Supplementary material for: Learning Supervised Topic Models for Classification and Regression from Crowds

Filipe Rodrigues (rodr@dtu.dk)
Mariana Lourenço (mrlouren@student.dei.uc.pt),
Bernardete Ribeiro (bribeiro@dei.uc.pt,
Francisco C. Pereira (camara@smart.mit.edu)

In this supplementary material we provide more details on deriving the (batch) variational Bayesian EM algorithm presented in the paper (Section 3), and also provide additional results for the experiments on the LabelMe dataset.

1 Deriving the lower bound

The variational objective function (or the evidence lower bound or ELBO) is given by

$$\begin{aligned}
&\log p(\mathbf{w}_{1:D}, \mathbf{y}_{1:D} | \alpha, \tau, \omega, \boldsymbol{\eta}) \\
&\quad = \log \int_{\pi} \int_{\beta} \int_{\theta} \sum_{\mathbf{z}} \sum_{c} \frac{p(\boldsymbol{\theta}, \mathbf{z}_{1:D}, \boldsymbol{c}, \mathbf{w}_{1:D}, \mathbf{y}_{1:D}, \boldsymbol{\beta}, \boldsymbol{\pi}_{1:R} | \Theta) q(\boldsymbol{\theta}, \mathbf{z}_{1:D}, \boldsymbol{c}, \boldsymbol{\beta}, \boldsymbol{\pi}_{1:R})}{q(\boldsymbol{\theta}, \mathbf{z}_{1:D}, \boldsymbol{c}, \boldsymbol{\beta}, \boldsymbol{\pi}_{1:R})} \qquad (1) \\
&\quad \geqslant \mathcal{L}(\mathbf{w}_{1:D}, \mathbf{y}_{1:D} | \Theta) \\
&\quad = \mathbb{E}_q[\log p(\boldsymbol{\theta}, \mathbf{z}_{1:D}, \boldsymbol{c}, \mathbf{w}_{1:D}, \mathbf{y}_{1:D}, \boldsymbol{\beta}, \boldsymbol{\pi}_{1:R} | \Theta)] \underbrace{- \mathbb{E}_q[\log q(\boldsymbol{\theta}, \mathbf{z}_{1:D}, \boldsymbol{c}, \boldsymbol{\beta}, \boldsymbol{\pi}_{1:R})]}_{\mathcal{H}(q)} \qquad (2)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^{K} \mathbb{E}_q[\log p(\beta_i | \tau)] + \sum_{r=1}^{R} \sum_{c=1}^{C} \mathbb{E}_q[\log p(\pi_c^r | \omega)] + \sum_{d=1}^{D} \Bigg(\mathbb{E}_q[\log p(\theta^d | \alpha)] + \sum_{n=1}^{N^d} \mathbb{E}_q[\log p(z_n^d | \theta^d)] \\
&+ \sum_{n=1}^{N^d} \mathbb{E}_q[\log p(w_n^d | z_n^d, \boldsymbol{\beta})] + \mathbb{E}_q[\log p(c^d | \bar{z}^d, \boldsymbol{\eta})] + \sum_{r=1}^{R} \mathbb{E}_q[\log p(y^{d,r} | c^d, \boldsymbol{\pi}^r)] \Bigg) + \mathcal{H}(q) \qquad (3)
\end{aligned}$$

where the entropy $\mathcal{H}(q)$ of the variational distribution is given by

$$\begin{aligned}
\mathcal{H}(q) = &- \sum_{r=1}^{R} \sum_{c=1}^{C} \mathbb{E}_q[\log q(\pi_c^r | \xi_c^r)] - \sum_{i=1}^{K} \mathbb{E}_q[\log q(\beta_i | \zeta_i)] \\
&- \sum_{d=1}^{D} \Bigg(\mathbb{E}_q[\log q(\theta^d | \gamma^d)] - \sum_{n=1}^{N^d} \mathbb{E}_q[\log q(z_n^d | \phi_n^d)] - \mathbb{E}_q[\log q(c^d | \lambda^d)] \Bigg). \qquad (4)
\end{aligned}$$

The terms needed for the lower bound are given by

$$
\begin{aligned}
\mathbb{E}_q[\log p(\beta_i|\tau)] &= \mathbb{E}_q\left[\log \frac{\Gamma(\sum_{k=1}^{V}\tau_k)}{\prod_{j=1}^{V}\Gamma(\tau_j)}\prod_{j=1}^{V}\beta_{i,j}^{(\tau_j-1)}\right] \\
&= \log\Gamma\left(\sum_{k=1}^{V}\tau_k\right) - \sum_{j=1}^{V}\log\Gamma(\tau_j) + \sum_{j=1}^{V}(\tau_j-1)\mathbb{E}_q[\log\beta_{i,j}] \qquad (5) \\
\mathbb{E}_q[\log p(\pi_c^r|\omega)] &= \mathbb{E}_q\left[\log \frac{\Gamma(\sum_{t=1}^{C}\omega_t)}{\prod_{l=1}^{C}\Gamma(\omega_l)}\prod_{l=1}^{C}\left(\pi_{c,l}^r\right)^{(\omega_l-1)}\right] \\
&= \log\Gamma\left(\sum_{t=1}^{C}\omega_t\right) - \sum_{l=1}^{C}\log\Gamma(\omega_l) + \sum_{l=1}^{C}(\omega_l-1)\mathbb{E}_q[\log\pi_{c,l}^r] \qquad (6) \\
\mathbb{E}_q[\log p(\theta^d|\alpha)] &= \mathbb{E}_q\left[\log \frac{\Gamma(\sum_{j=1}^{K}\alpha_j)}{\prod_{i=1}^{K}\Gamma(\alpha_i)}\prod_{i=1}^{K}(\theta_i^d)^{(\alpha_i-1)}\right] \\
&= \log\Gamma\left(\sum_{j=1}^{K}\alpha_j\right) - \sum_{i=1}^{K}\log\Gamma(\alpha_i) + \sum_{i=1}^{K}(\alpha_i-1)\mathbb{E}_q[\log\theta_i^d] \qquad (7) \\
\mathbb{E}_q[\log p(z_n^d|\theta^d)] &= \mathbb{E}_q\left[\log\prod_{i=1}^{K}\left(\theta_i^d\right)^{z_{n,i}^d}\right] = \sum_{i=1}^{K}\phi_{n,i}^d\mathbb{E}_q[\log\theta_i^d] \qquad (8) \\
\mathbb{E}_q[\log p(w_n^d|z_n^d,\boldsymbol{\beta})] &= \mathbb{E}_q\left[\log\prod_{j=1}^{V}(\beta_{z_n^d,j})^{w_{n,j}^d}\right] = \sum_{j=1}^{V}\sum_{i=1}^{K}w_{n,j}^d\phi_{n,i}^d\mathbb{E}_q[\log\beta_{i,j}] \qquad (9) \\
\mathbb{E}_q[\log p(y^{d,r}|c^d,\boldsymbol{\pi}^r)] &= \mathbb{E}_q\left[\log\prod_{l=1}^{C}(\pi_{c^d,l}^r)^{y_l^{d,r}}\right] = \sum_{c=1}^{C}\sum_{l=1}^{C}\lambda_c^d y_l^{d,r}\mathbb{E}_q[\log\pi_{c,l}^r] \qquad (10) \\
\mathbb{E}_q[\log q(\pi_c^r|\xi_c^r)] &= \mathbb{E}_q\left[\log \frac{\Gamma(\sum_{t=1}^{C}\xi_{c,t}^r)}{\prod_{l=1}^{C}\Gamma(\xi_{c,l}^r)}\prod_{l=1}^{C}\left(\pi_{c,l}^r\right)^{(\xi_{c,l}^r-1)}\right] \\
&= \log\Gamma\left(\sum_{t=1}^{C}\xi_{c,t}^r\right) - \sum_{l=1}^{C}\log\Gamma(\xi_{c,l}^r) + \sum_{l=1}^{C}(\xi_{c,l}^r-1)\mathbb{E}_q[\log\pi_{c,l}^r] \qquad (11) \\
\mathbb{E}_q[\log q(\beta_i|\zeta_i)] &= \mathbb{E}_q\left[\log \frac{\Gamma(\sum_{k=1}^{V}\zeta_{i,k})}{\prod_{j=1}^{V}\Gamma(\zeta_{i,j})}\prod_{j=1}^{V}\left(\beta_{i,j}\right)^{(\zeta_{i,j}-1)}\right] \\
&= \log\Gamma\left(\sum_{k=1}^{V}\zeta_{i,k}\right) - \sum_{j=1}^{V}\log\Gamma(\zeta_{i,j}) + \sum_{j=1}^{V}(\zeta_{i,j}-1)\mathbb{E}_q[\log\beta_{i,j}] \qquad (12) \\
\mathbb{E}_q[\log q(\theta^d|\gamma^d)] &= \mathbb{E}_q\left[\log \frac{\Gamma(\sum_{j=1}^{K}\gamma_j^d)}{\prod_{i=1}^{K}\Gamma(\gamma_i^d)}\prod_{i=1}^{K}\left(\theta_i^d\right)^{(\gamma_i^d-1)}\right] \\
&= \log\Gamma\left(\sum_{j=1}^{K}\gamma_j^d\right) - \sum_{i=1}^{K}\log\Gamma(\gamma_i^d) + \sum_{i=1}^{K}(\gamma_i^d-1)\mathbb{E}_q[\log\theta_i^d] \qquad (13) \\
& \qquad (14)
\end{aligned}
$$

$$\mathbb{E}_q[\log q(z_n^d|\phi_n^d)] = \mathbb{E}_q\Bigg[\log \prod_{i=1}^{K} \left(\phi_{n,i}^d\right)^{z_{n,i}^d}\Bigg] = \sum_{i=1}^{K} \phi_{n,i}^d \log \phi_{n,i}^d \tag{15}$$

$$\mathbb{E}_q[\log q(c^d|\lambda^d)] = \mathbb{E}_q\Bigg[\log \prod_{l=1}^{C} \left(\lambda_l^d\right)^{c_l^d}\Bigg] = \sum_{l=1}^{C} \lambda_l^d \log \lambda_l^d, \tag{16}$$

where

$$\mathbb{E}_q[\log \theta_i^d] = \Psi(\gamma_i^d) - \Psi\Bigg(\sum_{j=1}^{K} \gamma_j^d\Bigg) \tag{17}$$

$$\mathbb{E}_q[\log \beta_{i,j}] = \Psi(\zeta_{i,j}) - \Psi\Bigg(\sum_{k=1}^{V} \zeta_{i,k}\Bigg) \tag{18}$$

$$\mathbb{E}_q[\log \pi_{c,l}^r] = \Psi(\xi_{c,l}^r) - \Psi\Bigg(\sum_{t=1}^{C} \xi_{c,t}^r\Bigg). \tag{19}$$

Finally, the expectation of the log probability of the latent classes is given by:

$$\mathbb{E}_q[\log p(c^d|\bar{z}^d, \boldsymbol{\eta})] = \mathbb{E}_q\Bigg[\log \frac{\exp(\eta_{c^d}^T \bar{z}^d)}{\sum_{l=1}^{C} \exp(\eta_l^T \bar{z}^d)}\Bigg] = \mathbb{E}_q[\eta_{c^d}^T \bar{z}^d] - \mathbb{E}_q\Bigg[\log \sum_{l=1}^{C} \exp(\eta_l^T \bar{z}^d)\Bigg], \tag{20}$$

where the first term can be easily computed as $\mathbb{E}_q[\eta_{c^d}^T \bar{z}^d] = \sum_{l=1}^{C} \lambda_l^d \eta_l^T \bar{\phi}^d$ and the second term can be lower-bounded by appealing again to the Jensen's inequality as follows:

$$\begin{aligned}
-\mathbb{E}_q\Bigg[\log \sum_{l=1}^{C} \exp(\eta_l^T \bar{z}^d)\Bigg] &\geqslant -\log \sum_{l=1}^{C} \mathbb{E}_q[\exp(\eta_l^T \bar{z}^d)] \\
&= -\log \sum_{l=1}^{C} \mathbb{E}_q\Bigg[\exp(\eta_l^T \frac{1}{N^d} \sum_{j=1}^{N^d} z_j^d)\Bigg] \\
&= -\log \sum_{l=1}^{C} \prod_{j=1}^{N^d} \left(\phi_j^d\right)^T \exp\left(\frac{1}{N^d}\eta_l\right) \\
&= -\log \sum_{l=1}^{C} \left(\phi_n^d\right)^T \exp\left(\frac{1}{N^d}\eta_l\right) \prod_{j=1, j\neq n}^{N^d} \left(\phi_j^d\right)^T \exp\left(\frac{1}{N^d}\eta_l\right) \\
&= -\log \left(\phi_n^d\right)^T \underbrace{\sum_{l=1}^{C} \exp\left(\frac{1}{N^d}\eta_l\right) \prod_{j=1, j\neq n}^{N^d} \left(\phi_j^d\right)^T \exp\left(\frac{1}{N^d}\eta_l\right)}_{=a} \\
&= -\log a^T \phi_n^d,
\end{aligned} \tag{21}$$

where we defined $a = \sum_{l=1}^{C} \exp(\frac{1}{N^d}\eta_l) \prod_{j=1, j\neq n}^{N^d} \left(\phi_j^d\right)^T \exp\left(\frac{1}{N^d}\eta_l\right)$.

Putting all the terms together, the lower-bound becomes:

$$
\begin{aligned}
&\mathcal{L}(\mathbf{w}_{1:D}, \mathbf{y}_{1:D}|\Theta) \\
&\quad = \sum_{i=1}^{K}\left(\log\Gamma\left(\sum_{k=1}^{V}\tau_k\right) - \sum_{j=1}^{V}\log\Gamma(\tau_j) + \sum_{j=1}^{V}(\tau_j - 1)\left(\Psi(\zeta_{i,j}) - \Psi\left(\sum_{k=1}^{V}\zeta_{i,k}\right)\right)\right) \\
&\quad + \sum_{r=1}^{R}\sum_{c=1}^{C}\left(\log\Gamma\left(\sum_{t=1}^{C}\omega_t\right) - \sum_{l=1}^{C}\log\Gamma(\omega_l) + \sum_{l=1}^{C}(\omega_l - 1)\left(\Psi(\xi^r_{c,l}) - \Psi\left(\sum_{t=1}^{C}\xi^r_{c,t}\right)\right)\right) \\
&\quad + \sum_{d=1}^{D}\left(\log\Gamma\left(\sum_{j=1}^{K}\alpha_j\right) - \sum_{i=1}^{K}\log\Gamma(\alpha_i) + \sum_{i=1}^{K}(\alpha_i - 1)\left(\Psi(\gamma^d_i) - \Psi\left(\sum_{j=1}^{K}\gamma^d_j\right)\right)\right) \\
&\quad + \sum_{d=1}^{D}\sum_{n=1}^{N^d}\sum_{i=1}^{K}\phi^d_{n,i}\left(\Psi(\gamma^d_i) - \Psi\left(\sum_{j=1}^{K}\gamma^d_j\right)\right) \\
&\quad + \sum_{d=1}^{D}\sum_{n=1}^{N^d}\sum_{j=1}^{V}\sum_{i=1}^{K}w^d_{n,j}\phi^d_{n,i}\left(\Psi(\zeta_{i,j}) - \Psi\left(\sum_{k=1}^{V}\zeta_{i,k}\right)\right) \\
&\quad + \sum_{d=1}^{D}\left(\sum_{l=1}^{C}\lambda^d_l\eta^T_l\bar{\phi}^d - (a^T(\phi^d_n)^{old})^{-1}(a^T\phi^d_n) - \log(a^T(\phi^d_n)^{old}) + 1\right) \\
&\quad + \sum_{d=1}^{D}\sum_{r=1}^{R}\sum_{c=1}^{C}\sum_{l=1}^{C}\lambda^d_c y^{d,r}_l\left(\Psi(\xi^r_{c,l}) - \Psi\left(\sum_{t=1}^{C}\xi^r_{c,t}\right)\right) \\
&\quad - \sum_{r=1}^{R}\sum_{c=1}^{C}\left(\log\Gamma\left(\sum_{t=1}^{C}\xi^r_{c,t}\right) - \sum_{l=1}^{C}\log\Gamma(\xi^r_{c,l}) + \sum_{l=1}^{C}(\xi^r_{c,l} - 1)\left(\Psi(\xi^r_{c,l}) - \Psi\left(\sum_{t=1}^{C}\xi^r_{c,t}\right)\right)\right) \\
&\quad - \sum_{i=1}^{K}\left(\log\Gamma\left(\sum_{k=1}^{V}\zeta_{i,k}\right) - \sum_{j=1}^{V}\log\Gamma(\zeta_{i,j}) + \sum_{j=1}^{V}(\zeta_{i,j} - 1)\left(\Psi(\zeta_{i,j}) - \Psi\left(\sum_{k=1}^{V}\zeta_{i,k}\right)\right)\right) \\
&\quad - \sum_{d=1}^{D}\left(\log\Gamma\left(\sum_{j=1}^{K}\gamma^d_j\right) - \sum_{i=1}^{K}\log\Gamma(\gamma^d_i) + \sum_{i=1}^{K}(\gamma^d_i - 1)\left(\Psi(\gamma^d_i) - \Psi\left(\sum_{j=1}^{K}\gamma^d_j\right)\right)\right) \\
&\quad - \sum_{d=1}^{D}\sum_{n=1}^{N^d}\sum_{i=1}^{K}\phi^d_{n,i}\log\phi^d_{n,i} \\
&\quad - \sum_{d=1}^{D}\sum_{l=1}^{C}\lambda^d_l\log\lambda^d_l.
\end{aligned}
\tag{22}
$$

2 Optimizing the lower bound (E-step)

2.1 Optimizing w.r.t. γ_i^d

Collecting only the terms in the bound that contain γ gives

$$\mathcal{L}_{[\gamma]} = \sum_{d=1}^{D}\sum_{i=1}^{K}\Psi(\gamma_i^d)\left(\alpha_i + \sum_{n=1}^{N^d}\phi_{n,i}^d - \gamma_i^d\right) - \sum_{d=1}^{D}\sum_{i=1}^{K}\Psi\left(\sum_{j=1}^{K}\gamma_j^d\right)\left(\alpha_i + \sum_{n=1}^{N^d}\phi_{n,i}^d - \gamma_i^d\right)$$
$$- \sum_{d=1}^{D}\log\Gamma\left(\sum_{j=1}^{K}\gamma_j^d\right) + \sum_{d=1}^{D}\sum_{i=1}^{K}\log\Gamma(\gamma_i^d). \tag{23}$$

Taking derivatives w.r.t. γ_i^d gives

$$\frac{\partial\mathcal{L}_{[\gamma]}}{\partial\gamma_i^d} = \Psi'(\gamma_i^d)\left(\alpha_i + \sum_{n=1}^{N^d}\phi_{n,i}^d - \gamma_i^d\right) - \Psi'\left(\sum_{j=1}^{K}\gamma_j^d\right)\sum_{j=1}^{K}\left(\alpha_j + \sum_{n=1}^{N^d}\phi_{n,j}^d - \gamma_j^d\right). \tag{24}$$

Setting this derivative to zero in order to get a maximum, we get the solution

$$\gamma_i^d = \alpha_i + \sum_{n=1}^{N^d}\phi_{n,i}^d. \tag{25}$$

2.2 Optimizing w.r.t. $\phi_{n,i}^d$

Collecting only the terms in the bound that contain $\phi_{n,i}^d$ and adding Lagrange multipliers gives

$$\mathcal{L}_{[\phi_{n,i}^d]} = \sum_{d=1}^{D}\sum_{n=1}^{N^d}\sum_{i=1}^{K}\phi_{n,i}^d\left(\Psi(\gamma_i^d) - \Psi\left(\sum_{j=1}^{K}\gamma_j^d\right)\right) + \sum_{d=1}^{D}\sum_{n=1}^{N^d}\sum_{j=1}^{V}\sum_{i=1}^{K}w_{n,j}^d\phi_{n,i}^d\left(\Psi(\zeta_{i,j}) - \Psi\left(\sum_{k=1}^{V}\zeta_{i,k}\right)\right)$$
$$+ \sum_{d=1}^{D}\left(\frac{1}{N^d}\sum_{l=1}^{C}\lambda_l^d\sum_{n=1}^{N^d}\eta_l^T\phi_n^d - (a^T(\phi_n^d)^{old})^{-1}(a^T\phi_n^d) - \log(a^T(\phi_n^d)^{old})\right)$$
$$- \sum_{d=1}^{D}\sum_{n=1}^{N^d}\sum_{i=1}^{K}\phi_{n,i}^d\log\phi_{n,i}^d + \mu\left(\sum_{k=1}^{K}\phi_{n,k}^d - 1\right). \tag{26}$$

Taking derivatives w.r.t. $\phi_{n,i}^d$ gives

$$\frac{\partial\mathcal{L}_{[\phi_{n,i}^d]}}{\partial\phi_{n,i}^d} = \Psi(\gamma_i^d) - \Psi\left(\sum_{j=1}^{K}\gamma_j^d\right) + \sum_{j=1}^{V}w_{n,j}^d\Psi(\zeta_{i,j}) - \sum_{j=1}^{V}w_{n,j}^d\Psi\left(\sum_{k=1}^{V}\zeta_{i,k}\right)$$
$$+ \frac{1}{N^d}\sum_{l=1}^{C}\lambda_l^d\eta_{l,i} - (a^T(\phi_n^d)^{old})^{-1}a_i - \log\phi_{n,i}^d - 1 + \mu. \tag{27}$$

The updates for $\phi^d_{n,i}$ are then given by

$$\phi^d_{n,i} \propto \exp\left(\Psi(\gamma_i) + \sum_{j=1}^{V} w^d_{n,j}\Psi(\zeta_{i,j}) - \sum_{j=1}^{V} w^d_{n,j}\Psi\left(\sum_{k=1}^{V}\zeta_{i,k}\right) + \frac{1}{N^d}\sum_{l=1}^{C}\lambda^d_l \eta_{l,i} - (a^T(\phi^d_n)^{old})^{-1}a_i\right). \tag{28}$$

2.3 Optimizing w.r.t. λ^d_l

Collecting only the terms in the bound that contain λ^d_l and adding Lagrange multipliers gives

$$\mathcal{L}_{[\lambda^d_l]} = \sum_{d=1}^{D}\sum_{l=1}^{C}\lambda^d_l \eta^T_l \bar{\phi}^d + \sum_{d=1}^{D}\sum_{r=1}^{R}\sum_{l=1}^{C}\sum_{c=1}^{C}\lambda^d_l y^{d,r}_c \left(\Psi(\xi^r_{l,c}) - \Psi\left(\sum_{t=1}^{C}\xi^r_{l,t}\right)\right)$$
$$- \sum_{l=1}^{C}\lambda^d_l \log\lambda^d_l + \mu\left(\sum_{k=1}^{C}\lambda^d_k - 1\right). \tag{29}$$

Taking derivatives w.r.t. λ^d_l gives

$$\frac{\partial\mathcal{L}_{[\lambda^d_l]}}{\partial\lambda^d_l} = \eta^T_l\bar{\phi}^d + \sum_{r=1}^{R}\sum_{c=1}^{C} y^{d,r}_c\Psi(\xi^r_{l,c}) - \sum_{r=1}^{R}\sum_{c=1}^{C} y^{d,r}_c\Psi\left(\sum_{t=1}^{C}\xi^r_{l,t}\right) - \log\lambda^d_l - 1 + \mu. \tag{30}$$

The updates for λ^d_l are then given by

$$\lambda^d_l \propto \exp\left(\eta^T_l\bar{\phi}^d + \sum_{r=1}^{R}\sum_{c=1}^{C} y^{d,r}_c\Psi(\xi^r_{l,c}) - \sum_{r=1}^{R}\sum_{c=1}^{C} y^{d,r}_c\Psi\left(\sum_{t=1}^{C}\xi^r_{l,t}\right)\right). \tag{31}$$

2.4 Optimizing w.r.t. $\zeta_{i,j}$

Collecting only the terms in the bound that contain ζ gives

$$\mathcal{L}_{[\zeta]} = \sum_{i=1}^{K}\sum_{j=1}^{V}\Psi(\zeta_{i,j})\left(\tau_j + \sum_{d=1}^{D}\sum_{n=1}^{N^d} w^d_{n,j}\phi^d_{n,i} - \zeta_{i,j}\right) - \sum_{i=1}^{K}\sum_{j=1}^{V}\Psi\left(\sum_{k=1}^{V}\zeta_{i,k}\right)\left(\tau_j + \sum_{d=1}^{D}\sum_{n=1}^{N^d} w^d_{n,j}\phi^d_{n,i} - \zeta_{i,j}\right)$$
$$- \sum_{i=1}^{K}\log\Gamma\left(\sum_{k=1}^{V}\zeta_{i,k}\right) + \sum_{i=1}^{K}\sum_{j=1}^{V}\log\Gamma(\zeta_{i,j}). \tag{32}$$

Taking derivatives w.r.t. $\zeta_{i,j}$ gives

$$\frac{\partial\mathcal{L}_{[\zeta]}}{\partial\zeta_{i,j}} = \Psi'(\zeta_{i,j})\left(\tau_j + \sum_{d=1}^{D}\sum_{n=1}^{N^d} w^d_{n,j}\phi^d_{n,i} - \zeta_{i,j}\right) - \Psi'\left(\sum_{k=1}^{V}\zeta_{i,k}\right)\sum_{k=1}^{V}\left(\tau_k + \sum_{d=1}^{D}\sum_{n=1}^{N^d} w^d_{n,k}\phi^d_{n,i} - \zeta_{i,k}\right). \tag{33}$$

Setting this derivative to zero in order to get a maximum, we get the solution

$$\zeta_{i,j} = \tau_j + \sum_{d=1}^{D}\sum_{n=1}^{N^d} w^d_{n,j}\phi^d_{n,i}. \tag{34}$$

2.5 Optimizing w.r.t. $\xi_{c,l}^r$

Collecting only the terms in the bound that contain ξ gives

$$\mathcal{L}_{[\xi]} = \sum_{r=1}^{R}\sum_{c=1}^{C}\sum_{l=1}^{C} \Psi(\xi_{c,l}^r)\left(\omega_l + \sum_{d=1}^{D} \lambda_c^d y_l^{d,r} - \xi_{c,l}^r\right) - \sum_{r=1}^{R}\sum_{c=1}^{C}\sum_{l=1}^{C} \Psi\left(\sum_{t=1}^{C} \xi_{c,t}^r\right)\left(\omega_l + \sum_{d=1}^{D} \lambda_c^d y_l^{d,r} - \xi_{c,l}^r\right)$$
$$- \sum_{r=1}^{R}\sum_{c=1}^{C} \log\Gamma\left(\sum_{t=1}^{C} \xi_{c,t}^r\right) + \sum_{r=1}^{R}\sum_{c=1}^{C}\sum_{l=1}^{C} \log\Gamma(\xi_{c,l}^r). \tag{35}$$

Taking derivatives w.r.t. $\xi_{c,l}^r$ gives

$$\frac{\partial \mathcal{L}_{[\xi]}}{\partial \xi_{c,l}^r} = \Psi'(\xi_{c,l}^r)\left(\omega_l + \sum_{d=1}^{D} \lambda_c^d y_l^{d,r} - \xi_{c,l}^r\right) - \Psi'\left(\sum_{t=1}^{C} \xi_{c,t}^r\right)\sum_{t=1}^{C}\left(\omega_t + \sum_{d=1}^{D} \lambda_c^d y_t^{d,r} - \xi_{c,t}^r\right). \tag{36}$$

Setting this derivative to zero in order to get a maximum, we get the solution

$$\xi_{c,l}^r = \omega_l + \sum_{d=1}^{D} \lambda_c^d y_l^{d,r}. \tag{37}$$

2.6 Parameter estimation (M-step)

Given a corpus of D documents labeled by R different annotators, $\mathcal{D} = \{\mathbf{w}^d, \mathbf{y}^d\}_{d=1}^D$, we find maximum likelihood estimates for the class coefficients $\boldsymbol{\eta}$ by maximizing the lower bound on the log-likelihood w.r.t. $\boldsymbol{\eta}$. Collecting only the terms in the bound that contain η_l gives

$$\mathcal{L}_{[\eta_{l,i}]} = \sum_{d=1}^{D}\left(\sum_{l=1}^{C} \lambda_l^d \eta_l^T \bar{\phi}^d - \log \sum_{l=1}^{C}\prod_{j=1}^{N^d}\left(\sum_{i=1}^{K} \phi_{j,i}^d \exp\left(\frac{1}{N^d}\eta_{l,i}\right)\right)\right). \tag{38}$$

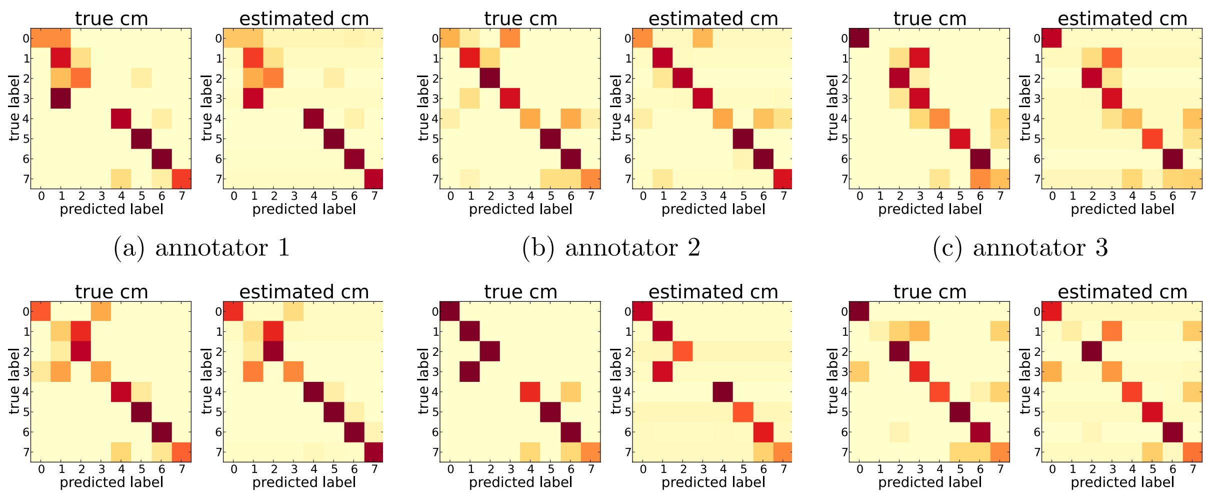
Taking derivatives w.r.t. $\eta_{l,i}$ gives

$$\frac{\partial \mathcal{L}_{[\eta_{l,i}]}}{\partial \eta_{l,i}} = \sum_{d=1}^{D}\left(\lambda_{l,i}^d \bar{\phi}_i^d - \frac{\prod_{n=1}^{N^d}\left(\sum_{i=1}^{K} \phi_{n,i}^d \exp\left(\frac{1}{N^d}\eta_{l,i}\right)\right)}{\sum_{t=1}^{C}\prod_{n=1}^{N^d}\left(\sum_{i=1}^{K} \phi_{n,i}^d \exp\left(\frac{1}{N^d}\eta_{t,i}\right)\right)}\left(\sum_{n=1}^{N^d} \frac{\frac{1}{N^d}\phi_{n,i}^d \exp(\frac{1}{N^d}\eta_{l,i})}{\sum_{j=1}^{K} \phi_{n,j}^d \exp(\frac{1}{N^d}\eta_{l,j})}\right)\right).$$

Setting this derivative to zero does not lead to a closed-form solution, hence a numerical optimization routine (L-BFGS) is used.

3 Additional results on the LabelMe dataset

In order to verify that the proposed model was estimating the (normalized) confusion matrices $\boldsymbol{\pi}^r$ of the different workers correctly, a random sample of them was plotted against the true confusion matrices (i.e. the normalized confusion matrices evaluated against the true labels). Figure 1 shows the results obtained with 60 topics on the LabelMe dataset, where the colour intensity of the cells increases with the magnitude of the value of $p(y^{d,r} = l|c^d) = \pi_{c,l}^r$. These results complement the results presented in the main article for the Reuters-21578 dataset.



(a) annotator 1 (b) annotator 2 (c) annotator 3

(d) annotator 4 (e) annotator 5 (f) annotator 6

Figure 1: True vs. estimated confusion matrix (cm) of 6 different workers of the LabelMe dataset.