

# Assignment 2

## Statistical Methods in Applied Computer Science

### DD2447

Jim Holmström 890503-7571  
[jimho@kth.se](mailto: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  $D = \{x_1, \dots, x_N\}$  (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

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

*Solution.* The M step is to optimize the auxiliary function  $Q$  with respect to  $\pi, \theta'$ .  $Q$  is the expected posterior log-likelihood<sup>1</sup> with respect to the last parameter  $\theta$  and the observed data  $D$ .<sup>2</sup> The expression for this is

$$Q(\theta', \theta) = \mathbb{E} [\log \mathcal{L}_{MAP}(\theta') | D, \theta] = \mathbb{E} [\log (\mathcal{L}(\theta') p(\theta')) | D, \theta] = \quad (3)$$

$$= \mathbb{E} [\ell(\theta') + \log p(\theta') | D, \theta] = \mathbb{E} [\ell(\theta') | D, \theta] + \log p(\theta') \quad (4)$$

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  $k$ <sup>3</sup>. We start with MAP and then derive ML by setting a uniform priori.

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

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

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

then log the inner parts and let  $\mathbb{E}$  operate on the expression

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

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

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

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  $\pi_k$  and  $\theta'_k$  separately and only the  $\theta'_k$ -term is asked for in the exercise which we denote  $\ell(\theta'_k)$ .

The model parameters in this case is denoted by  $\mu_k$  which is a vector and in index-notation  $\mu_{kj}$ <sup>4</sup>.

<sup>1</sup>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.

<sup>2</sup>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.

<sup>3</sup>Responsibility  $r_{ik} \triangleq p(z_i = k | x_i, \theta)$

<sup>4</sup>Not using Einstein notation for tensor product.

We start with the MAP

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

which we find by  $\frac{\partial \ell(\mu'_k)}{\partial \mu'_k} = 0$ , where  $p(x_i | \mu'_k) = \prod_j p(x_{ij} | \mu'_{kj})$  is the multivariate Bernoulli<sup>5</sup> 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  $\beta(\alpha, \beta) = \frac{\mu'_{kj}{}^{\alpha-1} (1-\mu'_{kj})^{\beta-1}}{B(\alpha, \beta)}$ .

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

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

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

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

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

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

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

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

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

and now find the zero of this expression

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

---

<sup>5</sup>Often called multinoulli.

$$\sum_i r_{ik} \mu'_{kj} - r_{ik} x_{ij} = 0 \quad (20)$$

$$\mu'_{kj} \sum_i r_{ik} = \sum_i r_{ik} x_{ij} \quad (21)$$

$$\mu'_{kj} = \frac{\sum_i r_{ik} x_{ij}}{\sum_i r_{ik}} \quad (22)$$

which is the same as the equation (2).

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

$$\beta(1, 1) = U(0, 1) \quad (23)$$

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

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

which is the same as the equation (1).