

initialise $q_{\phi}(\boldsymbol{\theta}|\mathbf{x})$ with K components

for $n = 1 \dots N$ **do**

 | sample $\boldsymbol{\theta}_n \sim p(\boldsymbol{\theta})$
 | sample $\mathbf{x}_n \sim p(\mathbf{x}|\boldsymbol{\theta}_n)$

train $q_{\phi}(\boldsymbol{\theta}|\mathbf{x})$ with $\mathcal{L}(\phi) = -\frac{1}{N} \sum_n \log q_{\phi}(\boldsymbol{\theta}_n|\mathbf{x}_n)$

set $\hat{p}(\boldsymbol{\theta}|\mathbf{x} = \mathbf{x}_o) := q_{\phi}(\boldsymbol{\theta}|\mathbf{x}_o)$