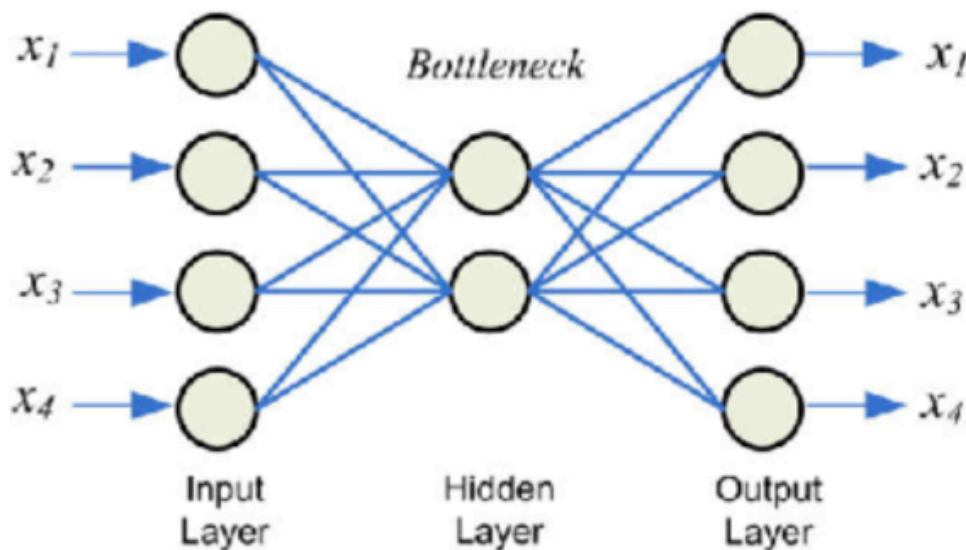


VAE

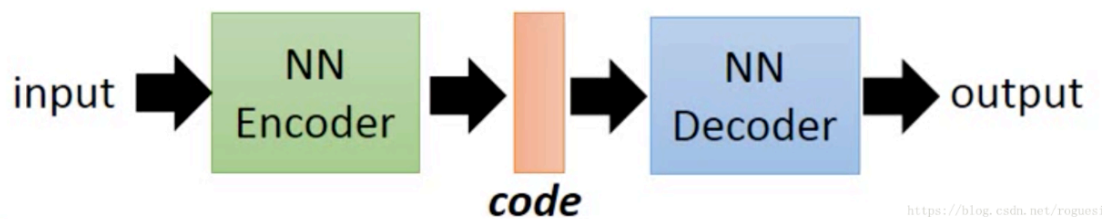
VAE(variational autoencoder)就是指 Auto Encoder 的编码输出服从正态分布。

Auto Encoder

Auto Encoder(后面用 AE 简称)是一个自动编码器，它是通过重建输入的神经网络训练过程，它的隐藏层的向量具有降维的作用。它的特点是编码器会创建一个隐藏层（或者多个隐藏层）包含了输入数据含义的低维向量，然后有一个解码器，会通过隐藏层的低维向量重建输入数据。通过神经网络的训练最后 AE 会在隐藏层中得到一个代表输入数据的低维向量。它可以帮助数据分类、可视化、存储。AE 是一个自动编码器是一个非监督的学习模式，只需要输入数据，不需要 label 或者输入输出对的数据。



AutoEncoder 的基本原理是通过一个 encoder 和一个 decoder 实现的，在卷积的 Autoencoder 中，encoder 通过卷积池化进行降维，decoder 通过反卷积反池化进行升维，目标是使输入和输出图片的差异性最小，这也使得自编码器生成的图片多余输入相似，无法生成新的图片，下图是 Autoencoder 的结构示意图：



虽然 AE 是一个非监督学习算法，但如果它的解码器是线性重建数据，可以用 MSE 来表示它的损失函数：

$$L(x, y) = \sum (x - h_{W,b}(x))^2 y = h_{W,b}(x)$$

如果解码器用 sigmoid 的激活函数，那主要用交叉损失函数：

$$L(x, y) = - \sum_{i=1}^{d_x} x_i \log(y_i) + (1 - x_i) \log(1 - y_i)$$

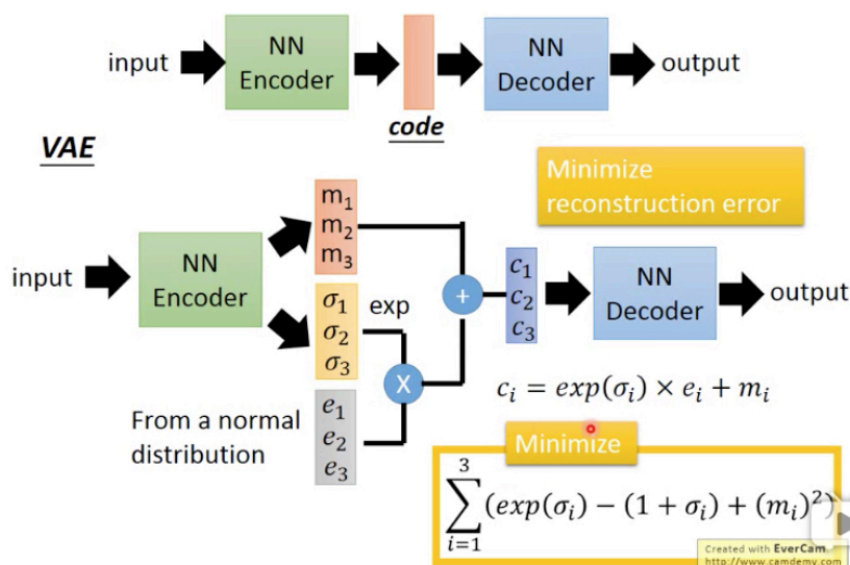
VAE

VAE(variational autoencoder)和普通的 auto encoding 相比，VAE 的 encoder 会输出两个向量，可以把其中的一个看做 mean,另一个看做 variance.同时还要加入一个误差 error 【也可以视为噪声 noise】，这个 error 是从一个高斯分布 sample 出来的，最终把这 3 个向量合成一个 code 【编码】。

VAE 的具体实现过程如下：与 AE 相比，encoder 过程和 decoder 过程都不变，在中间加入了一些 trick, encoder 输出两个 vector，假设原本 AE 中生成的 coder 是十维，生成的两个即 $(m_1, m_2 \dots m_{10})$ 和 $(\sigma_1, \sigma_2 \dots \sigma_{10})$ ，之后再从正太分布(normalize distribution)中生成一个 vector $(e_1, e_2 \dots e_{10})$ ，执行计算如下：

$$c_i = \exp(\sigma_i) \cdot e_i + m_i$$

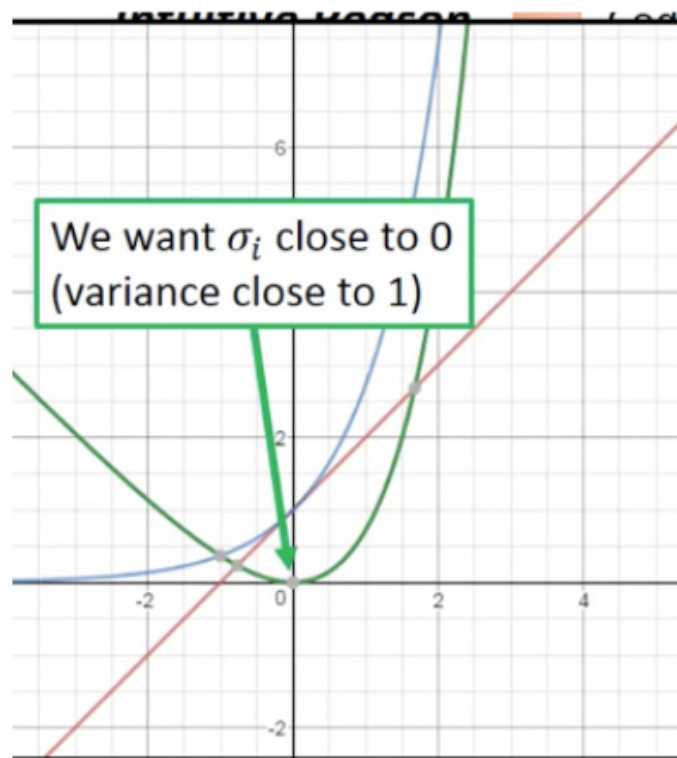
从而生成 $(c_1, c_2, \dots, c_{10})$,再把 $(c_1, c_2, \dots, c_{10})$ 丢进 decoder 中得到输出,通过“使 input 和 output 越接近越好”这个目标来优化模型。



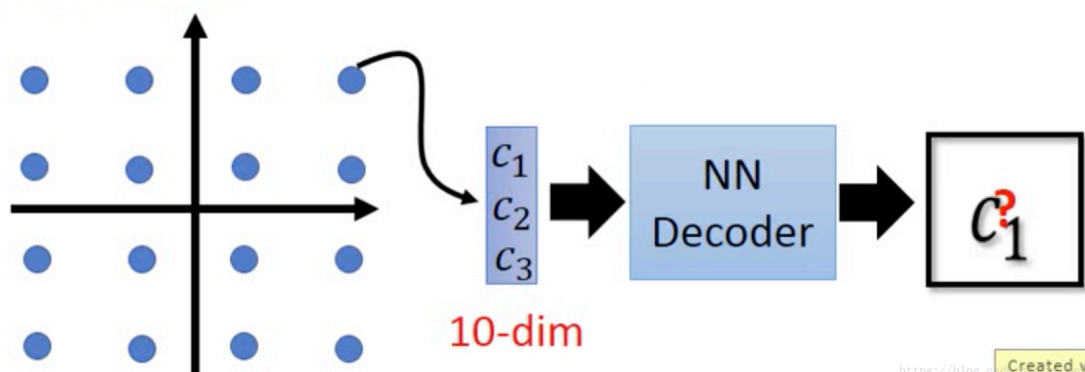
上面说的 noise, 就是上面示意图中的乘积部分: m 对应于原来 AE 中的 code, σ 是从输入图片中生成的, 但他表示 noise 的 variance, e 是从正态分布中抽样得到的, 二者相乘即得到了影响 code 的数值区间的 noise, variance 影响了 noise 的大小, 而且是有 encoder 训练得到的, 如果只考虑“使 Input 和 output 越接近越好”这个目标来优化模型, 那训练得到的 variance 为 0 时候模型理论最佳, 也就变成了原来的 AE。因此需要对这个计算过程做一个限制, 如下:

$$\text{minimize} \sum_{i=1}^{10} (\exp(\sigma_i) - (1 + \sigma_i) + (m_i)^2) \quad (2)$$

如下图所示, 其中 $\exp(\sigma_i)$ 的图像如蓝色所示, $1 + \sigma_i$ 的图像如下图红色部分所示, 他们的差值是绿色图像, 即当 σ_i 为 0 时, 二者的差值会取得最小值, 因此需要加上一个 L2 正则项, 强迫这个 variance 不能太小。



VAE 过程的一个优势就是：可以通过调整 code 中的某维来确定 code 的维度所代表的东西，假设得到的 code 是十维的向量，我们控制其中的八维，来讨论剩下的两维，如下图所示，现在坐标系中均匀采点，然后把取样的点连同固定的八维一起丢到 decoder 中，这样就可以观察这两维对于 VAE 生成图片的作用效果。



如下图所示，分别为人脸的生成映射和 MNIST 数据集上固定其中两维进行图像生成，从左侧的人脸图片中，可以猜测两维特征可能分别控制着人脸的角度和表情；右侧 MNIST 数据集中，两维特征可能控制着数字中笔画的弯曲程度和数字的倾斜程度。

