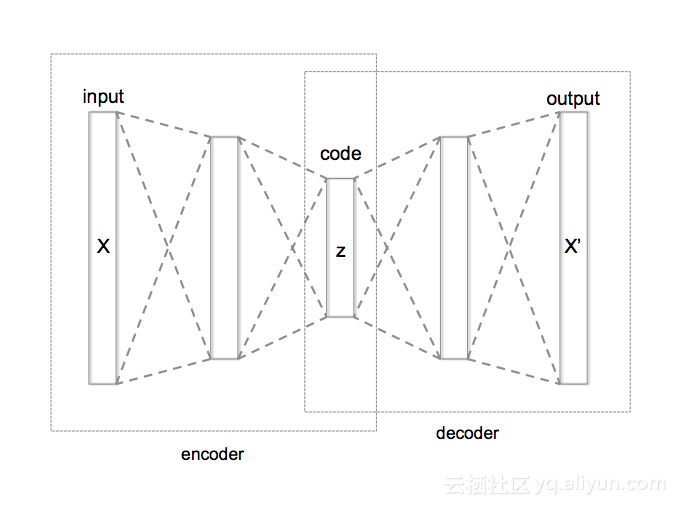
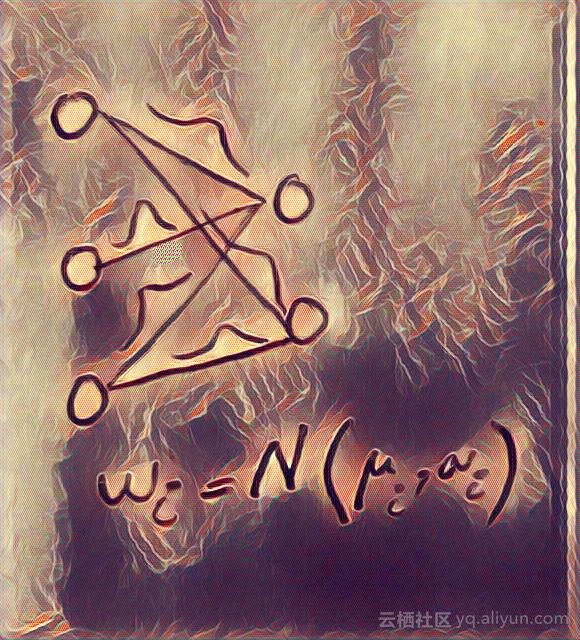
autoencoders作为一种非常直观的无监督的学习方法是很受欢迎的，最简单的情况是三层的神经网络，第一层是数据输入，第二层的节点数一般少于输入层，并且第三层与输入层类似，层与层之间互相全连接，这种网络被称作自动编码器，因为该网络将输入“编码”成一个隐藏代码，然后从隐藏表示中“译码”出来。通过简单的测量重构误差和[反传](https://en.wikipedia.org/wiki/Autoencoder)网络参数能够很好的训练该类网络。



上图是另一种版本的自动编码器，称为“[变分自动编码器-VAE](https://arxiv.org/abs/1312.6114)”，不同点在于其隐藏节点来自于训练期间学习到的概率分布。

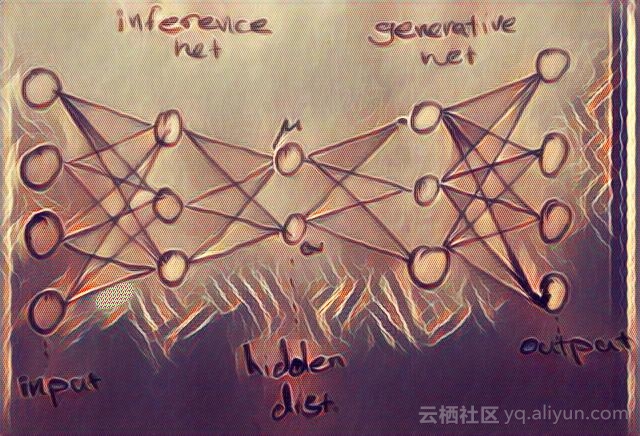
在90年代，一些研究人员提出一种概率解释的神经网络模型，在该模型中，提供了适当的贝叶斯方法。然而，学习这些模型的参数是困难的，直到深度学习研究的新进展导致了用于这种概率方法能更有效的学习参数。

[概率解释](https://en.wikipedia.org/wiki/Probabilistic_neural_network)通过假设每个参数的概率分布来降低网络中每个参数的单个值的刚性约束。例如，如果在经典神经网络中计算权重w\_i=0.7，在概率版本中，计算均值大约为u\_i = 0.7和方差为v\_i = 0.1的高斯分布，即w\_i =N（0.7,0.1）。这个假设将输入，隐藏表示以及神经网络的输出转换为[概率随机变量](https://en.wikipedia.org/wiki/Random_variable)。这类网络被称为贝叶斯神经网络或BNN。



学习的目标是找到上述分布的参数。这种学习被称为“推理”，贝叶斯网络中的推论对应于计算潜在变量相对于数据的条件概率。这类模型引入了变分逼近推理方法后，将该计算问题转换为优化问题，可以使用随机梯度下降法来解决。

在贝叶斯网络中，网络可以基于分布参数重新参数化。在变分自动编码器中，仅在隐藏节点上假设这些分布。因此，编码器变成一个变分推理网络，而且译码器变成一个将隐藏节点映射回数据分布的生成网络。



将分布的均值和方差视为传统网络的参数，并将方差乘以来自噪声发生器的样本以增加随机性。通过参数化隐藏分布，可以反向传播梯度得到编码器的参数，并用随机梯度下降训练整个网络。此过程能够学习到隐藏代码的均值与方差值，这就是所谓的“[重新调参技巧](http://stats.stackexchange.com/questions/199605/how-does-the-reparameterization-trick-for-vaes-work-and-why-is-it-important)”。

在经典版的神经网络中，可以用均方误差（MSE）简单测量网络输出与期望的目标值之间的误差。但在处理分布时，MSE不再是一个好的误差度量，因此用KL­散度测量两个分布之间的差异。事实证明变分近似和真实后验分布之间的距离不是很容易被最小化。它包括两个主要部分。因此，可以最大化较小的项（ELBO）。从自动编码器的角度来看，ELBO函数可以看作是输入的重建代价与正则化项的和

在最大化ELBO之后，数据的下限接近数据分布，则距离接近零，间接地最小化了误差距离。最大化下界的算法与梯度下降的完全相反。沿着梯度的正方向达到最大值，这整个算法被称为“自动编码变分贝叶斯”。