{\partial \mu'_{kj}} = \left\{ \frac{\partial \sum_{j \neq j'} (\cdot)}{\partial \mu'_{kj}} = 0 \right\} = \tag{9}$$

$$= \frac{\partial \sum_i r_{ik} \log \left( {\mu'_{kj}}^{x_{ij}} (1 - \mu'_{kj})^{(1-x_{ij})} \right)}{\partial \mu'_{kj}} = \frac{\partial \sum_i r_{ik} x_{ij} \log \mu'_{kj} + r_{ik}(1 - x_{ij}) \log(1 - \mu'_{kj})}{\partial \mu'_k} = \tag{10}$$

$$= \sum_i r_{ik} \frac{x_{ij}}{\mu'_{kj}} - r_{ik} \frac{1 - x_{ij}}{1 - \mu'_{kj}} = \sum_i r_{ik} \frac{\mu'_{kj} - x_{ij}}{\mu'_{kj}(\mu'_{kj} - 1)} \tag{11}$$

now find the zero of this expression

$$\sum_i r_{ik} \frac{\mu'_{kj} - x_{ij}}{\mu'_{kj}(\mu'_{kj} - 1)} = 0 \tag{12}$$

$$\sum_i r_{ik} \mu'_{kj} - r_{ik} x_{ij} = 0 \tag{13}$$

$$\mu'_{kj} \sum_i r_{ik} = \sum_i r_{ik} x_{ij} \tag{14}$$

$$\mu'_{kj} = \frac{\sum_i r_{ik} x_{ij}}{\sum_i r_{ik}} \tag{15}$$

which is the same as the equation (1).