

## 0.1 The Model

Let  $D_{nj}$  be the data corresponding to  $n$ -th sample and  $j$ -th feature, where  $n$  runs from 1 to  $N$  and  $j$  runs from 1 to  $G$ . We assume as a first pass that the data is Gaussian in its distribution.

Let us consider a graph model with  $K$  nodes (factors) and  $E$  edges (in our model right now,  $E = K(K - 1)/2$ ).

Let us define latent variables  $Z$  and  $\Lambda$  such that  $Z_n$  is a  $E \times 1$  vector which is one-hot, where  $E$  is the total number of edges in the graph. Again  $\Lambda_n$  is a  $Q \times 1$  one-hot vector, where  $Q$  is the cardinality of the set of values that  $q$  can take.

$$Pr(Z|\pi) = \prod_{n=1}^N \prod_{k_1 < k_2} \pi_{k_1, k_2}^{z_{n, k_1, k_2}}$$

We assume that  $q$  can take a finite set of values between 0 and 1, say  $1/100, 2/100, \dots, 90/100, 1$ . Also

$$Pr(\Lambda|\delta) = \prod_{n=1}^N \prod_{q=1}^Q \delta_q^{\lambda_{nq}}$$

here  $Q$  is 100, if there are 100 values that  $q$  can take.

We assume the model

$$E[D_{nj}|Z_{n, k_1, k_2} = 1, \lambda_{n, q} = 1, F] = qF_{k_1, j} + (1 - q)F_{k_2, j} \quad (1)$$

The prior we define over  $\Lambda$  is equivalent to writing

$$Pr[\Lambda_{n, q} = 1] = \pi_q$$

Then we can write

$$Pr[D_n|Z_{n, k_1, k_2} = 1, F, s_{j=1, 2, \dots, J}^2] = \sum_q \pi_q Pr[D_n|Z_{n, k_1, k_2} = 1, \Lambda_{n, q} = 1, F, s_{j=1, 2, \dots, J}^2]$$

where  $s_j^2$  is the variance of the  $j$ th feature.

We also get from the prior on  $Z_{n, k_1, k_2}$ ,

$$Pr[Z_{n, k_1, k_2} = 1] = \pi_{k_1, k_2} \quad k_1 < k_2$$

Then we can write

$$Pr[D_n|\pi, F] = \sum_{k_1 < k_2} \pi_{k_1, k_2} Pr[D_n|Z_{n, k_1, k_2} = 1, F, s_{j=1, 2, \dots, J}^2]$$

We define the joint prior over the edges and the fraction of the edge represented as

$$\pi_{k_1, k_2, q} = \pi_{k_1, k_2} \pi_q \quad k_1 < k_2$$

The overall likelihood

$$L(\pi, F) = \prod_{n=1}^N Pr [D_n | \pi, F, s_{j=1,2,\dots,J}^2]$$

or we can write it as

$$L(\pi, F) = \prod_{n=1}^N \sum_{k_1 < k_2} \sum_q \left[ \pi_{k_1, k_2, q} \times \prod_{j=1}^G N(D_{nj}; qF_{k_1, j} + (1-q)F_{k_2, j}, s_j^2) \right]$$

And the log likelihood

$$\ln L(\pi, F) = \sum_{n=1}^N \ln \left( \sum_{k_1 < k_2} \sum_q \left[ \pi_{k_1, k_2, q} \times \prod_{j=1}^G N(D_{nj}; qF_{k_1, j} + (1-q)F_{k_2, j}, s_j^2) \right] \right) \quad (2)$$

This is the quantity we want to maximize.

## 0.2 EM updates - E step

We assume that  $q$  can take a finite set of values between 0 and 1, say  $1/100, 2/100, \dots, 90/100, 1$ .

Suppose we have run upto  $m$  iterations. For the  $(m+1)$ th iteration, we have

$$\begin{aligned} \delta_{n, k_1, k_2, q}^{(m+1)} &= Pr [Z_{n, k_1, k_2} = 1, \Lambda_{n, q} = 1 | \pi^{(m)}, F^{(m)}, s_{j=1,2,\dots,J}^{(m)}, D_n] \\ &\propto Pr [Z_{n, k_1, k_2} = 1] Pr [\Lambda_{n, q} = 1] Pr [D_n | \pi^{(m)}, F^{(m)}, s_{j=1,2,\dots,J}^{(m)}, Z_{n, k_1, k_2} = 1, \Lambda_{n, q} = 1] \\ &\propto \pi_{k_1, k_2, q}^{(m)} \prod_j N(D_{nj} | qF_{k_1, j}^{(m)} + (1-q)F_{k_2, j}^{(m)}, s_j^{(m)2}) \end{aligned}$$

where  $s_j^{(m)2}$  is the residual variance of feature  $j$ .

We normalize  $\delta$  so that

$$\sum_{k_1 < k_2} \sum_q \delta_{n, k_1, k_2, q}^{(m+1)} = 1 \quad \forall n$$

We define

$$\pi_{k_1, k_2, q}^{(m+1)} = \frac{1}{N} \sum_{n=1}^N \delta_{n, k_1, k_2, q}^{(m+1)}$$

We have therefore updated  $\pi_{k_1, k_2, q}^{(m)}$  to  $\pi_{k_1, k_2, q}^{(m+1)}$ .

### 0.3 Variational EM - Model 1

In this set up, we assume prior distributions of  $\pi$  and  $\delta$  as follows

$$Pr(\pi|\alpha_0) = C(\alpha_0) \prod_{k_1 < k_2} \pi_{k_1, k_2}^{\alpha_0 - 1}$$

Similarly the prior distribution for  $\delta$  is

$$Pr(\delta|\beta_0) = C(\beta_0) \prod_{q=1}^Q \delta_q^{\beta_0 - 1}$$

The likelihood above can be written as

$$p(D|Z, \Lambda, F, s_{j=1,2,\dots,J}) = \prod_{n=1}^N \prod_{k_1 < k_2} \prod_{q=1}^Q \prod_{j=1}^G [N(D_{nj}|qF_{k_1,g} + (1-q)F_{k_2,g}, s_g^2)]^{\Lambda_{nq} Z_{n,k_1,k_2}}$$

The joint probability distribution is given by

$$p(D, Z, \Lambda, \pi, \delta|F, s_{j=1,2,\dots,J}, \alpha_0, \beta_0) = p(\pi|\alpha_0)p(\delta|\beta_0)p(\Lambda|\delta)p(Z|\pi)p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})$$

We assume the following mean field variational distribution. In the first model, we assume the two latent variables  $Z$  and  $\Lambda$  are independent.

$$q(Z, \Lambda, \pi, \delta) = q(Z)q(\Lambda)q(\pi)q(\delta)$$

The variational distribution for  $Z$

$$\begin{aligned} \ln q^*(Z) &= E_{\pi, \delta, \Lambda} [\ln p(\pi|\alpha_0) + \ln p(\delta|\beta_0) + \ln p(\Lambda|\delta) + \ln p(Z|\pi) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] \\ &= E_{\pi, \delta, \Lambda} [\ln p(Z|\pi) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] + \text{constant} \\ &= \sum_{n=1}^N \sum_{k_1 < k_2} z_{n,k_1,k_2} E_{\pi} [\ln(\pi_{k_1,k_2})] + \sum_{n=1}^N \sum_{k_1 < k_2} z_{n,k_1,k_2} \sum_q E_{\Lambda}(\lambda_{nq}) \left[ - \sum_{j=1}^G \ln(s_j) \right. \\ &\quad \left. - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^G \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \end{aligned}$$

$$\begin{aligned} \ln q^*(\Lambda) &= E_{\pi, \delta, Z} [\ln p(\pi|\alpha_0) + \ln p(\delta|\beta_0) + \ln p(\Lambda|\delta) + \ln p(Z|\pi) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] \\ &= E_{\pi, \delta, Z} [\ln p(\Lambda|\delta) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] + \text{constant} \\ &= \sum_{n=1}^N \sum_{q=1}^Q \lambda_{n,q} E_{\delta} [\ln(\delta_q)] + \sum_{n=1}^N \sum_q \lambda_{nq} \sum_{k_1 < k_2} E_Z(z_{n,k_1,k_2}) \left[ - \sum_{j=1}^G \ln(s_j) \right. \\ &\quad \left. - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^G \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \end{aligned}$$

So we get

$$q^*(Z) \propto \prod_{n=1}^N \prod_{k_1 < k_2} \rho_{n,k_1,k_2}^{Z_{n,k_1,k_2}}$$

where we define

$$\rho_{n,k_1,k_2} \propto \exp \left( E_\pi [\ln(\pi_{k_1,k_2})] + \sum_q E_\Lambda(\lambda_{nq}) \left[ - \sum_{j=1}^G \ln(s_j) - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^G \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \right)$$

$$\rho_{n,k_1,k_2} \propto \exp \left( E_\pi [\ln(\pi_{k_1,k_2})] + \sum_q \nu_{nq} \left[ - \sum_{j=1}^G \ln(s_j) - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^G \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \right)$$

$$\rho_{n,k_1,k_2} \propto \exp \left( \psi_{a_{k_1,k_2}} - \psi \left( \sum_{k_1 < k_2} a_{k_1,k_2} \right) + \left[ - \sum_{j=1}^G \ln(s_j) - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^G \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \right)$$

$\rho_{n,k_1,k_2}$  is normalized to sum to 1 for each  $n$  over  $k_1$  and  $k_2$ .

We also get

$$q^*(\Lambda) \propto \prod_{n=1}^N \prod_{q=1}^Q \nu_{nq}^{\Lambda_{nq}}$$

where

$$\nu_{nq} \propto \exp \left( E_\delta [\ln(\delta_q)] + \sum_{k_1 < k_2} E_Z(z_{n,k_1,k_2}) \left[ - \sum_{j=1}^G \ln(s_j) - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^G \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \right)$$

$$\nu_{nq} \propto \exp \left( E_\delta [\ln(\delta_q)] + \sum_{k_1 < k_2} \rho_{n,k_1,k_2} \left[ - \sum_{j=1}^G \ln(s_j) - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^G \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \right)$$

$$\nu_{nq} \propto \exp \left( \psi(b_q) - \psi \left( \sum_{q=1}^Q b_q \right) + \left[ - \sum_{j=1}^G \ln(s_j) - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^G \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \right)$$

$\nu_{nq}$  are normalized to sum to 1.

We can also derive variational distributions similarly for  $\pi$  and  $\delta$ .

$$\begin{aligned}
\ln q^*(\pi) &= E_{\Lambda, Z, \delta} [\ln p(\pi|\alpha_0) + \ln p(\delta|\beta_0) + \ln p(\Lambda|\delta) + \ln p(Z|\pi) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] \\
&= E_Z [\ln p(Z|\pi)] + \ln p(\pi|\alpha_0) + \text{constant} \\
&= \sum_{n=1}^N \sum_{k1 < k2} E(z_{n,k1,k2}) \ln \pi_{k1,k2} + (\alpha_0 - 1) \sum_{k1 < k2} \ln \pi_{k1,k2} \\
&= \sum_{k1 < k2} \left[ \sum_{n=1}^N \rho_{n,k1,k2} + (\alpha_0 - 1) \right] \ln \pi_k
\end{aligned}$$

We define

$$\begin{aligned}
a_{k1,k2} &= \alpha_0 + \sum_{n=1}^N \rho_{n,k1,k2} \\
q^*(\pi) &= \text{Dir}(\pi|a)
\end{aligned}$$

$$\begin{aligned}
\ln q^*(\delta) &= E_{\Lambda, Z, \pi} [\ln p(\pi|\alpha_0) + \ln p(\delta|\beta_0) + \ln p(\Lambda|\delta) + \ln p(Z|\pi) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] \\
&= E_{\Lambda} [\ln p(\Lambda|\delta)] + \ln p(\delta|\beta_0) + \text{constant} \\
&= \sum_{n=1}^N \sum_{q=1}^Q E(\lambda_{n,q}) \ln \delta_q + (\beta_0 - 1) \sum_{q=1}^Q \ln \delta_q \\
&= \sum_{q=1}^Q \left[ \sum_{n=1}^N \nu_{n,q} + (\beta_0 - 1) \right] \ln \delta_q
\end{aligned}$$

We define

$$\begin{aligned}
b_q &= \beta_0 + \sum_{n=1}^N \nu_{n,q} \\
q^*(\delta) &= \text{Dir}(\delta|b)
\end{aligned}$$

We alternate between the Variational E and M steps,  $E$  steps being the ones where we compute the responsibilities  $\rho_{n,k1,k2}$  and  $\nu_{n,q}$  and the M step is where we update the variational distribution of the parameters  $\pi$  and  $\delta$ .

We can start with  $a = \alpha_0$  and  $b = \beta_0$ . We can then estimate  $\rho_{n,k1,k2}$  and also  $\nu_{nq}$  and then then product of these two terms to get new responsibility

$$\delta_{n,k1,k2,q} = \rho_{n,k1,k2} \nu_{nq}$$

and we use this  $\delta_{n,k1,k2,q}$  as the responsibility for the M-step of the original EM updates.

## 0.4 Variational EM - Model 2

In model 2, we do not assume independence of the latent variables  $Z$  and  $\Lambda$  and instead estimate their joint variational distribution.

$$q(Z, \Lambda, \pi, \delta) = q(Z, \Lambda)q(\pi)q(\delta)$$

$$\begin{aligned} \ln q^*(Z, \Lambda) &= E_{\pi, \delta} [\ln p(\pi|\alpha_0) + \ln p(\delta|\beta_0) + \ln p(\Lambda|\delta) + \ln p(Z|\pi) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] \\ &= E_{\pi, \delta, \Lambda} [\ln p(Z|\pi) + \ln p(\Lambda|\delta) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] + \text{constant} \\ &= \sum_{n=1}^N \sum_{k_1 < k_2} \sum_{q=1}^Q z_{n,k_1,k_2} \lambda_{nq} E_{\pi} [\ln(\pi_{k_1,k_2})] + \sum_{n=1}^N \sum_{q=1}^Q \sum_{k_1 < k_2} \lambda_{n,q} z_{n,k_1,k_2} E_{\delta} [\ln(\delta_q)] \\ &\quad + \sum_{n=1}^N \sum_{k_1 < k_2} \sum_q z_{n,k_1,k_2} \lambda_{nq} \left[ - \sum_{j=1}^G \ln(s_j) - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^G \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \end{aligned} \quad (3)$$

From here one can get

$$q^*(Z, \Lambda) \propto \prod_{n=1}^N \prod_{k_1 < k_2} \prod_{q=1}^Q \delta_{n,k_1,k_2,q}^{Z_{n,k_1,k_2} \Lambda_{n,q}}$$

then

$$\begin{aligned} \delta_{n,k_1,k_2,q} &\propto \exp(E_{\pi} [\ln(\pi_{k_1,k_2})] + E_{\delta} [\ln(\delta_q)] \\ &\quad + \left[ - \sum_{j=1}^G \ln(s_j) - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^G \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right]) \end{aligned} \quad (4)$$

$$\begin{aligned} \delta_{n,k_1,k_2,q} &\propto \exp \left( \psi_{a_{k_1,k_2}} - \psi \left( \sum_{k_1 < k_2} a_{k_1,k_2} \right) + \psi(b_q) - \psi \left( \sum_{q=1}^Q b_q \right) \right. \\ &\quad \left. \left[ - \sum_{j=1}^G \ln(s_j) - \frac{G}{2} \ln(2\pi) - \sum_{j=1}^G \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2s_j^2} \right] \right) \end{aligned} \quad (5)$$

The updates for  $\pi$  and  $\delta$  are same as before. Here also we have the same way of initializing  $\pi$  and  $\delta$  first, then use  $a_{k_1,k_2} = \alpha_0$  and  $b_q = \beta_0$  to begin with and estimate  $\delta_{n,k_1,k_2,q}$ . Then use the  $\delta_{n,k_1,k_2,q}$  to update  $a_{k_1,k_2}$  and  $b_q$  and proceed in this way. In this case, we do not assume independence of the  $\Lambda$  and  $Z$  variational distributions, so this model is more generalized.

## 0.5 M step EM

We define the parameter

$$\theta := (\pi_{k_1, k_2, q}, F, s_{j=1,2,\dots,J})$$

We define the complete loglikelihood

$$\log L_c(\theta; D, Z, \lambda) = \log \pi_{k_1, k_2, q} + \log L(D|Z, \lambda, q, F)$$

We take the expectation of this quantity with respect to  $[Z, \lambda|D, \theta^{(m)}]$ .

$$Q(\theta|\theta^{(m)}) \propto - \sum_{n=1}^N \sum_{k_1 < k_2} \sum_q \delta_{n, k_1, k_2, q}^{(m+1)} \sum_j \left[ \log s_j^{(m+1)} + \frac{(D_{nj} - qF_{k_1, j} - (1-q)F_{k_2, j})^2}{2s_j^{(m+1)^2}} \right] \quad (6)$$

We try to maximize this quantity with respect to  $F$ , So, we can take derivative with respect to  $F$  and try to solve the resulting normal equation.

This equation, conditional on  $[Z, \lambda|D, \theta^{(m)}]$ , can be written as

$$D_{N \times J} = L_{N \times K} F_{K \times J} + E_{N \times J} \quad (7)$$

where

$$e_{nj} \sim N(0, s_j^2)$$

We define

$$D'_{nj} := \frac{D_{nj}}{s_j}$$

If we consider finding the factors on a feature-by-feature basis, we do not need to worry about  $s_j$ .

$$L_{nk} = \begin{cases} q \text{ or } (1-q) & \lambda_n = q \\ 0 & \text{o.w.} \end{cases}$$

We have

$$\begin{aligned} E_{Z, \lambda|D, \theta^{(m)}} [L_{nk}] &= \sum_q \sum_{k_2 > k} q \delta_{n, k, k_2, q}^{(m+1)} + \sum_q \sum_{k_1 < k} (1-q) \delta_{n, k_1, k, q}^{(m+1)} \\ E_{Z, \lambda|D, \theta^{(m)}} [L_{nk}^2] &= \sum_q \sum_{k_2 > k} q^2 \delta_{n, k, k_2, q}^{(m+1)} + \sum_q \sum_{k_1 < k} (1-q)^2 \delta_{n, k_1, k, q}^{(m+1)} \end{aligned}$$

Also for any  $k \neq l$ ,

$$E_{Z,\lambda|D,\theta^{(m)}} [L_{nk}L_{nl}] = \sum_q q(1-q)\delta_{n,k,l,q}^{(m+1)}$$

We use these to solve for the equation

$$[E_{Z,\lambda|D,\theta^{(m)}} (L^T L)] F \approx [E_{Z,\lambda|D,\theta^{(m)}} (L)]^T D$$

The solution therefore is

$$F \approx [E_{Z,\lambda|D,\theta^{(m)}} (L^T L)]^{-1} [E_{Z,\lambda|D,\theta^{(m)}} (L)]^T D$$

For  $W = L^T L$

$$W_{kl} = \sum_n L_{kn}L_{nl}$$

$$E_{Z,\lambda|D,\theta^{(m)}} (W_{kl}) = \sum_n E_{Z,\lambda|D,\theta^{(m)}} (L_{kn}L_{nl})$$

We use the definition of  $E_{Z,\lambda|D,\theta^{(m)}} [L_{nk}^2]$  and  $E_{Z,\lambda|D,\theta^{(m)}} [L_{nk}L_{nl}]$  from above to solve  $F$ .

In the same way as we computed  $F$  by solving for the normal equation obtained from taking derivative of the function  $Q(\theta|\theta^{(m)})$ , we take derivative of the latter with respect to  $s_j^2$  to obtain EM updates of the residual variance terms. Taking the derivative, we obtain the estimate as

$$\widehat{s_j^{(m+1)}}^2 = \frac{1}{N} \sum_{n=1}^N \sum_{k_1 < k_2} \sum_q \delta_{n,k_1,k_2,q}^{(m+1)} (D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2 \quad (8)$$

where the  $F$  are the estimated values of the factors from the previous step.

We then continue this procedure described above for multiple iterations.