consulted solution

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

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

i = datapoint index, k = component, j = dimension index of D-dim bit vectors

$$\ell(\mu) = \sum_i \sum_k r_{ik} \log \mathbb{P}(x_i | \theta_k) = \sum_i \sum_k r_{ik} \sum_j x_{ij} \log \mu_{kj}$$
$$+ (1-x_{ij}) \log(1-\mu_{kj})$$

Take deriv w.r.t. μ_{kj} =>

$$\frac{\partial \ell}{\partial \mu_{kj}} = \sum_i r_{ik} \left(\frac{x_{ij}}{\mu_{kj}} - \frac{1-x_{ij}}{1-\mu_{kj}} \right) = \sum_i r_{ik} \left(\frac{x_{ij} - \mu_{kj}}{\mu_{kj}(1-\mu_{kj})} \right) =$$

$$\frac{1}{\mu_{kj}(1-\mu_{kj})} \sum_i r_{ik} (x_{ij} - \mu_{kj}) = 0,$$ which gives the optimality condition $\sum_i r_{ik} x_{ij} = \mu_{kj} \sum_i$, which gives us the desired result.

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

$$\mu_{kj} = \frac{(\sum_i r_{ik} x_{ij}) + a - 1}{(\sum_i r_{ik}) + a + b - 2}$$

$$\ell(\mu) = \sum_i \sum_k r_{ik} \log \mathbb{P}(x_i | \mu_k) + \log \mathbb{P}(\mu_k) =$$

$$\sum_i \sum_k r_{ik} \left(\sum_j x_{ij} \log \mu_{kj} + (1-x_{ij}) \log(1-\mu_{kj}) \right) + (a-1) \cdot \log \mu_{kj}$$

$+ (b-1) \log(1-\mu_{kj})$, taking derivatives we obtain

$$\frac{\partial l}{\partial \mu} = \sum_i \left(\frac{r_{ik} x_{ij} + a - 1}{\mu_{kj}} - \frac{r_{ik}(1-x_{ij}) + b - 1}{1-\mu_{kj}} \right) =$$

$$\frac{1}{\mu_{kj}(1-\mu_{kj})} \sum_i r_{ik} x_{ij} - r_{ik} \mu_{kj} + a - 1 - \mu_{kj} a + \mu_{kj} - \mu_{kj} b + \mu_{kj}$$

$$= \frac{1}{\mu_{kj}(1-\mu_{kj})} \left(\sum_i r_{ik} x_{ij} - \left(\sum_i r_{ik} + a + b - 2 \right) \mu_{kj} + a - 1 \right) = 0$$

which gives the optimality condition $\sum_i r_{ik} x_{ij} + a - 1 = \left(\sum_i r_{ik} + a + b - 2 \right) \mu_{kj}$.