Small Exercises 11

Inference in latent variable models

Problem 1: Explain the relation between training via the EM algorithm and training by variational auto-encoding in 140 characters or less.

> EM explicitly minimizes the KL between posterior and approximate posterior, which implicitly maximizes the ELBO. The VAE does the opposite.
>
> (This last sentence has 139 characters.)

Problem 2: Explain why the posterior distribution of the latent variables given the observed data is of such high interest in machine learning.

> The latent variables can be interpreted as the core information that generates the data. Any data processing is about extracting (parts of) this information from raw data. The posterior distribution formalizes this concept, which allows us to develop theoretically grounded algorithms.

Problem 3: EM and variational auto-encoding are iterative training algorithms. Name problems that are caused by such procedures.

> Initialisation matters a lot. Both methods, depending on the data, can get stuck in local minima. As all point estimates of parameters, there is a risk of overfitting—although the inbuilt regularisation eases some of that risk.

Problem 4: What is the qualitative difference between the latent space when choosing (i) a mixture of Gaussians and a (ii) variational auto-encoder for modelling the MNIST data set as in the slides?

> This is a very open question, but probably the most important difference is that the variational auto-encoder does not impose much structure on how information is stored in the latent variables, whereas the approach with mixture of Gaussians heavily depends on our assumptions that the driving latent factor are 10 different classes (in our case digits, which is why we chose a mixture of 10 Gaussians).