第七周周汇报

按照第六周的安排，第七周我主要对VAE(Variational autoencoder)做了详细的整理与理解。VAE中文翻译为变分自动编码器，而其中利用到的自编码器(autoencoder)是机器学习的一个基础知识。为了更好的理解VAE我把autoencoder部分又进行了消化。

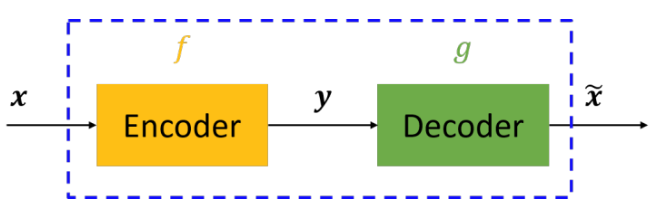
自编码器是通过对输入X进行编码后得到一个低维的向量y，然后根据这个向量还原出输入X，。通过对比X与X，的误差，再利用神经网络去训练使得误差逐渐减小从而达到非监督学习的目的，由图一展示autoencoder过程。

图1.autoencoder流程图

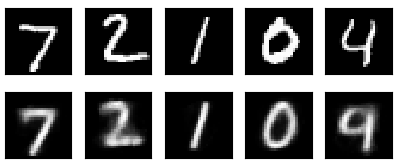
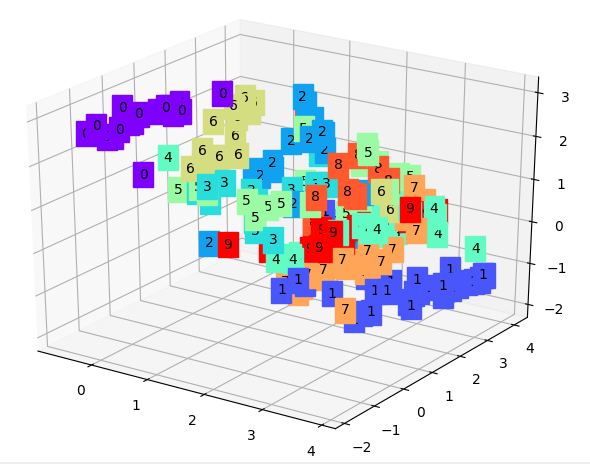
这样训练好的模型，就可以提取输入的特征从而更好的还原输入。我也是根据这中思想参考了一些知识写了一个可视化的autoencoder的小代码，在minist通过matplotlib显示原始数据与autoencoder训练后的数据的图片的对比来观察autoencoder的过程，再提取训练后的encoder的特征分类通过rainbow显示分类的结果，这里也是用到了matplotlib中的3D显示效果，代码在我的主页：<https://github.com/TwistedW/Tensorflow-noting/blob/master/tensorflow_autoencoder_show.py>，图2，图3为实验结果。

图2.autuencoder输入输出对比图

图3.encoder训练后分类效果

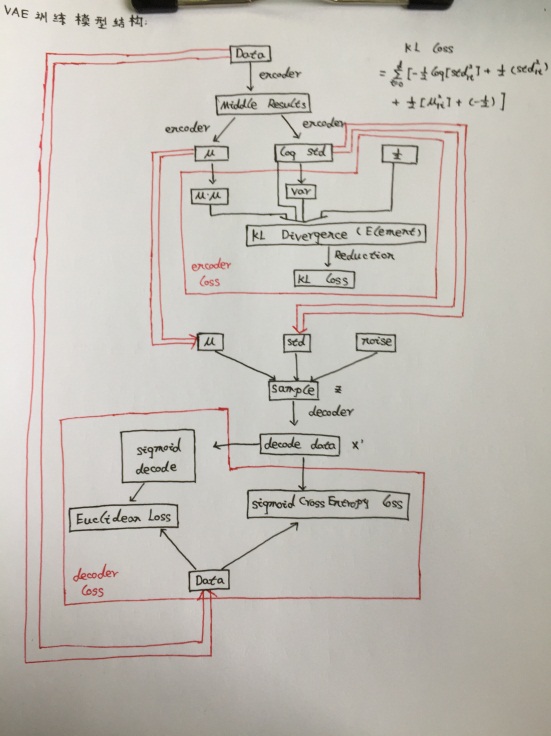
通过实验结果可以看出autoencoder的特征分类效果还是可以的。虽然自编码器可以实现无监督学习但是也存在着一个较大的问题就是模型不能产生任何未知的东西，因为模型不能随意产生合理的潜在变量。因为合理的潜在变量都是编码器从原始图片中产生的。所以在多位大牛的不断努力下就生成了VAE模型，从大方向说的话就是对编码器添加约束，就是强迫它产生服从单位高斯分布的潜在变量。正是这种约束，把VAE和自编码器给区分开来了。VAE模型下产生新的图片也相对容易些，只要从单位高斯分布中进行采样，然后把它传给decoder就可以了。我根据对VAE的理解手动画了一个VAE训练模型结构：

图4.VAE训练模型结构

这张图片的基本思想源于<https://github.com/cdoersch/vae_tutorial>，在讨论VAE的服从单位高斯分布的潜在变量时，论文中的处理方式是利用KL散度来度量潜在变量（图4中sample）的分布和单位高斯分布的差异。KL散度的公式推导我也整理在笔记上了，在仔细推导后对其中的原理也是有了自己的理解。利用KL散度确定encoder的loss，利用交叉熵来衡量decoder的loss，具体的Loss公式应用论文中<https://arxiv.org/pdf/1312.6114.pdf>。

 (1)

此公式的具体推导我会在附录中贴上，右边前一项为encoder部分的KL散度的计算，右边第二项为decoder中Loss的表示。

VAE与GAN的不同之处在于，VAE可以通过编码解码的步骤，直接比较重建图片和原始图片的差异，这一点GAN是做不到。但是VAE的一个劣势就是没有使用对抗网络，所以会更趋向于产生模糊的图片。这一点在实验结果上可以很好的反映出来，VAE虽然可能在生成图片的质量上不如GAN但是VAE一样具有很大的应用前景，VAE相对训练起来比较稳定同时在网络结构不复杂的情况下训练的时间也是比GAN要节省一点时间。

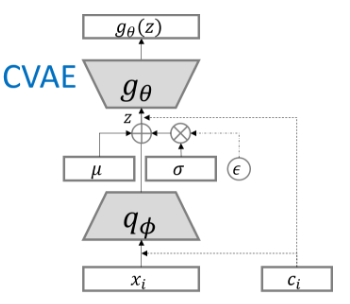
当然VAE也有一些模型上的创新，这一周我还花了一点时间看了[CVAE](https://arxiv.org/pdf/1406.5298.pdf)（条件变分自编码器），条件变分自编码器是在训练、测试的时候加入了一个one-hot向量，用于表示标签的向量。在训练阶段和解码上都要加入one-hot向量，图5为CVAE模型图。

图5.CVAE模型

小结：VAE是无监督学习中一个蛮重要的生成模型，VAE与GAN之间还是有着很多的不同，同样的VAE与GAN各有着自己的优缺点，两种模型的理论知识在两周时间的阅读下已经有了基础的认识。VAE利用KL散度衡量潜在变量的分布和单位高斯分布的差异，通过encoder和decoder之间的损失函数训练可以在生成图片上有较好的对比效果，在encoder端的特征分类上也具有一定的优势。本周的时间用于对VAE这篇论文和其中原理进行了详细的解读，在KL散度和VAE损失函数的推导都有了自己的理解。

这两周只是对GAN和VAE的基础知识上有了一步的了解但是其中的一些细节和论文的核心还有着一些地方没有理解透。所以在下一周我准备在阅读VAE与GAN结合在一起的模型文章的同时把GAN和其拓展模型中的损失函数的理解上再精进一些。

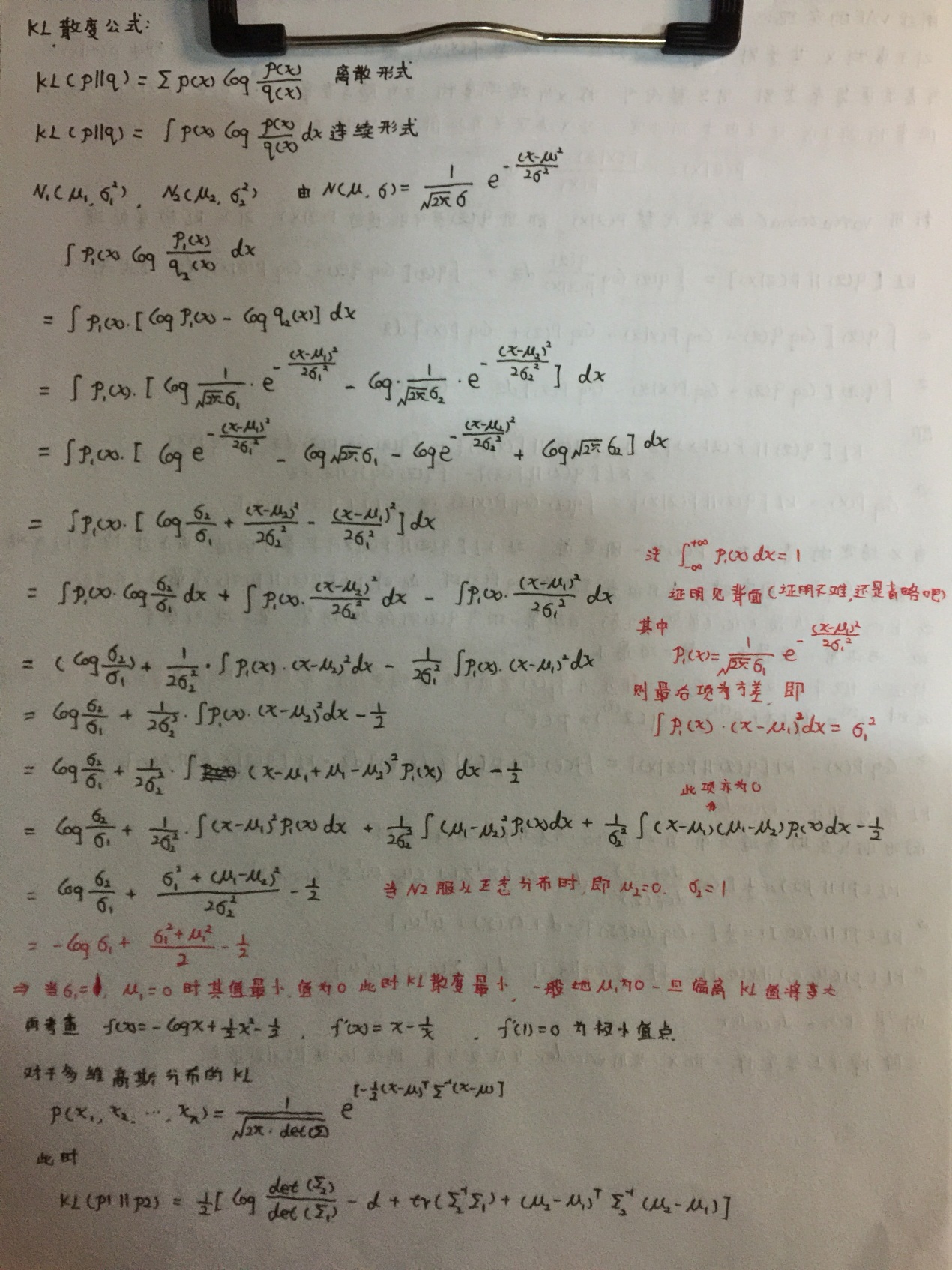
附录：

图6.KL散度公式证明

图7.VAE的实现