How would you train the model?

Log-likelihood optimization:

$$\log \prod_{d \in D} p(d) \prod_{w \in d} p(w|d)^{n_{dw}} \to \max_{\Phi, \Theta}$$

$$\Updownarrow$$

$$\sum_{d \in D} \sum_{w \in d} n_{dw} \log \sum_{t \in T} \phi_{wt} \theta_{td} \to \max_{\Phi, \Theta}$$

Given non-negativity and normalization constraints:

$$\phi_{wt} \geq 0 \qquad \theta_{td} \geq 0$$

$$\sum_{w \in W} \phi_{wt} = 1 \qquad \sum_{t \in T} \theta_{td} = 1$$