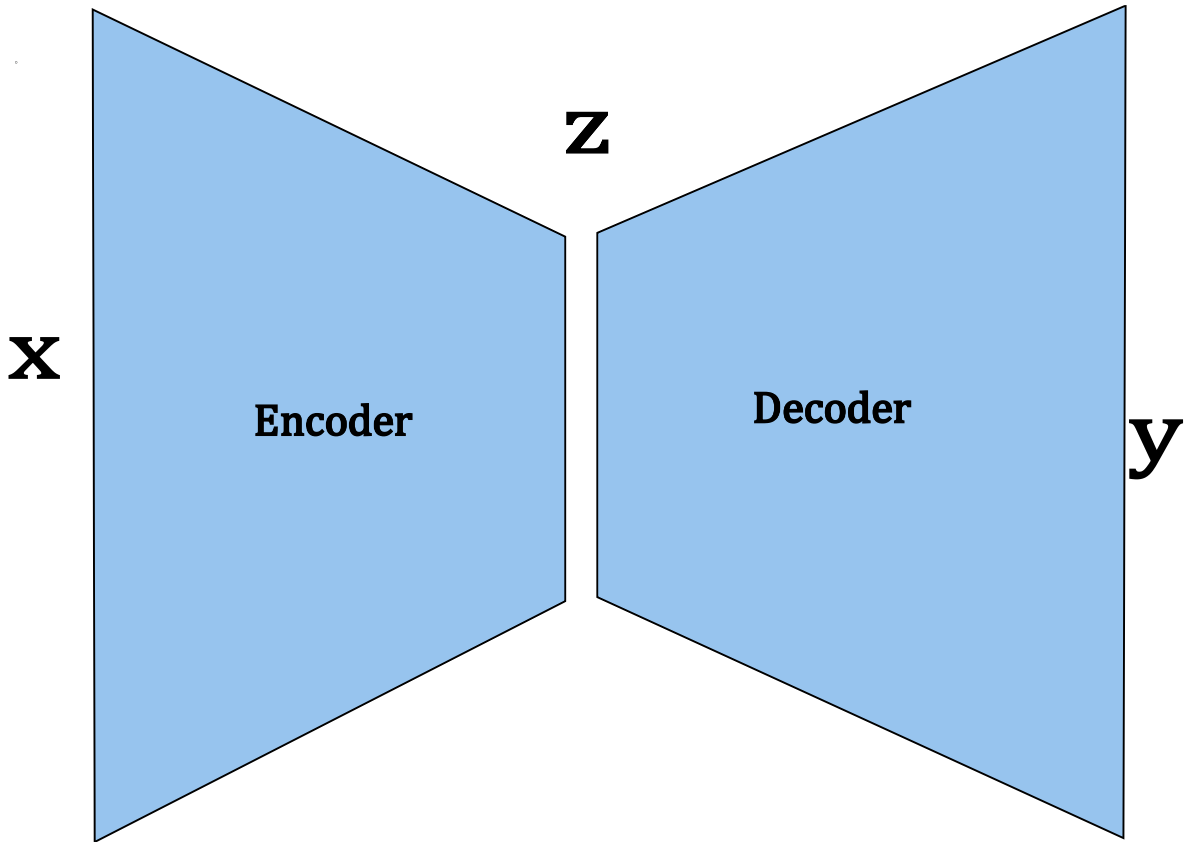
AutoEncoders

MUSTAFA HAJIJ

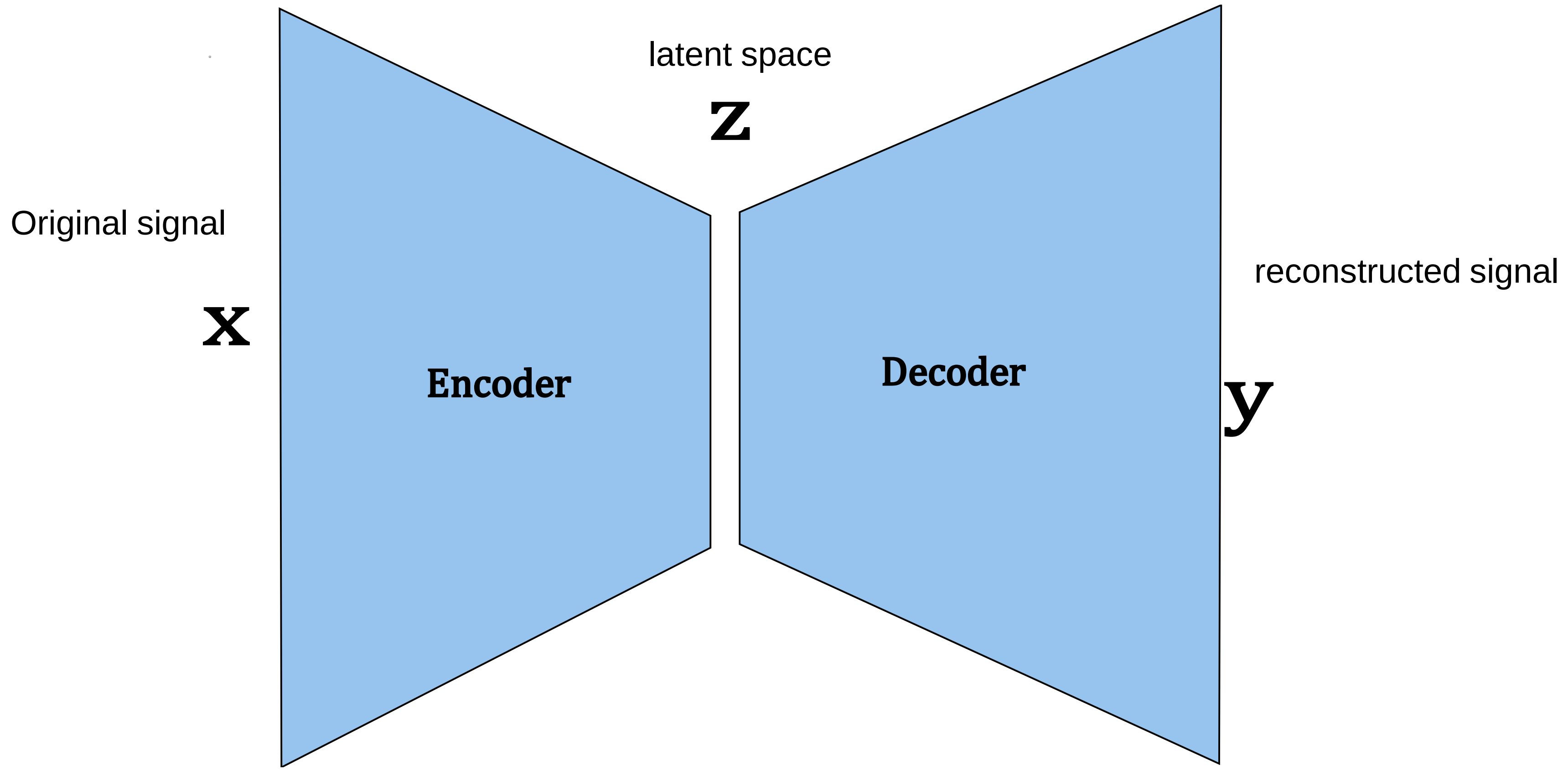
AutoEncoder

Autoencoders are special neural networks that try to construct a signal x by passing it through a bottleneck z. The final reconstruction signal is denoted by y



AutoEncoder

Autoencoders are special neural networks that try to construct a signal x by passing it through a bottleneck z. The final reconstruction signal is denoted by y

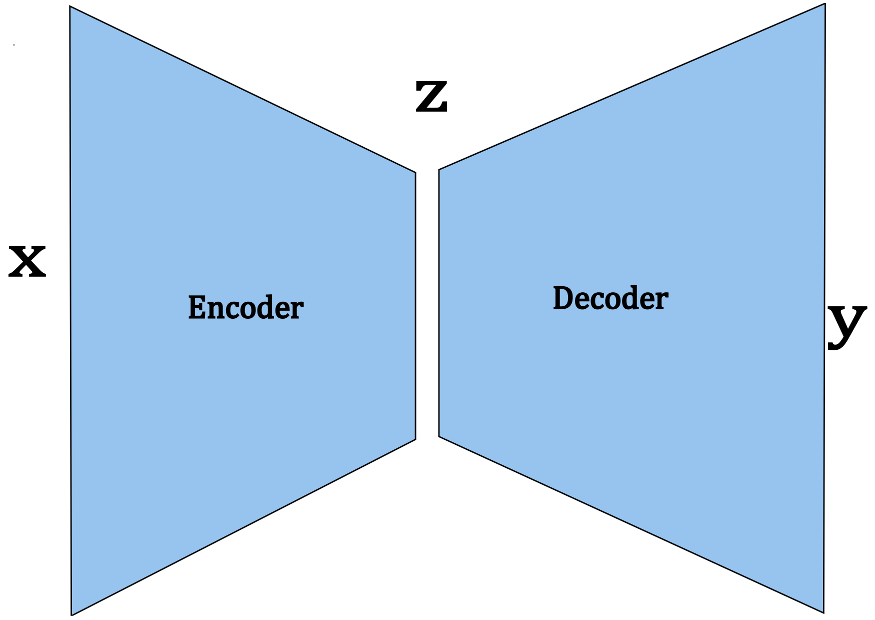


Mathematically, an autoencoder is a tuple of NNs (Encoder, Decoder) such that y=Decoder(Encoder(x))=x. In practice we realize this equation as an an MSE loss |Decoder(Encoder(x))-y| and train the tuple (Encoder, Decoder) to minimize that loss.

AutoEncoders and PCA

In the special case, an autoencoder can give us Principal Component Analysis (PCA)-like behavior when certain conditions are met.

An autoencoder is a type of neural network that learns to encode and decode data. It consists of an encoder network that compresses the input data into a lower-dimensional representation and a decoder network that reconstructs the original data from the encoded representation.

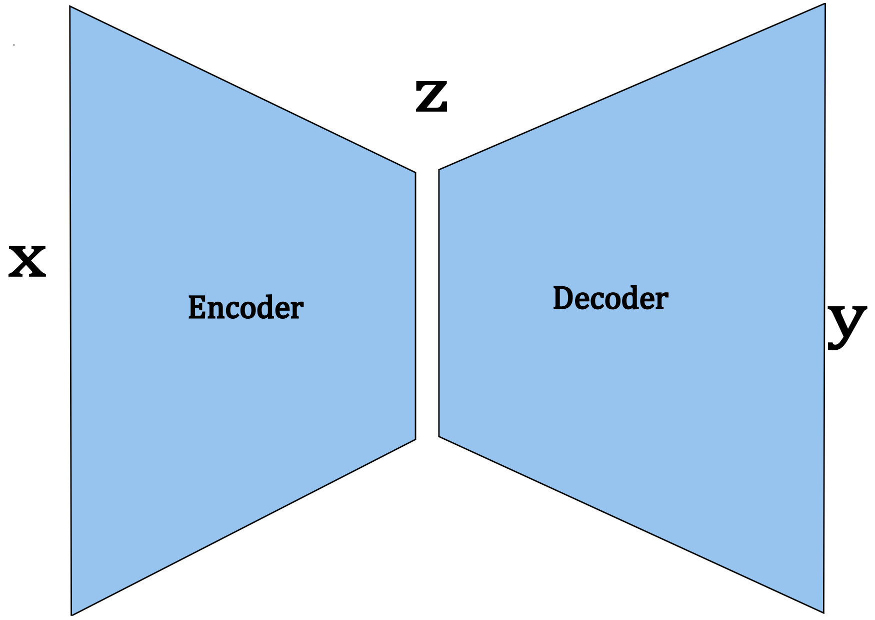


AutoEncoders and PCA

In the special case, an autoencoder can give us Principal Component Analysis (PCA)-like behavior when certain conditions are met.

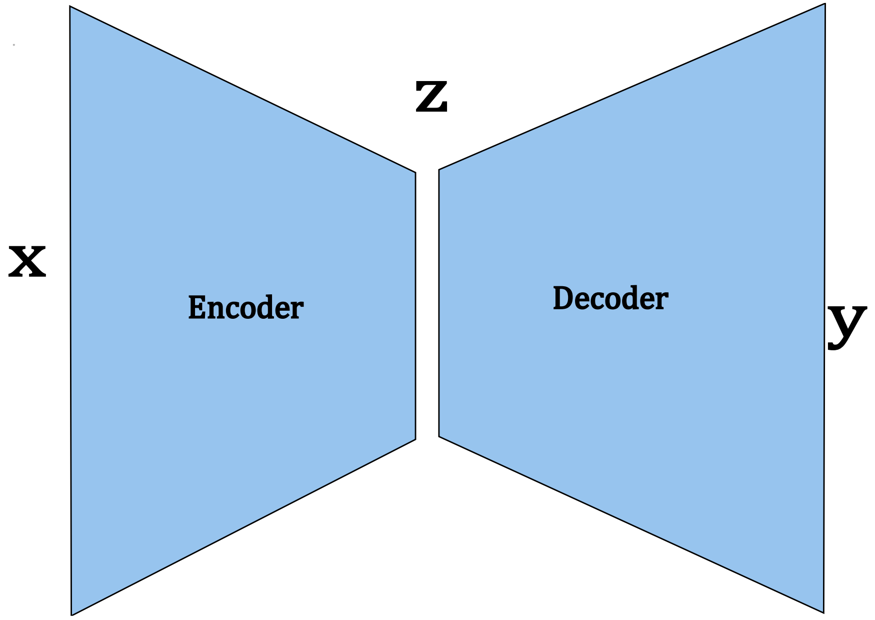
An autoencoder is a type of neural network that learns to encode and decode data. It consists of an encoder network that compresses the input data into a lower-dimensional representation and a decoder network that reconstructs the original data from the encoded representation.

The encoder part of the autoencoder can learn to project the input data onto a lower-dimensional subspace, capturing the most important features or patterns. The decoder part then reconstructs the original data from the encoded representation. If the reconstruction loss is minimized, the autoencoder tries to recreate the original input as accurately as possible.



AutoEncoders and PCA

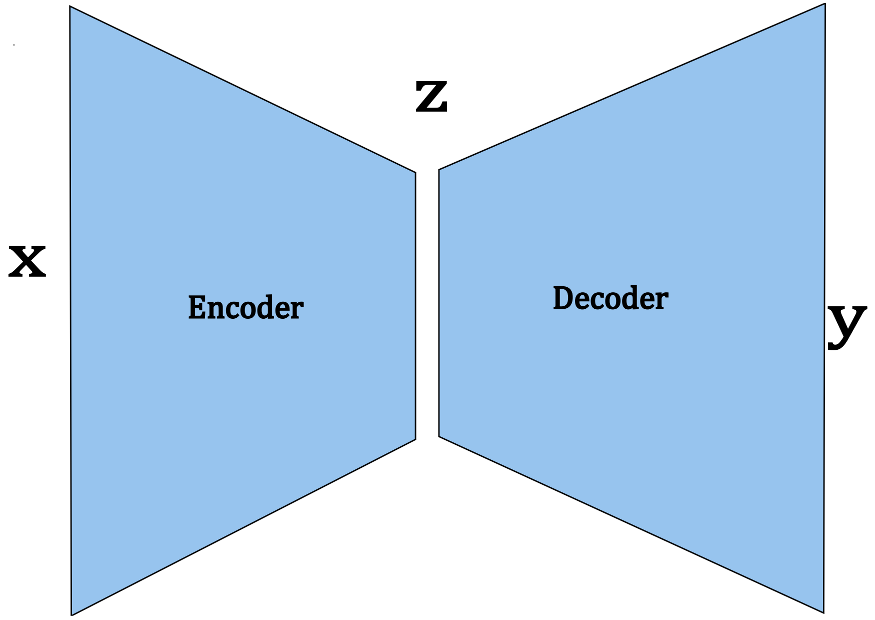
The encoder part of the autoencoder can learn to project the input data onto a lower-dimensional subspace, capturing the most important features or patterns. The decoder part then reconstructs the original data from the encoded representation. If the reconstruction loss is minimized, the autoencoder tries to recreate the original input as accurately as possible.



AutoEncoders and PCA

The encoder part of the autoencoder can learn to project the input data onto a lower-dimensional subspace, capturing the most important features or patterns. The decoder part then reconstructs the original data from the encoded representation. If the reconstruction loss is minimized, the autoencoder tries to recreate the original input as accurately as possible.

This process encourages the autoencoder to find a low-dimensional representation of the data that preserves the most important information, similar to how PCA finds the directions of maximum variance. The latent space of the autoencoder can be considered as the principal components, and the encoder weights can be interpreted as the loading vectors of the principal components.



Variational Autoencoders

$AutoEncoders \qquad x \rightarrow z = e(x) \rightarrow y = d(z)$

Image source

Variational Autoencoders

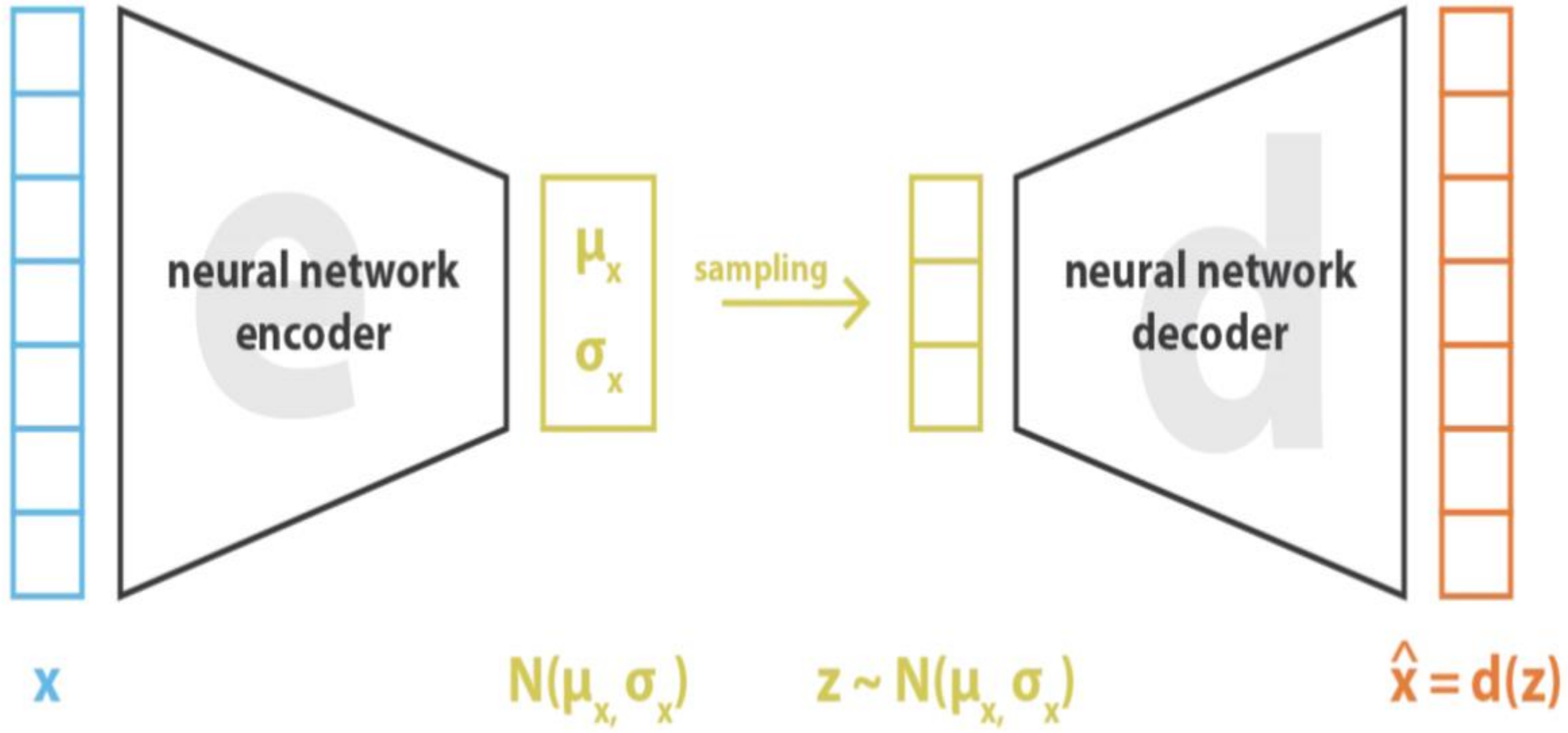
$AutoEncoders \qquad x \rightarrow z = e(x) \rightarrow y = d(z)$

Rather than encoding the input as a single point, we encode it as a distribution over the latent space.

$$Variational\ AutoEncoders \qquad x \rightarrow \underbrace{p(z|x)}_{latent\ distribution} \rightarrow \underbrace{z \sim p(z|x)}_{sample\ from\ latent\ distribution} \rightarrow \underbrace{d(z)}_{reconstruction}$$

Image source

Variational Autoencoders



$$\text{loss} = \| x - \hat{x} \|^2 + KL[N(\mu_x, \sigma_x), N(0, I)] = \| x - d(z) \|^2 + KL[N(\mu_x, \sigma_x), N(0, I)]$$

Image source

Variational Autoencoders

1.Encoding Variational Distribution: The VAE framework assumes that the encoder network learns to map the input data to a distribution in the latent space. This distribution is usually assumed to be a multivariate Gaussian with a mean and a diagonal covariance matrix.

Variational Autoencoders

1. Encoding Variational Distribution: The VAE framework assumes that the encoder network learns to map the input data to a distribution in the latent space. This distribution is usually assumed to be a multivariate Gaussian with a mean and a diagonal covariance matrix.

2. Sampling from the Latent Space: During the training process, VAEs introduce a sampling step to generate a latent variable (sample) from the learned distribution in the latent space. This sample is used as an input to the decoder network for reconstructing the data.

Variational Autoencoders

1. Encoding Variational Distribution: The VAE framework assumes that the encoder network learns to map the input data to a distribution in the latent space. This distribution is usually assumed to be a multivariate Gaussian with a mean and a diagonal covariance matrix.

2. Sampling from the Latent Space: During the training process, VAEs introduce a sampling step to generate a latent variable (sample) from the learned distribution in the latent space. This sample is used as an input to the decoder network for reconstructing the data.

3. Learning the Latent Space Distribution: The goal of the VAE is to learn the parameters (mean and covariance) of the distribution in the latent space that best represents the training data. This is achieved by minimizing the reconstruction error (reconstruction loss) and simultaneously regularizing the distribution in the latent space using the KL loss.

Variational Autoencoders

1.Encoding Variational Distribution: The VAE framework assumes that the encoder network learns to map the input data to a distribution in the latent space. This distribution is usually assumed to be a multivariate Gaussian with a mean and a diagonal covariance matrix.

2.Sampling from the Latent Space: During the training process, VAEs introduce a sampling step to generate a latent variable (sample) from the learned distribution in the latent space. This sample is used as an input to the decoder network for reconstructing the data.

3.Learning the Latent Space Distribution: The goal of the VAE is to learn the parameters (mean and covariance) of the distribution in the latent space that best represents the training data. This is achieved by minimizing the reconstruction error (reconstruction loss) and simultaneously regularizing the distribution in the latent space using the KL loss.

4.Regularizing Latent Space Distribution: The KL loss serves as a regularization term that encourages the learned distribution in the latent space to resemble a desired latent distribution, which is usually a standard Gaussian distribution (zero mean and unit variance). It measures the divergence (difference) between the learned distribution and the latent distribution.

Variational Autoencoders

1.Encoding Variational Distribution: The VAE framework assumes that the encoder network learns to map the input data to a distribution in the latent space. This distribution is usually assumed to be a multivariate Gaussian with a mean and a diagonal covariance matrix.

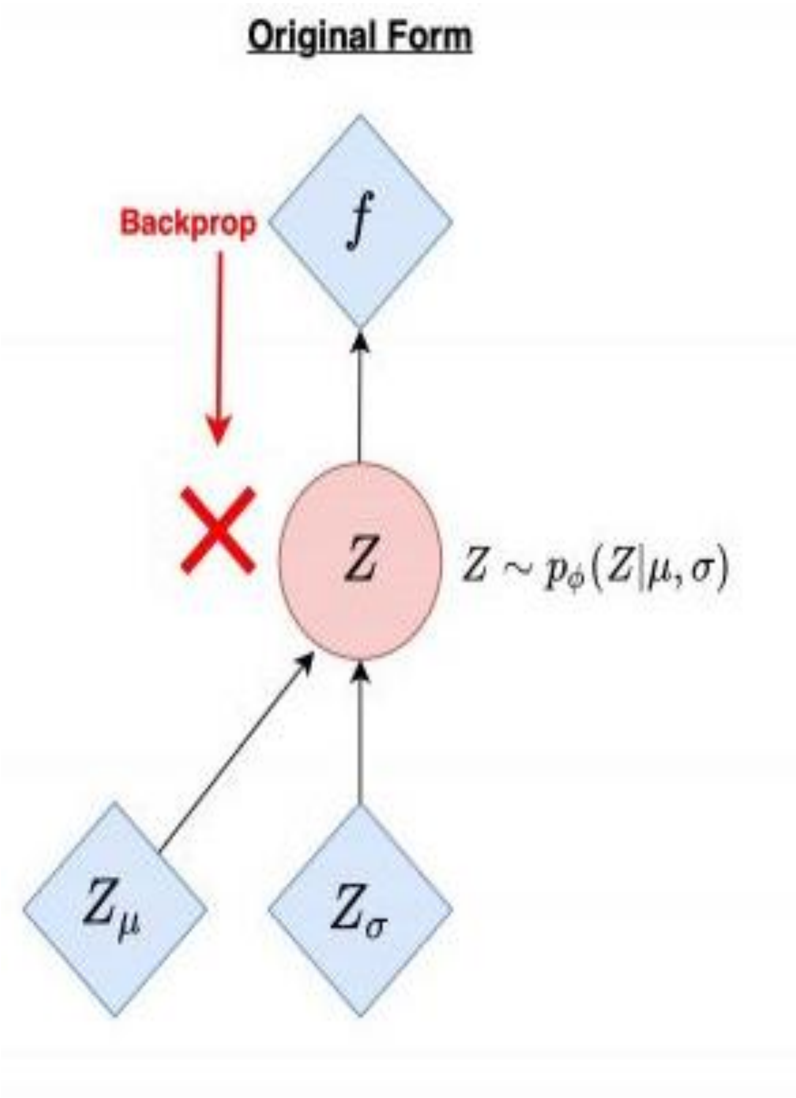
2.Sampling from the Latent Space: During the training process, VAEs introduce a sampling step to generate a latent variable (sample) from the learned distribution in the latent space. This sample is used as an input to the decoder network for reconstructing the data.

3.Learning the Latent Space Distribution: The goal of the VAE is to learn the parameters (mean and covariance) of the distribution in the latent space that best represents the training data. This is achieved by minimizing the reconstruction error (reconstruction loss) and simultaneously regularizing the distribution in the latent space using the KL loss.

4.Regularizing Latent Space Distribution: The KL loss serves as a regularization term that encourages the learned distribution in the latent space to resemble a desired latent distribution, which is usually a standard Gaussian distribution (zero mean and unit variance). It measures the divergence (difference) between the learned distribution and the latent distribution.

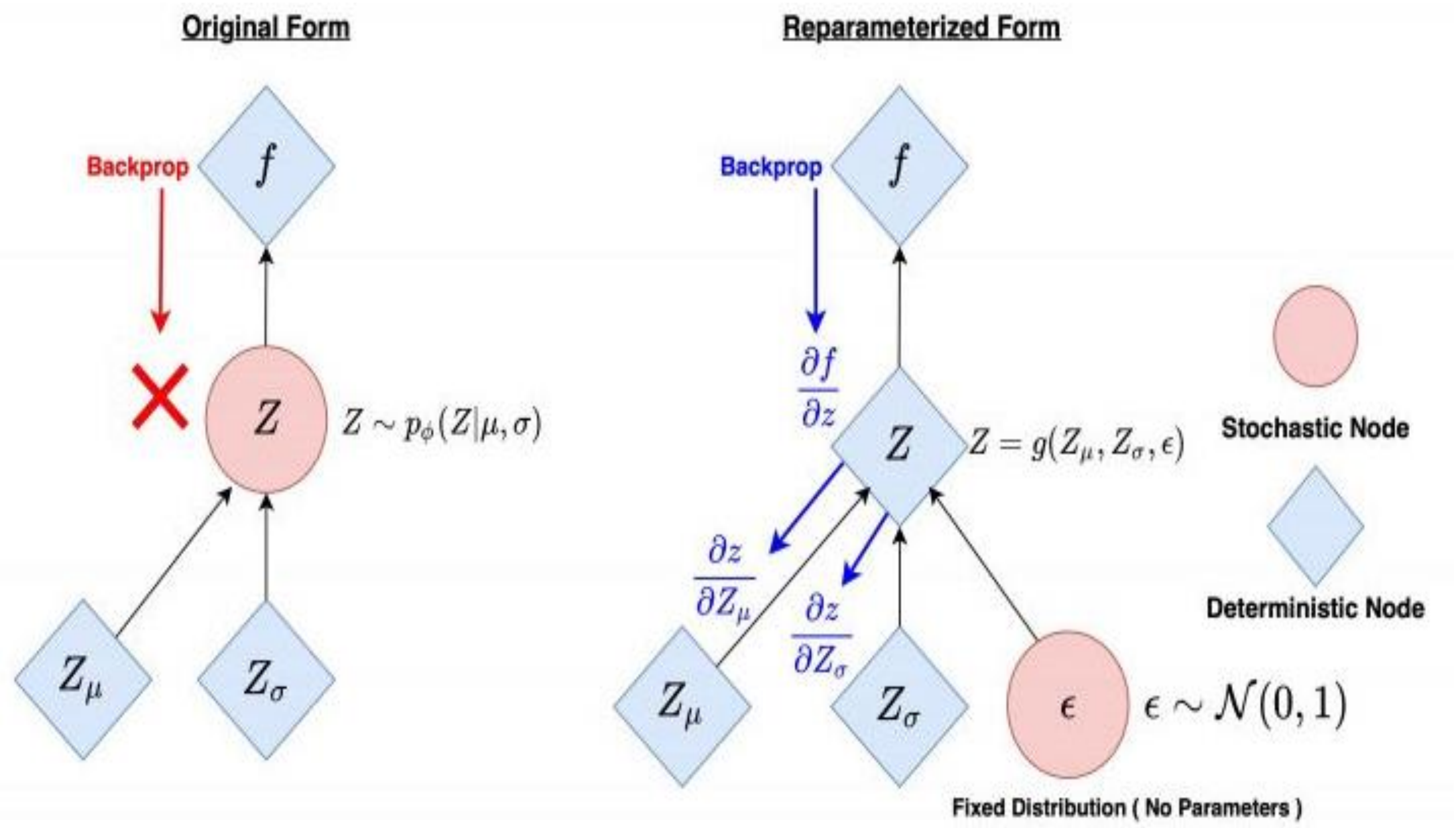
5.Balancing Reconstruction and Regularization: By including the KL loss in the training objective, the VAE strikes a balance between accurate data reconstruction (minimizing reconstruction loss) and ensuring the learned distribution in the latent space follows the desired prior distribution (minimizing KL loss).

Variational Autoencoders



[source]

Variational Autoencoders



$$Z = Z_\mu + Z_\sigma^2 \odot \varepsilon$$

Here, $\varepsilon \sim \mathcal{N}(0,1)$ and $\odot$ is element-wise multiplication.

source

Refs

[Variational Autoencoder in TensorFlow (Python Code) (learnopencv.com)]