NAME: Tanmay Gupta **NET ID: tgupta6**
NAME: Peter Maginnis **NET ID: maginni1**

IE598: Inference in Graphical Models
Homework 3

EM without prior

We treat $\mathbf{A}$ as our observed variables, $\mathbf{t}$ as latent variables and $\mathbf{p}$ as parameters. Then the joint distribution will be given as:

$$\begin{aligned}\mathbb{P}(\mathbf{A},\mathbf{t}|\mathbf{p}) &= \mathbb{P}(\mathbf{A}|\mathbf{t},\mathbf{p})\mathbb{P}(\mathbf{t}|\mathbf{p})\\ &= \mathbb{P}(\mathbf{A}|\mathbf{t},\mathbf{p})\mathbb{P}(\mathbf{t}) && \text{(Since true labels are independent of worker abilities)}\\ &= \prod_{(i,j)\in E}\mathbb{P}(A_{ij}|t_i,p_j)\prod_{i}^{n}\mathbb{P}(t_i)\end{aligned}$$

where,

$$\begin{aligned}\mathbb{P}(A_{ij}|t_i,p_j) &= p_j\mathbb{I}(A_{ij}=t_i) + (1-p_j)\mathbb{I}(A_{ij}=-t_i)\\ \mathbb{P}(t_i) &= \frac{3}{4}\mathbb{I}(t_i=1) + \frac{1}{4}\mathbb{I}(t_i=-1)\end{aligned}$$

E-Step:

In the **E** step we find the distribution of the latent variables as a function of the observed variables and parameters estimated in the previous iteration.

$$\mathbb{P}(\mathbf{t}|\mathbf{A},\mathbf{p}^{\text{old}}) = \prod_{i=1}^{n}\mathbb{P}(t_i|\mathbf{A},\mathbf{p}^{\text{old}})$$

where,

$$\begin{aligned}\mathbb{P}(t_i|\mathbf{A},\mathbf{p}^{\text{old}}) &= \mathbb{P}(t_i|\mathbf{A}_{i\partial i},\mathbf{p}^{\text{old}}_{\partial i})\\ &= \frac{\mathbb{P}(\mathbf{A}_{i\partial i},\mathbf{t}_i|\mathbf{p}^{\text{old}}_{\partial i})}{\mathbb{P}(\mathbf{A}_{i\partial i}|\mathbf{p}^{\text{old}}_{\partial i})}\\ &= \frac{\mathbb{P}(t_i)\prod_{j\in\partial i}\mathbb{P}(A_{ij}|t_i,p_j^{\text{old}})}{\sum_{t_i}\mathbb{P}(t_i)\prod_{j\in\partial i}\mathbb{P}(A_{ij}|t_i,p_j^{\text{old}})}\\ &= \gamma_i(t_i)\end{aligned}$$

In the sequel, allowing for a slight abuse of notation, we need only refer to

$$\begin{aligned}\gamma_i := \gamma_i(1) &= \mathbb{P}(t_i=1|\mathbf{A},\mathbf{p}^{\text{old}})\\ &= \frac{\mathbb{P}(A_{i\partial i}|t_i=1,\mathbf{p})\mathbb{P}(t_i=1)}{\mathbb{P}(A_{i\partial i}|t_i=1,\mathbf{p})\mathbb{P}(t_i=1)+\mathbb{P}(A_{i\partial i}|t_i=-1,\mathbf{p})\mathbb{P}(t_i=-1)}\\ &= \frac{\frac{3}{4}\prod_{j\in\partial i}p_j^{\mathbb{I}(A_{ij}=1)}(1-p_j)^{\mathbb{I}(A_{ij}=-1)}}{\frac{3}{4}\prod_{j\in\partial i}p_j^{\mathbb{I}(A_{ij}=1)}(1-p_j)^{\mathbb{I}(A_{ij}=-1)}+\frac{1}{4}\prod_{j\in\partial i}p_j^{\mathbb{I}(A_{ij}=-1)}(1-p_j)^{\mathbb{I}(A_{ij}=1)}}\\ &= \frac{\frac{3}{4}\prod_{j\in\partial i}\left[\frac{1}{2}+\frac{2p_j-1}{2}A_{ij}\right]}{\frac{3}{4}\prod_{j\in\partial i}\left[\frac{1}{2}+\frac{2p_j-1}{2}A_{ij}\right]+\frac{1}{4}\prod_{j\in\partial i}\left[\frac{1}{2}-\frac{2p_j-1}{2}A_{ij}\right]}\end{aligned}$$

M-Step:

In this we approximate the log likelihood of the observed data using the expected log likelihood of the observed and latent variables together where the expectation is with respect to the distribution of the latent variables computed in the **E** step

$$
\begin{aligned}
Q(\mathbf{p}|\mathbf{p}^{\text{old}}) &= \mathbb{E}_\gamma \left[\log \mathbb{P}(\mathbf{A}, \mathbf{t}|\mathbf{p})\right] \\
&= \mathbb{E}_{\mathbf{t}|\mathbf{A},\mathbf{p}^{\text{old}}} \left[\log \mathbb{P}(\mathbf{A}, \mathbf{t}|\mathbf{p})\right] \\
&= \sum_{\mathbf{t}\in\{-1,1\}^n} \mathbb{P}(\mathbf{t}|\mathbf{A}, \mathbf{p}^{\text{old}}) \log \mathbb{P}(\mathbf{A}, \mathbf{t}|\mathbf{p}) \\
&= \sum_{\mathbf{t}\in\{-1,1\}^n} \left[\left(\prod_{i=1}^n \mathbb{P}(t_i|\mathbf{A}, \mathbf{p}^{\text{old}})\right) \sum_{i=1}^n \left(\log \mathbb{P}(t_i|\mathbf{p}) + \sum_{j\in\partial i} \log \mathbb{P}(A_{ij}|t_i, p_j)\right)\right] \\
&= \sum_{i=1}^n \sum_{t_i} \gamma_i(t_i) \left[\log \mathbb{P}(t_i|\mathbf{p}) + \sum_{j\in\partial i} \log \mathbb{P}(A_{ij}|t_i, p_j)\right] \\
&= \sum_{i=1}^n \Bigg\{ \gamma_i \left[\log \mathbb{P}(t_i = 1) + \sum_{j\in\partial i} \log \mathbb{P}(A_{ij}|t_i = 1, p_j)\right] \\
&\qquad + (1-\gamma_i) \left[\log \mathbb{P}(t_i = -1) + \sum_{j\in\partial i} \log \mathbb{P}(A_{ij}|t_i = -1, p_j)\right] \Bigg\}
\end{aligned}
$$

Now to maximize $Q(\mathbf{p}|\mathbf{p}^{\text{old}})$ with respect to $\mathbf{p}$ we set the derivatives to zero.

$$
\begin{aligned}
&\frac{\partial Q}{\partial p_j} = 0 \\
\implies &\sum_{i\in\partial j} \left[\gamma_i \left(\frac{\mathbb{I}(A_{ij} = 1)}{p_j} - \frac{\mathbb{I}(A_{ij} = -1)}{1 - p_j}\right) + (1-\gamma_i) \left(\frac{\mathbb{I}(A_{ij} = -1)}{p_j} - \frac{\mathbb{I}(A_{ij} = 1)}{1 - p_j}\right)\right] = 0 \\
\implies &\sum_{i\in\partial j} \left[(\gamma_i - p_j)\, \mathbb{I}(A_{ij} = 1) + (1 - \gamma_i - p_j)\, \mathbb{I}(A_{ij} = -1)\right] = 0 \\
\implies &p_j = \frac{1}{|\partial j|} \sum_{i\in\partial j} \left[\gamma_i \mathbb{I}(A_{ij} = 1) + (1-\gamma_i)\, \mathbb{I}(A_{ij} = -1)\right]
\end{aligned}
$$

EM with Beta Prior

Instead, we seek to maximize the objective

$$
\begin{aligned}
Q(\mathbf{p}|\mathbf{p}^{\text{old}}) &= \mathbb{E}_\gamma \left[\ln \mathbb{P}(\mathbf{A}|\mathbf{p}) + \ln \mathbb{P}(\mathbf{p})\right] \\
&= \sum_{i=1}^n \left[\gamma_i \ln \frac{3}{4} a_i + (1-\gamma_i) \ln \frac{1}{4} b_i\right] + \sum_{j=1}^m \ln f\left(\frac{p_j - 0.1}{0.9}\right),
\end{aligned}
$$

where

$$
\begin{aligned}
a_i &:= \mathbb{P}(A_{i\partial i}|t_i = 1, \mathbf{p}) = \prod_{j\in\partial i} p_j^{\mathbb{I}(A_{ij}=1)} (1-p_j)^{\mathbb{I}(A_{ij}=-1)} \\
b_i &:= \mathbb{P}(A_{i\partial i}|t_i = -1, \mathbf{p}) = \prod_{j\in\partial i} p_j^{\mathbb{I}(A_{ij}=-1)} (1-p_j)^{\mathbb{I}(A_{ij}=1)},
\end{aligned}
$$

and

$$f(z) = \frac{1}{B(\alpha,\beta)}(z)^{\alpha-1}(1-z)^{\beta-1}$$

We take the gradient

$$\begin{aligned}
\frac{\partial Q}{\partial p_j} &= \sum_{i\in\partial j}\left\{\gamma_i\left[\frac{\mathbb{I}(A_{ij}=1)}{p_j} - \frac{\mathbb{I}(A_{ij}=-1)}{1-p_j}\right] + (1-\gamma_i)\left[\frac{\mathbb{I}(A_{ij}=-1)}{p_j} - \frac{\mathbb{I}(A_{ij}=1)}{1-p_j}\right]\right\} + \frac{\alpha-1}{p_j-0.1} - \frac{\beta-1}{1-p_j} \\
&= \sum_{i\in\partial j}\left\{\frac{\gamma_i\mathbb{I}(A_{ij}=1) + (1-\gamma_i)(\mathbb{I}(A_{ij}=-1))}{p_j} - \frac{(1-\gamma_i)\mathbb{I}(A_{ij}=1) + \gamma_i(\mathbb{I}(A_{ij}=-1))}{1-p_j}\right\} + \frac{\alpha-1}{p_j-0.1} - \frac{\beta-1}{1-p_j} \\
&= \sum_{i\in\partial j}\left\{\frac{\gamma_i A_{ij} - A_{ij} + 1}{p_j} - \frac{A_{ij} - \gamma_i A_{ij} + 1}{1-p_j}\right\} + \frac{\alpha-1}{p_j-0.1} - \frac{\beta-1}{1-p_j}.
\end{aligned}$$

We then solve for the critical point

$$\frac{\partial Q}{\partial p_j} = 0$$

which implies

$$0 = (p_j - 0.1)\sum_{i\in\partial j}\{\gamma_i\mathbb{I}(A_{ij}=1) + (1-\gamma_i)\mathbb{I}(A_{ij}=-1) - p_j\} + (\alpha-1)p_j(1-p_j) - (\beta-1)p_j(p_j-0.1).$$

Define

$$\lambda_j := \frac{1}{|\partial j|}\sum_{i\in\partial j}\left[\gamma_i\mathbb{I}(A_{ij}=1) + (1-\gamma_i)\mathbb{I}(A_{ij}=-1)\right].$$

Then, we may more compactly write the quadratic equation

$$0 = (p_j - 0.1)|\partial j|[\lambda_j - p_j] + (\alpha-1)p_j(1-p_j) - (\beta-1)p_j(p_j-0.1).$$

Alternatively, suppose we approximated p_j as $\mathrm{Beta}(\alpha,\beta)$, then the above becomes

$$\begin{aligned}
0 &= |\partial j|(\lambda_j - p_j) + (\alpha-1)(1-p_j) - (\beta-1)p_j \\
&\implies \lambda_j|\partial j| + \alpha - 1 = (|\partial j| + \alpha + \beta - 2)p_j \\
&\implies p_j = \frac{\sum_{i\in\partial j}\left[\gamma_i\mathbb{I}(A_{ij}=1) + (1-\gamma_i)\mathbb{I}(A_{ij}=-1)\right] + \alpha - 1}{|\partial j| + \alpha + \beta - 2}
\end{aligned}$$