Mathematical Underpinnings Lab 8 17.04.2024, 24.04.2024

Task 1 (VAE)

Load the data: `(x_train, y_train), (x_test, y_test) = mnist.load_data()`.
Next, scale the data so the value of each pixel is in $[0, 1]$ (and not $[0, 255]$). Make a plot of the first digit in the dataset. Transform the matrices in which each image of a digit is stored into vectors. The length of that vector will be denoted as `original_dim` and in our case be equal to $28 \cdot 28$.

a) Define a function computing ELBO. It should contain two parts: the expectation of $p(x|z)$ according to q (input: a vector of observed pixels and posterior probability of a pixel being black after reparametrization trick) and Kullback-Leibler divergence of q and p (input: the means and log-variances of the latent variable z).

b) Define sampling function using the reparametrization trick.

c) Define the encoder (input: a vector of pixels, output: a vector with concatenated mean and log-variance of the hidden variable z).

d) Define the decoder (input: a vector of length that is equal to latent space dimension, output: a vector of `original_dim` length with numbers in $[0, 1]$).

e) Combine the encoder, sampling, decoder and the risk function and fit the weights using training data.

f) Visualise the results - compare the true digits with artificially generated using test data. For artificially generated image, take a true image, then find its representation in terms of a latent space, decode the representation using decoder, and then transform the result into matrix and draw it.

g) Find latent representation of the digits from the test dataset. Next, draw a scatterplot (if your latent space has more dimensions than 2 use an algorithm which will map it into 2 dimensions). Color the points according to the digit they represent.

h) Generate new samples (new images of digits) from your trained VAE. To get new samples you have to sample from the prior distribution $p(z)$ and then from the likelihood $p(x|z)$.

Task 2 (cVAE)

Implement conditional VAE. Make appropriate changes in the implementation of VAE.
The changes and what does not change:

- The input to the model is now a vector of pixels x concatenated with class information c (the digit).
- The encoder $q(z|x, c)$ now uses the information about the class. The prior of the latent space doesn't change and is standard normal.
- The loss function (ELBO) doesn't change.
- The decoder takes as input a vector z concatenated with c (to enable conditional sampling).

Next, do f), g) and h). In h) first sample from $p(z)$ and then from $p(x|z, c)$. Add the information about the class to the decoder.

Task 3 (optional)

Experiment with the parameters, encoder's and decoder's architectures, sample new images 'around' an image from the training data, visualise the transition from one digit to another (by choosing appropriate path in the latent representation) etc.

Simple yet comprehensive description of VAE is here.