Mixture of Boltzmann Machines Input x: a document's bag-of-words encoder $q_{\phi}(\boldsymbol{h}|\boldsymbol{x})$ $|\operatorname{decoder}\, p_{\mathbf{\Theta}^{(1)}}(oldsymbol{x}|oldsymbol{h})|$ \boldsymbol{x} \boldsymbol{h} \boldsymbol{h} $\mathbf{\Theta}^{(1)} = [\mathbf{R}^{(1)}, \mathbf{b}^{(1)}]$ $\log \boldsymbol{lpha}^2$ sampling $g(f_{A_2}(g(f_{A_1}(x))))$ decoder $p_{\mathbf{\Theta}^{(K)}}(oldsymbol{x}|oldsymbol{h})$ $h \sim N(h|\mu, \operatorname{diag}(\alpha))$ \boldsymbol{h} encoder $q_{\eta}(\boldsymbol{c}|\boldsymbol{x})$ \boldsymbol{x} $\log B$ $\boldsymbol{\Theta}^{(K)} = [\boldsymbol{R}^{(K)}, \boldsymbol{b}^{(K)}]$ $igspace p(oldsymbol{x}|oldsymbol{c},oldsymbol{h}) = \prod p_{oldsymbol{\Theta}^{(k)}}(oldsymbol{x}|oldsymbol{h})^{c_k}$ sampling $p_{\mathbf{\Theta}^{(k)}}(\boldsymbol{x}|\boldsymbol{h}) \propto \frac{1}{Z} \exp(-E(\boldsymbol{x};\boldsymbol{h},\mathbf{\Theta}^{(k)}))$ $g(f_{O_2}(g(f_{O_1}(x))))$ $c \sim \text{Multinomial}(\pi)$ $E(\boldsymbol{x};\boldsymbol{h},\boldsymbol{\Theta}^{(k)}) = -\boldsymbol{h}^{\top}\boldsymbol{R}^{(k)}\boldsymbol{x} - (\boldsymbol{b}^{(k)})^{\top}\boldsymbol{x}$