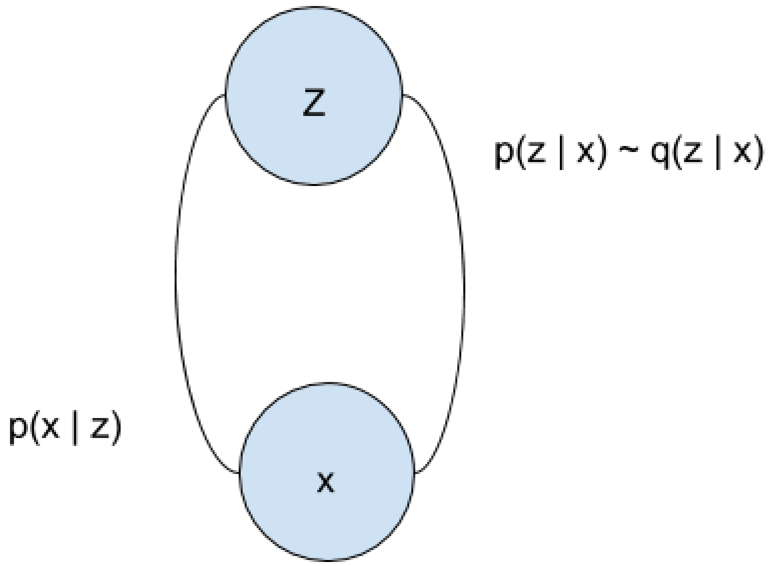
Using variational autoencoders, our goal is to generate the input with the least error possible.



z is the hidden or latent space and **x** is the input. $p(x \mid z)$ is the posterior probability of x given z .

The problem arises in calculation of $p(z|x) \ = \ p(x,z) \,/\, p(x)$ where $p(x) \ = \ \int_z p(x \mid z).\, p(z) \,.\, dz$.

Estimation of $p(x)$ can be intractable if z is high dimensional, which usually is the case. To overcome this, we try to estimate the unknown $p(z \mid x)$ distribution using a known tractable distribution $q(z \mid x)$ like a normal distribution.

For making the $q(z \mid x)$ as close as possible to the $p(z \mid x)$ distribution we want to minimize the KL divergence of q with respect to p

Calculating the lower bound

$$
\begin{aligned}
KL(\, q(z \mid x) \,||\, p(z \mid x)\,) \quad &= \ -\sum_z q(z|x) \,.\, log\, \frac{p(z|x)}{q(z|x)} \\
&= \ -\sum_z q(z|x) \,.\, log\, \frac{p(x,z)}{q(z|x)} . \frac{1}{p(x)} \\
&= \ -\sum_z q(z|x) \,.[\, log\, \frac{p(x,z)}{q(z|x)} \ + log . \frac{1}{p(x)} \ ] \\
&= \ -\sum_z q(z|x) \,.[\, log\, \frac{p(x,z)}{q(z|x)} \ - log\, p(x) \,] \\
&= \ -\sum_z q(z|x) \,.\, log\, \frac{p(x,z)}{q(z|x)} \ + \sum_z q(z|x) \,.log\, p(x) \,] \\
&= \ -\sum_z q(z|x) \,.\, log\, \frac{p(x,z)}{q(z|x)} \ + log\, p(x) . \sum_z q(z|x) \ \ ] \\
&= \ -\sum_z q(z|x) \,.\, log\, \frac{p(x,z)}{q(z|x)} \ + log\, p(x).\ 1 \ \ ] \quad \text{as } \sum_z q(z|x) \ = \ 1
\end{aligned}
$$

$$KL(\ q(z \mid x) \parallel p(z \mid x)\ ) \ = \ -\sum_z q(z|x) \ . \ log \ \frac{p(x,z)}{q(z|x)} \ + \ log\ p(x)$$

$$log\ p(x) \ = KL(\ q(z \mid x) \parallel p(z \mid x)\ ) \ + \sum_z q(z|x) \ . \ log \ \frac{p(x,z)}{q(z|x)}$$

We know $KL \geq 0$, so in order to minimize the KL divergence between q *and* p we need to maximize the second quantity, also called the lower bound L as $log\ p(x)$ is fixed for any given x**.** In the process of maximizing the lower bound L, we are also maximizing the lower bound of the log probability of the training sample x.

Maximizing the lower bound

$$\begin{aligned}
\sum_z q(z|x) \ . \ log \ \frac{p(x,z)}{q(z|x)} \ &= \sum_z q(z|x) \ . \ log \ \frac{p(x|z)p(z)}{q(z|x)} \\
&= \ \sum_z q(z|x) \ . \ [log\ p(x|z) \ + \ log \ \frac{p(z)}{q(z|x)} \ ] \\
&= \ \sum_z q(z|x) \ . \ log\ p(x|z) \ + \ \sum_z q(z|x) \ . \ log \ \frac{p(z)}{q(z|x)} \ ] \\
&= \ E_{q(z|x)} log\ p(x|z) \ - \ KL(\ q(z|x) \parallel p(z)\ )
\end{aligned}$$

Objective Function

The objective function for the variational encoder is as follows

$$E_{q(z|x)} log\ p(x|z) \ - \ KL(\ q(z|x) \parallel p(z)\ )$$

Which can be interpreted as maximizing the expectation of the log likelihood of the training data under the tractable distribution q and at the same time ensuring that the KL divergence of our distribution q is as close to the actual distribution as possible.

If we assume the $p(z)$ distribution to be a normal distribution, our objective function can be written as follows

$$E_{q(z|x)} log\ p(x|z) \ - \ KL(\ q(z|\ x) \parallel N(z;\ \mu,\ \Sigma)$$

For our case, we can assume the $p(z)$ normal distribution to be $N(0, 1)$

$$E_{q(z|x)} log\ p(x|z) \ - \ KL(\ q(z|\ x) \parallel N(z;\ 0, 1\ )\ )$$

The q function is parametrized as follows (in the implementation it would be modeled using a neural network)

$$q(z \mid x) = N(z;\ \mu(x),\ \Sigma(x))$$

Architecture