VSB TECHNICAL UNIVERSITY OF OSTRAVA | FACULTY OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE | DEPARTMENT OF COMPUTER SCIENCE

Deep Learning

Variational Autoencoders

Jan Platoš, Radek Svoboda

March 24, 2024

Department of Computer Science
Faculty of Electrical Engineering and Computer Science
VŠB - Technical University of Ostrava

Variational Autoencoders

Variational Autoencoders

- The autoencoder itself works well in a task of reconstruction of the input.

Variational Autoencoders

- The autoencoder itself works well in a task of reconstruction of the input.
- What about encoder?
  - Encoder produces a compressed representation of the input.
  - The size of the output is defined by the model.
  - The representation follows the needs of the encoder.

Variational Autoencoders

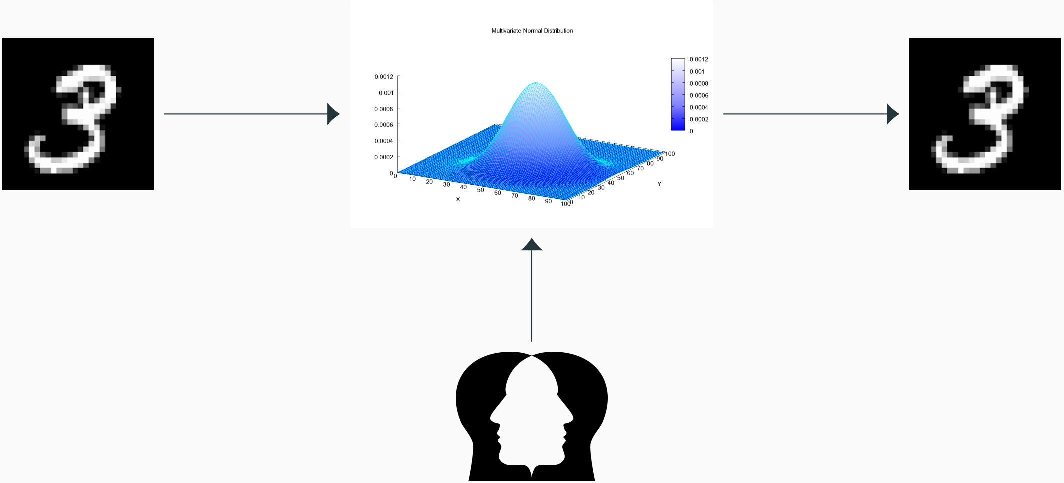
- The autoencoder itself works well in a task of reconstruction of the input.
- What about encoder?
    - Encoder produces a compressed representation of the input.
    - The size of the output is defined by the model.
    - The representation follows the needs of the encoder.
- What about decoder?
    - Decoder decodes the encoded representation into an original form (with some modifications).
    - The representation is defined as an ”agreement” between encoder and decoder.
    - May it be used for generation of new objects?

Variational Autoencoders

- The representation produced by encoder is too specific to be directly generated.
- The encoded representation is very sparse and discreet.
- A major modification need to be designed[1]
- The encoder process tries to create a latent representation that contains all the information that are needed to reconstruct it.
- Latent representation may be imagined as the object, thickness of the line, color, etc...
- When we have these parameters, we may reconstruct the originals.

[1] Kingma DP, Welling M. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114. 2013 Dec 20.

Variational Autoencoders - Idea



Variational Autoencoders

- VAE model the process of the latent representation generation and the reconstruction using the Law of Total Probability:

$$P(x) = \int P(x|z)P(z)dz$$

- z is a candidate latent vector.
- $P(x|z)$ represents the probability that input x may be generated using the z.
- $P(z)$ represents the probability that the z exists in the latent space.

- VAE training objective is to maximize $P(x)$.
- $P(x|z)$ is modeled as a multi-variete Gaussian $\mathcal{N}\left(f\left(x\right), \sigma^2 \cdot I\right)$
- The f is what is modeled using the Neural network.
- The σ is a hyper-parameter.

- The latent space itself is very complex and hard to define or determine.
- Interpretation of each dimension is very hard.
- Latent dimension may be correlated.
- Reconstruction from the latent space is also very complex task.
- We need to define the function that maps from latent space into $P(x|z)$.

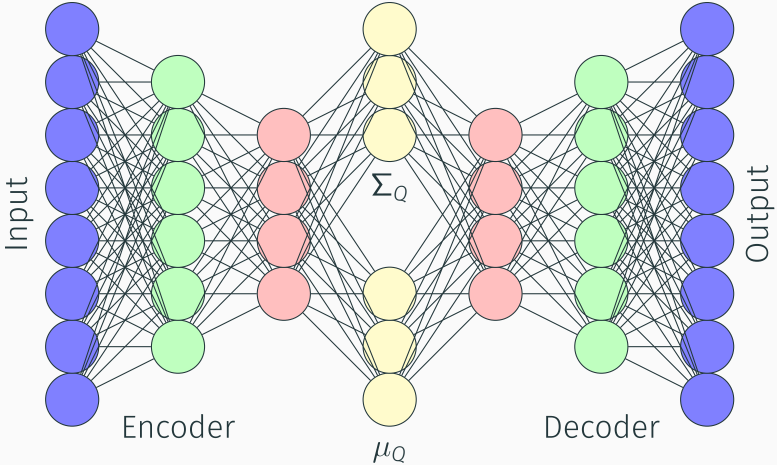
Variational Autoencoders

- Lets choose $P(z)$ to be standard multivariete Gaussian.
- Deep learning/Neural networks solves the problem of finding f using the following decomposition:

1. Defines the encoder that maps the Gaussian to the true distribution over latent space.
2. Defines the decoder that map the latent space to $P(x|z)$.

- It is difficult to derive the latent representation directly even using NN - not all samples has meaning to $P(x)$.
- We may substitute it with a different distribution $Q(z, x)$ that increases the likelihood of usability of z.
- Solution is to replace latent space with something that is easier, e.g. parameterless Gaussian.
- The input is the summarized using the mean μ_Q and diagonal covariance matrix Σ_Q - these are the encoded parameters.

Variotional Autoencoder



Variational Autoencoder

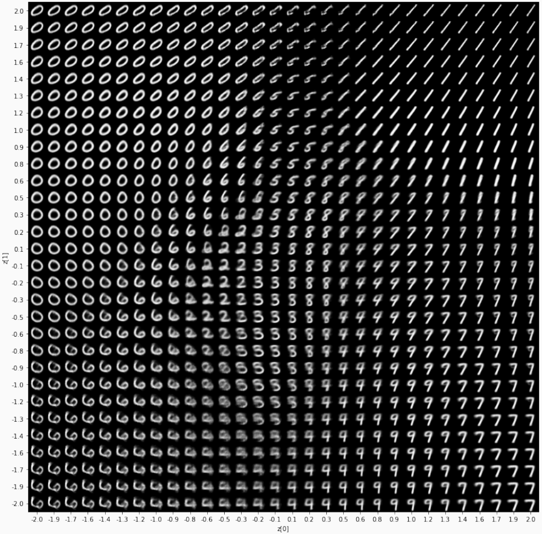
1. The input image is passed through an encoder network.
2. The encoder outputs parameters μ_Q and Σ_Q of the distribution $Q(z|x)$.
3. The latent vector z is sampled from $Q(z|x)$.
4. The decoder decodes the z into an image.
5. The loss function depends not only on the pixel reconstruction but also on the distribution learned (using Kullback–Leibler divergence).

$$KLd = \frac{1}{2}\sum_{i=0}^{n} \sigma_i^2 + \mu_i^2 - log(\sigma_i) - 1$$

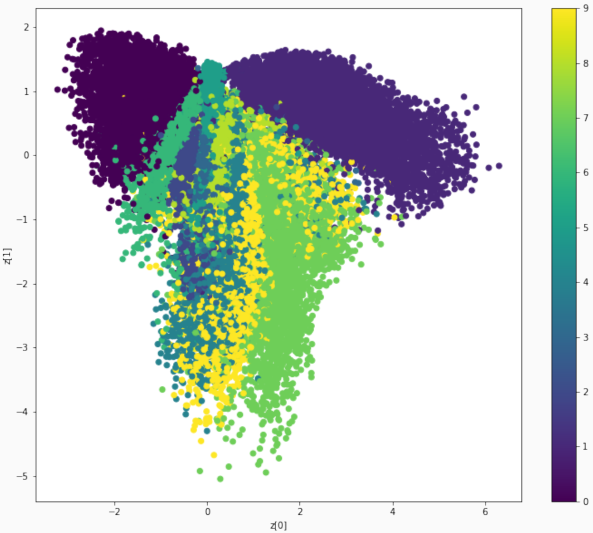
Variational Autoencoder

- The VAE learn how to set the distribution properly.
- The mean values should group similar objects and spread different objects across the latent space.
- The encoder complexity as well as decoder complexity may be large.
- Again, any type of network may be used (Dense, CNN, ...).

Variational Autoencoder



Variational Autoencoder



Questions?