Tutoring Session 11

Inference in latent variable models

1 The EM algorithm

Problem 1: Consider a general mixture model with K components. Assume that the distribution of components $p(\mathbf{z})$ is uniform.

Derive a simplified posterior inference (assignment of responsibilities) algorithm for this scenario.

Problem 2: Revisit the slides. Extract an explicit EM algorithm for the Mixture of Gaussians.

Problem 3: Consider a mixture model where the components are given by independent Bernoulli variables. This is useful when modelling, e.g., binary images, where each of the D dimensions of the image $\mathbf{x}$ corresponds to a different pixel that is either black or white. More formally, we have

$$p(\mathbf{x} \mid \mathbf{z} = k) = \prod_{d=1}^{D} \theta_{kd}^{x_d}(1 - \theta_{kd})^{1-x_d}.$$

That is, for a given mixture index $\mathbf{z} = k$, we have a product of independent Bernoullis, where θ_{kd} denotes the Bernoulli parameter for component k at pixel d.

Derive the EM algorithm for the parameters $\theta = \{\theta_{kd} \mid k = 1, \ldots, K, d = 1, \ldots, D\}$ of a mixture of Bernoullis.

2 Variational Auto-Encoders

Problem 4: A crucial trick when implementing Variational Auto-Encoders (VAEs) is the *reparameterisation trick*,

$$\nabla_\phi \, \mathbb{E}_{q(\mathbf{z}|\phi)}[f(\mathbf{z})] = \mathbb{E}_{q(\boldsymbol{\epsilon})}\big[\nabla_\phi \, f(g(\boldsymbol{\epsilon}, \phi))\big].$$

It is based on the fairly weak assumptions that

- f is differentiable,
- and we can draw samples from the distribution of $q(\mathbf{z} \mid \phi)$ by first drawing a sample $\boldsymbol{\epsilon} \sim q(\boldsymbol{\epsilon})$ and then applying a deterministic, differentiable transform $g(\boldsymbol{\epsilon}, \phi)$.

Prove that this trick is valid. Why is it so important?

Problem 5: Prove

$$\begin{aligned}
\mathcal{L}_{\text{ELBO}}(\phi, \theta) &= \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\ln p(\mathbf{x} \mid \mathbf{z}, \theta)] - \mathrm{KL}\big(q_\phi(\mathbf{z} \mid \mathbf{x}) \,\|\, p(\mathbf{z})\big) && (1)\\
&= \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\ln p(\mathbf{x}, \mathbf{z} \mid \theta)] + \mathbb{H}\left(q_\phi(\mathbf{z} \mid \mathbf{x})\right) && (2)\\
&= \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\big[\ln p(\mathbf{x}, \mathbf{z} \mid \theta) - \ln q_\phi(\mathbf{z} \mid \mathbf{x})\big]. && (3)
\end{aligned}$$

$\mathbb{H}(p(\mathbf{x})) = \mathbb{E}_{p(\mathbf{x})}[-\ln p(\mathbf{x})]$ denotes the entropy.

All three variants are found in practice when training variational inference models. Can you guess reasons?

Problem 6: Remember the results of a VAE on MNIST, cf. Figure 1.

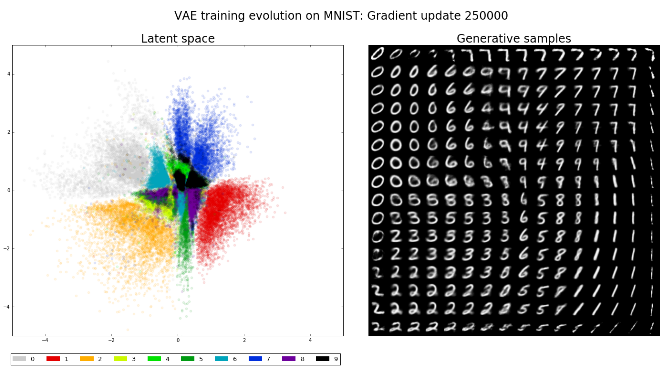

Figure 1: Left: Latent space samples from the approximate posterior $q_\phi(\mathbf{z} \mid \mathbf{x}_n)$ for $\mathbf{x}_n \in \mathcal{D}$. Right: Samples from the generative model with latent samples $\mathbf{z}$ randomly drawn from the prior.

Answer the following questions:

- Does the total distribution of the latent samples remind you of something?
- Can you justify your claim from the previous question?
- The latent space looks nicely ordered. Can you identify problems?