

Scaling up Variational EM

$$\underset{\substack{w, m_1, \dots, m_N \\ s_1, \dots, s_N}}{\text{maximize}} \quad \mathcal{L}(w, q_1, \dots, q_N)$$

$$\text{subject to} \quad q_i(t_i) = \mathcal{N}(m_i, \text{diag}(s_i^2))$$

But this way ~ 100 parameters for each training object