**variational autoencoder (VAE)**

A natural next step from the autoencoder is the variational autoencoder. Variational autoencoders was first proposed by Kingma and Welling [21], which suggested a probabilistic model over a latent space using an encoder-decoder structure, which allows us to sample from the autoencoder. A variational autoencoder is similar to an autoencoder since it utilizes an encoder-decoder structure with a latent space. but the variational autoencoder has some fundamental differences; a VAE is a probabilistic model. Additionally, the encoding distribution is regularized during training to guarantee that the properties of the latent space is suitable to generate new data. Simply put, the VAE learns the parameters of the distribution of the input data, which implies that it is possible to generate new data from the learned distribution. The encoder predicts the mean and standard deviation of the input, which means it can output a low-dimensional representation of the normal distribution of the latent space [18, 21].

The probabilistic feature and the variational inference of the VAE requires further loss calculations than a typical loss function, since the encoder output (latent space representation) can be considered a latent random variable Z makes the variational inference problem a Bayesian inference problem. The main issue of this problem is that the density p(x) =∫𝑝(𝑥|𝑧) 𝑝(𝑧) d𝑧 is normally intractable. The solution is to approximate 𝑝(z|x) with a tractable density q(z), so that 𝑝(z|x) ≈ 𝑞(𝑧). Using Jensen’s inequality which states that 𝜙(𝔼[X]) ≤ 𝔼[𝜙(X)] for a convex function 𝜙, we get the following solution:

A group of math equations

Description automatically generated

By maximizing the evidence lower bound (ELBO), we minimize the loss of the approximation. The ELBO consists of two parts: KL divergence and a reproduction loss. The reproduction loss makes sure the reconstructed inputs are as close as possible to the original image, and the KL divergence quantifies how one probability distribution diverges from another. In other words, the KL divergence minimizes the “distance” between the variational posterior and the prior distribution [18, nadli).

With the variational inference problem solved, the encoder can be seen as f(x) = q(z|x) and the decoder can be seen as g(z) = p(x|z). However, to use the KL-divergence, we need an estimate of p(z) that is comparable to q(z), which means the latent space representation z must be probabilistic. To ensure that the latent space representation z must be probabilistic, it is possible to sample a standardized random variable from the distribution and scale it utilizing expected parametrization (nadli – class notes). The model can then output parameters for the mean and standard deviation for each input during training. This can then be sampled from by sampling normal random noise 𝜖, so that 𝜖∼ N (0, 1) and making the latent space representation z\_i = mean\_i + 𝜖\_i \* std\_i ∼ N (mean, var). 𝑧\_𝑖 =𝜇\_𝑖 + 𝜖⋅𝜎\_𝑖 ∼N(𝜇,𝜎^2). [18, nadli]

Some advantages of VAEs are that they are easy to train, since it has one tractable likelihood loss and that they manage to produce diverse samples since the likelihood maximization makes sure to capture all modes of the input. However, VAEs falls short in terms of generation quality and reliability because the predicted distributions of the latent space can be very similar for multiple samples, meaning the generation will be an average of the original samples, in other words blurry generation. Additionally, the pixel-based loss is not transferable to the latent space, which makes the model predict pixel-averages, resulting in blurry generations [21]. The pixel loss, or reconstruction loss, is typically MSE which gives blurred images [18]

A diagram of a sample

Description automatically generated

Figure 1. <https://pub.towardsai.net/diffusion-models-vs-gans-vs-vaes-comparison-of-deep-generative-models-67ab93e0d9ae>

A diagram of a process

Description automatically generated

Figure 2. <https://medium.com/@judyyes10/generate-images-using-variational-autoencoder-vae-4d429d9bdb5>

A diagram of a process flow

Description automatically generated

Figure 3. <https://medium.com/@rushikesh.shende/autoencoders-variational-autoencoders-vae-and-%CE%B2-vae-ceba9998773d>