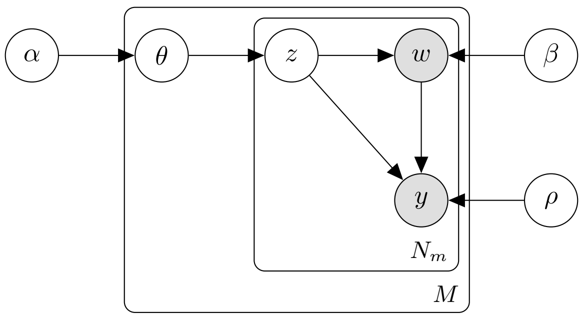
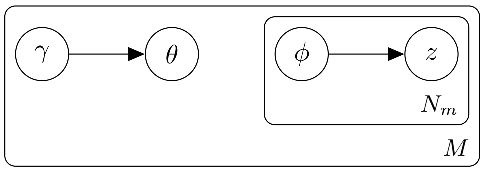
# polarLDA

Blue indicates new or additional terms to the original LDA.




(a) The model.




(b) Variational distribution.

## 1 Generative Process

For patient $m$,

1. draw a subtype mixture $\theta \sim \mathrm{Dir}(\alpha)$;
2. for each of his/her $N_m$ voxel involvements, independently
   (a) draw a subtype $z_n \sim \mathrm{Mult}(\theta)$;
   (b) draw a voxel $w_n \sim p\left(w_n \mid z_n, \beta\right)$;
   (c) draw an involvement type $y_n \sim p\left(y_n \mid w_n, z_n, \rho\right)$.

## 2 Constructing the Lower Bound

The variational distribution used to approximate the true posterior is factorizable as

$$q\left(\theta, z \mid \gamma, \phi\right) = q\left(\theta \mid \gamma\right) \prod_{n=1}^{N} q\left(z_n \mid \phi_n\right).$$

The lower bound $\mathcal{L}(\gamma, \phi \mid \alpha, \beta, \rho)$ of the single-brain log-likelihood $\log p(w \mid \alpha, \beta, \rho)$ is computed using Jensen's inequality as follows

$$\begin{aligned}
\log p(w, y \mid \alpha, \beta, \rho) &= \log \int \sum_z p(\theta, z, w, y \mid \alpha, \beta, \rho)\,\mathrm{d}\theta \\
&= \log \int \sum_z \frac{p(\theta, z, w, y \mid \alpha, \beta, \rho)\, q(\theta, z \mid \gamma, \phi)}{q(\theta, z \mid \gamma, \phi)} \mathrm{d}\theta \\
&= \log \int \sum_z q(\theta, z \mid \gamma, \phi) \frac{p(\theta, z, w, y \mid \alpha, \beta, \rho)}{q(\theta, z \mid \gamma, \phi)} \mathrm{d}\theta \\
&= \log \mathrm{E}_q \left\{ \frac{p(\theta, z, w, y \mid \alpha, \beta, \rho)}{q(\theta, z \mid \gamma, \phi)} \right\} \\
&\geq \mathrm{E}_q \{\log p(\theta, z, w, y \mid \alpha, \beta, \rho)\} - \mathrm{E}_q \{\log q(\theta, z \mid \gamma, \phi)\} \\
&\triangleq \mathcal{L}(\gamma, \phi \mid \alpha, \beta, \rho).
\end{aligned} \tag{1}$$

The difference between the log-likelihood and its lower bound can be proven to be in fact the KL divergence between the variational distribution and the true posterior.

$$\begin{aligned}
&\log p(w, y \mid \alpha, \beta, \rho) - \mathcal{L}(\gamma, \phi \mid \alpha, \beta, \rho) \\
&= \mathrm{E}_q \{\log p(w, y \mid \alpha, \beta, \rho)\} - \mathrm{E}_q \{\log p(\theta, z, w, y \mid \alpha, \beta, \rho)\} + \mathrm{E}_q \{\log q(\theta, z \mid \gamma, \phi)\} \\
&= \mathrm{E}_q \left\{ \log \frac{p(w, y \mid \alpha, \beta, \rho)\, q(\theta, z \mid \gamma, \phi)}{p(\theta, z, w, y \mid \alpha, \beta, \rho)} \right\} \\
&= \mathrm{E}_q \left\{ \log \frac{q(\theta, z \mid \gamma, \phi)}{p(\theta, z \mid w, y, \alpha, \beta, \rho)} \right\} \\
&= \mathrm{D}_{\mathrm{KL}}(q(\theta, z \mid \gamma, \phi) \parallel p(\theta, z \mid w, y, \alpha, \beta, \rho)).
\end{aligned}$$

Therefore, maximizing the lower bound is equivalent to minimizing this KL divergence. That is, the variational distribution automatically approaches to the real posterior as we maximize the lower bound.

## 3 Expanding the Lower Bound

To maximize the lower bound, we first need to spell out the lower bound $\mathcal{L}(\gamma, \phi \mid \alpha, \beta, \rho)$ in terms of the model parameters $(\alpha, \beta, \rho)$ and the variational parameters $(\gamma, \phi)$. Continuing from (1), we have

$$\begin{aligned}
\mathcal{L}(\gamma, \phi \mid \alpha, \beta) &= \mathrm{E}_q \{\log p(\theta, z, w, y \mid \alpha, \beta, \rho)\} - \mathrm{E}_q \{\log q(\theta, z \mid \gamma, \phi)\} \\
&= \mathrm{E}_q \left\{ \log \frac{p(\theta, z, w, y \mid \alpha, \beta, \rho)}{q(\theta, z \mid \gamma, \phi)} \right\} \\
&= \mathrm{E}_q \left\{ \log \frac{p(\theta \mid \alpha)\, p(z \mid \theta)\, p(w \mid z, \beta)\, p(y \mid z, w, \rho)}{q(\theta \mid \gamma)\, q(z \mid \phi)} \right\} \\
&= \mathrm{E}_q \{\log p(\theta \mid \alpha)\} + \mathrm{E}_q \{\log p(z \mid \theta)\} + \mathrm{E}_q \{\log p(w \mid z, \beta)\} \\
&+ \mathrm{E}_q \{\log p(y \mid z, w, \rho)\} \\
&- \mathrm{E}_q \{\log q(\theta \mid \gamma)\} - \mathrm{E}_q \{\log q(z \mid \phi)\}.
\end{aligned} \tag{2}$$

We now further expand each of the five terms in (2).

**The first term** is

$$
\begin{aligned}
\mathrm{E}_q\left\{\log p\left(\theta \mid \alpha\right)\right\} &= \mathrm{E}_q\left\{\log \frac{\Gamma\left(\sum_{i=1}^{K}\alpha_i\right)}{\prod_{k=1}^{K}\Gamma\left(\alpha_k\right)}\prod_{k=1}^{K}\theta_k^{\alpha_k-1}\right\} \\
&= \mathrm{E}_q\left\{\log\Gamma\left(\sum_{i=1}^{K}\alpha_i\right)+\sum_{k=1}^{K}(\alpha_k-1)\log\theta_k-\sum_{k=1}^{K}\log\Gamma\left(\alpha_k\right)\right\} \\
&= \sum_{k=1}^{K}(\alpha_k-1)\mathrm{E}_q\left\{\log\theta_k\right\}+\log\Gamma\left(\sum_{i=1}^{K}\alpha_i\right)-\sum_{k=1}^{K}\log\Gamma\left(\alpha_k\right) \\
&= \sum_{k=1}^{K}(\alpha_k-1)\left(\Psi(\gamma_k)-\Psi\left(\sum_{i=1}^{K}\gamma_i\right)\right)+\log\Gamma\left(\sum_{i=1}^{K}\alpha_i\right)-\sum_{k=1}^{K}\log\Gamma\left(\alpha_k\right),
\end{aligned}
$$

where $\Psi(\cdot)$ is the digamma function, the first derivative of the log Gamma function. The final line is due to the following property of the Dirichlet distribution as a member of the exponential family. If $\theta \sim \mathrm{Dir}(\alpha)$, then $\mathrm{E}_{p(\theta|\alpha)}\left\{\log\theta_i\right\} = \Psi(\alpha_i) - \Psi(\sum_{i=1}^{K}\alpha_i)$.

**The second term** is

$$
\begin{aligned}
\mathrm{E}_q\left\{\log p\left(z \mid \theta\right)\right\} &= \mathrm{E}_q\left\{\log\prod_{n=1}^{N}p\left(z_n \mid \theta\right)\right\} \\
&= \mathrm{E}_q\left\{\log\prod_{n=1}^{N}\prod_{k=1}^{K}\theta_k^{\mathbb{1}_z(n,k)}\right\} \\
&= \sum_{n=1}^{N}\sum_{k=1}^{K}\mathrm{E}_q\left\{\mathbb{1}_z(n,k)\log\theta_k\right\} \\
&= \sum_{n=1}^{N}\sum_{k=1}^{K}\mathrm{E}_q\left\{\mathbb{1}_z(n,k)\right\}\mathrm{E}_q\left\{\log\theta_k\right\} \\
&= \sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{nk}\left(\Psi(\gamma_k)-\Psi\left(\sum_{i=1}^{K}\gamma_i\right)\right),
\end{aligned}
$$

where $\phi_{nk}$ is the probability of the $n$th voxel being recruited by subtype $k$, and $\mathbb{1}(\cdot)$ is the indicator function.

We expand **the third term** as

$$
\begin{aligned}
\mathrm{E}_q\left\{\log p\left(w \mid z,\beta\right)\right\} &= \mathrm{E}_q\left\{\log\prod_{n=1}^{N}p\left(w_n \mid z_n,\beta\right)\right\} \\
&= \mathrm{E}_q\left\{\log\prod_{n=1}^{N}\prod_{k=1}^{K}\prod_{v=1}^{V}\beta_{kv}^{\mathbb{1}_z(n,k)\mathbb{1}_w(n,v)}\right\} \\
&= \mathrm{E}_q\left\{\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{v=1}^{V}\mathbb{1}_z(n,k)\mathbb{1}_w(n,v)\log\beta_{kv}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{v=1}^{V} \mathrm{E}_q\left\{\mathbb{1}_z(n,k)\right\}\mathbb{1}_w(n,v)\log\beta_{kv} \\
&= \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{v=1}^{V} \phi_{nk}\mathbb{1}_w(n,v)\log\beta_{kv}.
\end{aligned}
$$

**The fourth (new) term** is

$$
\begin{aligned}
\mathrm{E}_q\left\{\log p\left(y \mid z, w, \rho\right)\right\} &= \mathrm{E}_q\left\{\log \prod_{n=1}^{N} p\left(y_n \mid z_n, w_n, \rho\right)\right\} \\
&= \mathrm{E}_q\left\{\log \prod_{n=1}^{N}\prod_{k=1}^{K}\prod_{v=1}^{V}\left(\rho_{kv}^{\mathbb{1}_y(n)}\left(1-\rho_{kv}\right)^{1-\mathbb{1}_y(n)}\right)^{\mathbb{1}_z(n,k)\mathbb{1}_w(n,v)}\right\} \\
&= \mathrm{E}_q\left\{\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{v=1}^{V}\mathbb{1}_z(n,k)\mathbb{1}_w(n,v)\left(\mathbb{1}_y(n)\log\rho_{kv} + (1-\mathbb{1}_y(n))\log(1-\rho_{kv})\right)\right\} \\
&= \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{v=1}^{V}\phi_{nk}\mathbb{1}_w(n,v)\left(\mathbb{1}_y(n)\log\rho_{kv} + (1-\mathbb{1}_y(n))\log(1-\rho_{kv})\right)
\end{aligned}
$$

Very similar to the first term, **the fifth term** is expanded as

$$
\mathrm{E}_q\left\{\log q\left(\theta \mid \gamma\right)\right\} = \sum_{k=1}^{K}(\gamma_k - 1)\left(\Psi(\gamma_k) - \Psi\left(\sum_{i=1}^{K}\gamma_i\right)\right) + \log\Gamma\left(\sum_{k=1}^{K}\gamma_k\right) - \sum_{k=1}^{K}\log\Gamma\left(\gamma_k\right).
$$

Finally, **the sixth term** is expanded as

$$
\begin{aligned}
\mathrm{E}_q\left\{\log q\left(z \mid \phi\right)\right\} &= \mathrm{E}_q\left\{\log\prod_{n=1}^{N} q\left(z_n \mid \phi_n\right)\right\} \\
&= \mathrm{E}_q\left\{\log\prod_{n=1}^{N}\prod_{k=1}^{K}\phi_{nk}^{\mathbb{1}_z(n,k)}\right\} \\
&= \sum_{n=1}^{N}\sum_{k=1}^{K}\mathrm{E}_q\left\{\mathbb{1}_z(n,k)\right\}\log\phi_{nk} \\
&= \sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{nk}\log\phi_{nk}.
\end{aligned}
$$

Therefore, the fully expanded lower bound is

$$
\begin{aligned}
\mathcal{L}\left(\gamma, \phi \mid \alpha, \beta\right) = &\sum_{k=1}^{K}(\alpha_k - 1)\left(\Psi(\gamma_k) - \Psi\left(\sum_{i=1}^{K}\gamma_i\right)\right) + \log\Gamma\left(\sum_{i=1}^{K}\alpha_i\right) - \sum_{k=1}^{K}\log\Gamma\left(\alpha_k\right) \\
&+ \sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{nk}\left(\Psi(\gamma_k) - \Psi\left(\sum_{i=1}^{K}\gamma_i\right)\right) \\
&+ \sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{v=1}^{V}\phi_{nk}\mathbb{1}_w(n,v)\log\beta_{kv} \qquad (3)
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{n=1}^{N}\sum_{k=1}^{K}\sum_{v=1}^{V}\phi_{nk}\mathbb{1}_w(n,v)\left(\mathbb{1}_y(n)\log\rho_{kv}+(1-\mathbb{1}_y(n))\log(1-\rho_{kv})\right) \\
&-\sum_{k=1}^{K}(\gamma_k-1)\left(\Psi(\gamma_k)-\Psi\left(\sum_{i=1}^{K}\gamma_i\right)\right)-\log\Gamma\left(\sum_{k=1}^{K}\gamma_k\right)+\sum_{k=1}^{K}\log\Gamma\left(\gamma_k\right) \\
&-\sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{nk}\log\phi_{nk}.
\end{aligned}
$$

# 4 Maximizing the Lower Bound

In this section, we maximize the lower bound w.r.t. the variational parameters $\phi$ and $\gamma$. Recall that as the maximization runs, the KL divergence between the variational distribution and the true posterior drops (E-step of the variational EM algorithm).

## 4.1 Variational Multinomial

We first maximize Equation (3) w.r.t. $\phi_{nk}$. Since $\sum_{k=1}^{K}\phi_{nk}=1$, this is a constrained optimization problem that can be solved by the Lagrange multiplier method. The Lagrangian w.r.t. $\phi_{nk}$ is

$$
\begin{aligned}
\mathcal{L}_{[\phi_{nk}]}=\phi_{nk}&\left(\Psi(\gamma_k)-\Psi\left(\sum_{i=1}^{K}\gamma_i\right)\right) \\
&+\sum_{v=1}^{V}\phi_{nk}\mathbb{1}_w(n,v)\log\beta_{kv} \\
&+\sum_{v=1}^{V}\phi_{nk}\mathbb{1}_w(n,v)\left(\mathbb{1}_y(n)\log\rho_{kv}+(1-\mathbb{1}_y(n))\log(1-\rho_{kv})\right) \\
&-\phi_{nk}\log\phi_{nk}+\lambda_n\left(\sum_{i=1}^{K}\phi_{ni}-1\right),
\end{aligned}
$$

where $\lambda_n$ is the Lagrange multiplier. Taking the derivative, we get

$$
\begin{aligned}
\frac{\partial}{\partial\phi_{nk}}\mathcal{L}_{[\phi_{nk}]}=\Psi(\gamma_k)&-\Psi\left(\sum_{i=1}^{K}\gamma_i\right) \\
&+\sum_{v=1}^{V}\mathbb{1}_w(n,v)\log\beta_{kv} \\
&+\sum_{v=1}^{V}\mathbb{1}_w(n,v)(\mathbb{1}_y(n)\log\rho_{kv}+(1-\mathbb{1}_y(n))\log(1-\rho_{kv})) \\
&-\log\phi_{nk}-1+\lambda_n.
\end{aligned}
$$

Setting this derivative to zero yields

$$
\phi_{nk}=\exp\left(\Psi(\gamma_k)-\Psi\left(\sum_{i=1}^{K}\gamma_i\right)+\lambda_n-1\right)\prod_{v=1}^{V}\left(\beta_{kv}\cdot\rho_{kv}^{\mathbb{1}_y(n)}\cdot(1-\rho_{kv})^{1-\mathbb{1}_y(n)}\right)^{\mathbb{1}_w(n,v)}
$$

$$\propto \exp\left(\Psi(\gamma_k) - \Psi\left(\sum_{i=1}^{K}\gamma_i\right)\right)\prod_{v=1}^{V}\left(\beta_{kv}\cdot\rho_{kv}^{\mathbb{1}_y(n)}\cdot(1-\rho_{kv})^{1-\mathbb{1}_y(n)}\right)^{\mathbb{1}_w(n,v)}.$$

### 4.2 Variational Dirichlet

Now we maximize Equation (3) w.r.t. $\gamma_k$, the $k$th component of the Dirichlet parameter. Only the terms containing $\gamma_k$ are retained.

$$\begin{aligned}
\mathcal{L}_{[\gamma]} = &\sum_{k=1}^{K}(\alpha_k - 1)\left(\Psi(\gamma_k) - \Psi\left(\sum_{i=1}^{K}\gamma_i\right)\right)\\
&+\sum_{n=1}^{N}\sum_{k=1}^{K}\phi_{nk}\left(\Psi(\gamma_k) - \Psi\left(\sum_{i=1}^{K}\gamma_i\right)\right)\\
&-\sum_{k=1}^{K}(\gamma_k - 1)\left(\Psi(\gamma_k) - \Psi\left(\sum_{i=1}^{K}\gamma_i\right)\right) - \log\Gamma\left(\sum_{i=1}^{K}\gamma_i\right) + \sum_{k=1}^{K}\log\Gamma\left(\gamma_k\right)
\end{aligned}$$

Taking the derivative w.r.t. $\gamma_k$, we have

$$\begin{aligned}
\frac{\partial}{\partial\gamma_k}\mathcal{L}_{[\gamma]} = &\left(\Psi'(\gamma_k) - \Psi'\left(\sum_{i=1}^{K}\gamma_i\right)\right)(\alpha_k - 1)\\
&+\left(\Psi'(\gamma_k) - \Psi'\left(\sum_{i=1}^{K}\gamma_i\right)\right)\sum_{n=1}^{N}\phi_{nk}\\
&-\left(\Psi'(\gamma_k) - \Psi'\left(\sum_{i=1}^{K}\gamma_i\right)\right)(\gamma_k - 1) - \left(\Psi(\gamma_k) - \Psi\left(\sum_{i=1}^{K}\gamma_i\right)\right)\\
&-\frac{\Psi\left(\sum_{i=1}^{K}\gamma_i\right)}{\Gamma\left(\sum_{i=1}^{K}\gamma_i\right)} + \frac{\Psi(\gamma_k)}{\Gamma\left(\gamma_k\right)}\\
= &\left(\Psi'(\gamma_k) - \Psi'\left(\sum_{i=1}^{K}\gamma_i\right)\right)\left(\alpha_k + \sum_{n=1}^{N}\phi_{nk} - \gamma_k\right) - \Psi(\gamma_k) + \Psi\left(\sum_{i=1}^{K}\gamma_i\right)\\
&-\Psi\left(\sum_{i=1}^{K}\gamma_i\right) + \Psi(\gamma_k)\\
= &\left(\Psi'(\gamma_k) - \Psi'\left(\sum_{i=1}^{K}\gamma_i\right)\right)\left(\alpha_k + \sum_{n=1}^{N}\phi_{nk} - \gamma_k\right).
\end{aligned}$$

Setting it to zero, we have

$$\gamma_k = \alpha_k + \sum_{n=1}^{N}\phi_{nk}.$$

## 5 Estimating Model Parameters

The previous section is the E-step of the variational EM algorithm, where we tighten the lower bound w.r.t. the variational parameters; this section is the M-step, where we maximize the lower bound w.r.t. the model parameters $\alpha$, $\beta$, and $\rho$. Now we add back the document subscript to consider the whole corpus.

By the assumed exchangeability of the documents, the overall log-likelihood of the corpus is just the sum of all the documents' log-likelihoods, and the overall lower bound is just the sum of the individual lower bounds. Again, only the terms involving $\beta$ are left in the overall lower bound. Adding the Lagrange multipliers, we obtain

$$\mathcal{L}_{[\beta]} = \sum_{m=1}^{M}\sum_{n=1}^{N_m}\sum_{k=1}^{K}\sum_{v=1}^{V} \phi_{mnk}\mathbb{1}_w(m,n,v)\log\beta_{kv} + \sum_{k=1}^{K}\lambda_k\left(\sum_{v=1}^{V}\beta_{kv} - 1\right).$$

Taking the derivative w.r.t. $\beta_{kv}$ and setting it to zero, we have

$$\begin{aligned}
\frac{\partial}{\partial\beta_{kv}}\mathcal{L}_{[\beta]} &= \sum_{m=1}^{M}\sum_{n=1}^{N_m}\phi_{mnk}\mathbb{1}_w(m,n,v)\frac{1}{\beta_{kv}} + \lambda_k = 0 \\
\Rightarrow \beta_{kv} &= -\frac{1}{\lambda_k}\sum_{m=1}^{M}\sum_{n=1}^{N_m}\phi_{mnk}\mathbb{1}_w(m,n,v) \\
\Rightarrow \beta_{kv} &\propto \sum_{m=1}^{M}\sum_{n=1}^{N_m}\phi_{mnk}\mathbb{1}_w(m,n,v).
\end{aligned}$$

For $\alpha$,

$$\mathcal{L}_{[\alpha]} = \sum_{m=1}^{M}\left(\sum_{k=1}^{K}(\alpha_k - 1)\left(\Psi(\gamma_{mk}) - \Psi\left(\sum_{i=1}^{K}\gamma_{mi}\right)\right) + \log\Gamma\left(\sum_{i=1}^{K}\alpha_i\right) - \sum_{k=1}^{K}\log\Gamma\left(\alpha_k\right)\right)$$

Assuming a symmetric Dirichlet prior (i.e., $\alpha_k = \alpha, \forall k$), we have

$$\begin{aligned}
\mathcal{L}_{[\alpha]} &= \sum_{m=1}^{M}\left(\sum_{k=1}^{K}(\alpha - 1)\left(\Psi(\gamma_{mk}) - \Psi\left(\sum_{i=1}^{K}\gamma_{mi}\right)\right)\right) + M\log\Gamma\left(K\alpha\right) - MK\log\Gamma\left(\alpha\right) \\
\frac{\partial}{\partial\alpha_k}\mathcal{L}_{[\alpha]} &= \sum_{m=1}^{M}\sum_{j=1}^{K}\left(\Psi(\gamma_{mj}) - \Psi\left(\sum_{i=1}^{K}\gamma_{mi}\right)\right) + MK\Psi\left(K\alpha\right) - MK\Psi(\alpha).
\end{aligned}$$

Since the derivative also depends on other $\alpha_{k'\neq k}$ in general asymmetric cases, we compute the Hessian

$$\frac{\partial^2}{\partial\alpha_k\partial\alpha_{k'}}\mathcal{L}_{[\alpha]} = MK^2\Psi'\left(K\alpha\right) - MK\Psi'(\alpha),$$

and notice that its form allows for the linear-time Newton-Raphson algorithm.

Finally, for $\rho$,

$$\mathcal{L}_{[\rho]} = \sum_{m=1}^{M}\sum_{n=1}^{N_m}\sum_{k=1}^{K}\sum_{v=1}^{V}\phi_{mnk}\mathbb{1}_w(m,n,v)\left(\mathbb{1}_y(m,n)\log\rho_{kv} + (1 - \mathbb{1}_y(m,n))\log(1-\rho_{kv})\right)$$

Taking the derivative w.r.t. $\rho_{kv}$ and setting it to zero, we have

$$
\begin{aligned}
&\frac{\partial}{\partial \rho_{kv}} \mathcal{L}_{[\rho]} = \sum_{m=1}^{M} \sum_{n=1}^{N_m} \phi_{mnk} \mathbb{1}_w(m,n,v) \left( \frac{\mathbb{1}_y(m,n)}{\rho_{kv}} - \frac{1 - \mathbb{1}_y(m,n)}{1 - \rho_{kv}} \right) = 0 \\
\Rightarrow & \sum_{m=1}^{M} \sum_{n=1}^{N_m} \phi_{mnk} \mathbb{1}_w(m,n,v) \left( \mathbb{1}_y(m,n) - \rho_{kv} \right) = 0 \\
\Rightarrow & \rho_{kv} = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N_m} \phi_{mnk} \mathbb{1}_w(m,n,v) \mathbb{1}_y(m,n)}{\sum_{m=1}^{M} \sum_{n=1}^{N_m} \phi_{mnk} \mathbb{1}_w(m,n,v)}.
\end{aligned}
$$