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

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:

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

where,

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

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.

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

where,

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

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

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

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

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