By Shreyank Banakar

Variational Auto Encoders
Vae are used to generate outputs which are similar to the inputs
this is achieved with the help of an encoder and a decoder
to generate an out we need to specify a latent variable which helps in the creation of a specific input let it be z.

this latent variable is then used as an input to the functon f(z;v) where v is a learnable parameter, which produces an output similar to input, we generally prefer it to be a gaussian distribution which has mean f(z;v) and variance sigma*I I(identity matrix).

we sample z from a distribution in which each dimension is a gaussian distribution with mean 0 and sd 1

but sampling z from such distribution is difficult we would like to have a funtion q(z/x) which provides the output(desired gaussian form) basing on the input x which helps in giving
a better set of values of z which depend on x
this functon q helps us in sampling z from a smaller set of values of z which helps us as it removes the difficulty of p(x/z) being 0 most of the time and reducing computation difficulty

the above step is the encoding part which provides us with q(z/x) (preferably gaussian) which outputs us with 1)mean 2)variance(a covariance matrix)
this is then used to sample(gaussian distr.) and then used as an input to f(z) to produce the output

the main difficulty here is that during backprop the loss cannot travel back to the encoder part as we are sampling from a gaussian distribtuion
this problem is solved with the help of reparameterization

here we take the output from the encoder(mean, covariane matrix) , then we sample from a standard normal distribution N(0,1) (let the variable be r), then we provide the input to the decoder as r*variance(from encoder)+mean(from encoder)

by the above step of reparmeterization trick we can form a direct link between the encoder and the decoder which can be used during backprop

now as the output is gaussian T(f(z;v),sigma*I)()where sigma is a hyperparameter the negative log likelihood is dircetly proportional to the square of the distance between f(z) and
x but as in the example given in the paper we need to keep the sigma very small so as to make the output more similar to input