在正式开始汇报这周工作之前要把上周的那篇CVAE-GAN再简单的说一下。CVAE-GAN的一大创新就是均值特征匹配的提出，这种方法一定程度上可以提高生成器G的稳定性。通过提取网络的中间特征函数来缩小网络的生成和真实数据的特征差异达到网络的优化。文章的均值特征匹配围绕着生成器展开，目的也是为了稳定生成器，GD之间的中间层特征提取了D的最后一个全连接层作为特征函数，GC之间的中间层特征提取了分类器C的最后一个全连接层作为特征函数。多层特征结合提高网络收敛的同时还能稳定生成器，模型的提出确实意义很大。受到这篇文章启发，阅读了陈老师发的[generating image with perceptual similarity metrics](https://arxiv.org/abs/1602.02644)，对中间特征有了一定的了解。在这篇文章中提到了图片的特征上的差异，在图像空间差异的基础上是一个很大的提高，通过分析图像的特征差达到空间上的无法实现的差异的缩小更能提高图片的平滑程度。文章将提出的方法命名为“深层感知相似度量”(deep perceptual similarity metrics)，文章的重点是损失函数的设计，由图1展示。

图1.DeePSiM模型图

值得说的是这篇文章的C是一个比较器，只是用来提取图像差异的并不是CVAE-GAN中的分类器C的含义，在CVAE-GAN中在LG的设计上与这篇文章有着异曲同工之处，在一定程度上比较了图像的特征差异，这一块在后续代码的研读上再做进一步的理解和深化。

 (1)

按照上一周的计划，本周对比各类GAN的优缺点，在具体研读之前我在网上参考了有经验的学者建议。目前在各类GAN中，[infoGAN](https://arxiv.org/abs/1606.03657)和[WGAN](https://arxiv.org/abs/1701.07875)是近两年的重大创新。虽然之前有阅读过这两篇，但是都是泛泛的看了一下对于具体的实现方法和公式的理解还存在一定的问题。本周我对这两篇文章作了细致的研读，WGAN只读了前作。

从非监督学习研究角度来看，GAN存在着一些美中不足的地方，在实现图片的可编辑生成上GAN是做不到多样化图像的，在现有的很多具有很好效果的GAN并不是完全非监督的，而是人为的加入了很多带标签数据的半监督学习link。传统GAN生成数据是通过一组完全随机的z隐变量得到，这个z基本是不可控的，我们很难通过控制z中某数的大小变化让生成的图像变大变小或进行旋转等一些可编辑操作，infoGAN正是基于这一问题而提出的GAN修正模型，infoGAN被OPENAI评选为2016年五大突破之一。

传统的GAN在generator输入端z是一个连续的噪声信号，这个z是随机的不受约束的，这就导致GAN无法很好的利用，将z的具体维度与数据的语义特征对应起来不是一个Interpretable（可解释）Representation。infoGAN正是利用z的特性作为改进点，文章试图利用z，寻找一个可解释的表达，文中将z进行了拆解分为不可压缩的噪声z以及可作为解释的隐变量c，称作为latent code。这样就可以约束c与生成数据之间关系，使得c里面包含有对数据的可解释的信息，比如在处理mnist数据集上c可以分为categorical latent code来表示数字种类信息（0~9），在做人脸生成时可以将c作为人脸的倾斜角度，在笔画的粗细上也可以作为表征信息。

为了引入c，文章利用互信息来对c进行约束，这是因为如果c对于生成数据具有可解释性，此时c和具有高度相关性，即互信息大，这是文章希望达到的效果。而如果是无约束的话，那么它们之间没有特定的关系，即互信息接近于0。c与的互信息越大就可以保证解释度越高，此时GAN模型的目标函数也跟着改变了。

 (2)

此时互信息并不是容易计算的，真实的并不清楚，因此在具体的优化过程中，作者采用了变分推断的思想，引入了变分分布来逼近，在基于最优互信息下界的轮流迭代实现最终的求解。具体的公式推导过程我将在附录中展示。

传统GAN在自提出以后最大的冲击就是[WGAN](https://arxiv.org/abs/1701.07875)的提出，WGAN也引起了GAN的作者Ian Goodfellow的高度关注，并且在各大论坛转载分享建议大家一起学习。为什么WGAN能这么的火热呢？这就要归结于WGAN作者[Martin Arjovsky](https://www.linkedin.com/in/mart%C3%ADn-arjovsky-b7582350)强大的数学功底。作者利用严谨的数学公式推证了GAN存在的两大问题。第一个就是传统GAN在判别器训练的越好的情况下，生成器越容易出现梯度消失情况；第二个问题就是传统GAN的collapse mode，模型崩塌或者称为多样性不足的造成原因。在严谨的数学公式推证下连GAN的作者也正视了这个问题，鼓励大家一起学习WGAN。

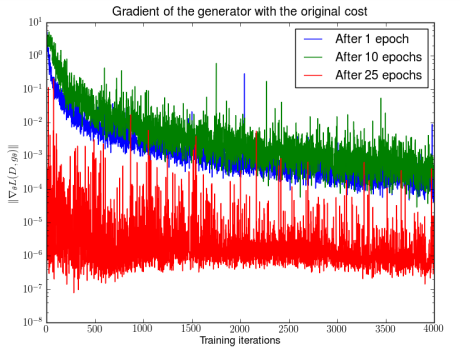
WGAN算是有两篇文章，[第一篇](https://arxiv.org/abs/1701.07875)是说明GAN存在的问题，[第二篇](https://arxiv.org/pdf/1701.04862.pdf)是提出新算法改进。坦白的说这两篇论文的公式多的有点离谱了，但是这么经典的东西我还是在这一周把第一篇文章读完了。根据已有知识和各大论坛的拜读以及文章的理解我算是用已有的数学能力把WGAN前作中的公式做了梳理和理解，具体推导在附录中。

简单阐述一下GAN的两大问题在数学基础上的问题原因。在传统GAN文章中有推导过判别器最优解和生成器的最优解，我之前报告上也在附录中推导过。正是由于对生成器的最优解的理想化造成大家陷入对模型的信赖上（大多数学者不相信也不行啊C:\Users\lenovo\AppData\Local\Temp\Z}4JA~3{S79KP7UYELJH){7.gif）。在生成器最优解的推导时，理想下的JS散度为0，但是这个JS会为0吗？作者提出一旦数据上升到高维情况下JS散度将不可能为0，而是趋近一个常数log2，这就导致G的目标函数值为0从而造成梯度消失情况，也就是GAN的第一种问题。第二种问题的造成也发生在G的损失函数上，GAN的作者在写G的损失函数时做了个变换。

 (3)

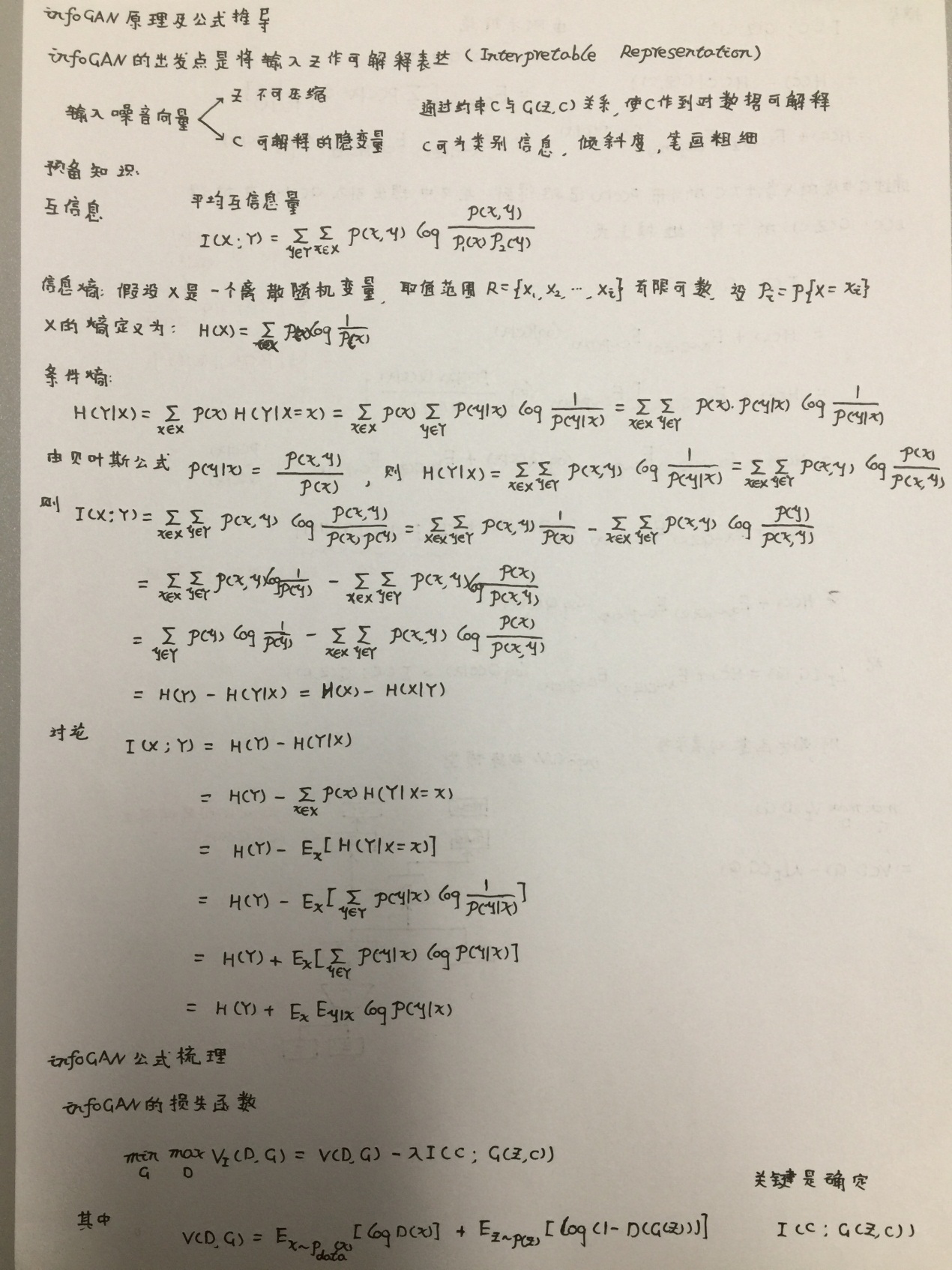
这个损失函数当用G的最优解和KL以及JS散度做变换后出现KL与JS之间的分布矛盾的现象，这样在生成器生成数据时，生成数据分布和原始数据分布在两种不同衡量度下产生矛盾，最后的结果是G为了不受大的惩罚更愿意犯惩罚度小的错误，经过不断训练后，这种错误将会导致G生成的数据缺乏多样性从而导致出现collapse mode。

WGAN论文洋洋洒洒的用了17张纸推证这两个问题，密密麻麻的公式让人望尘莫及。在阐述上的严谨确实说明了GAN的问题造成原因，作者通过重置训练阶段的weight值验证结论的正确性，从实验的结果上确实可以得到论文的推证也是没问题的。

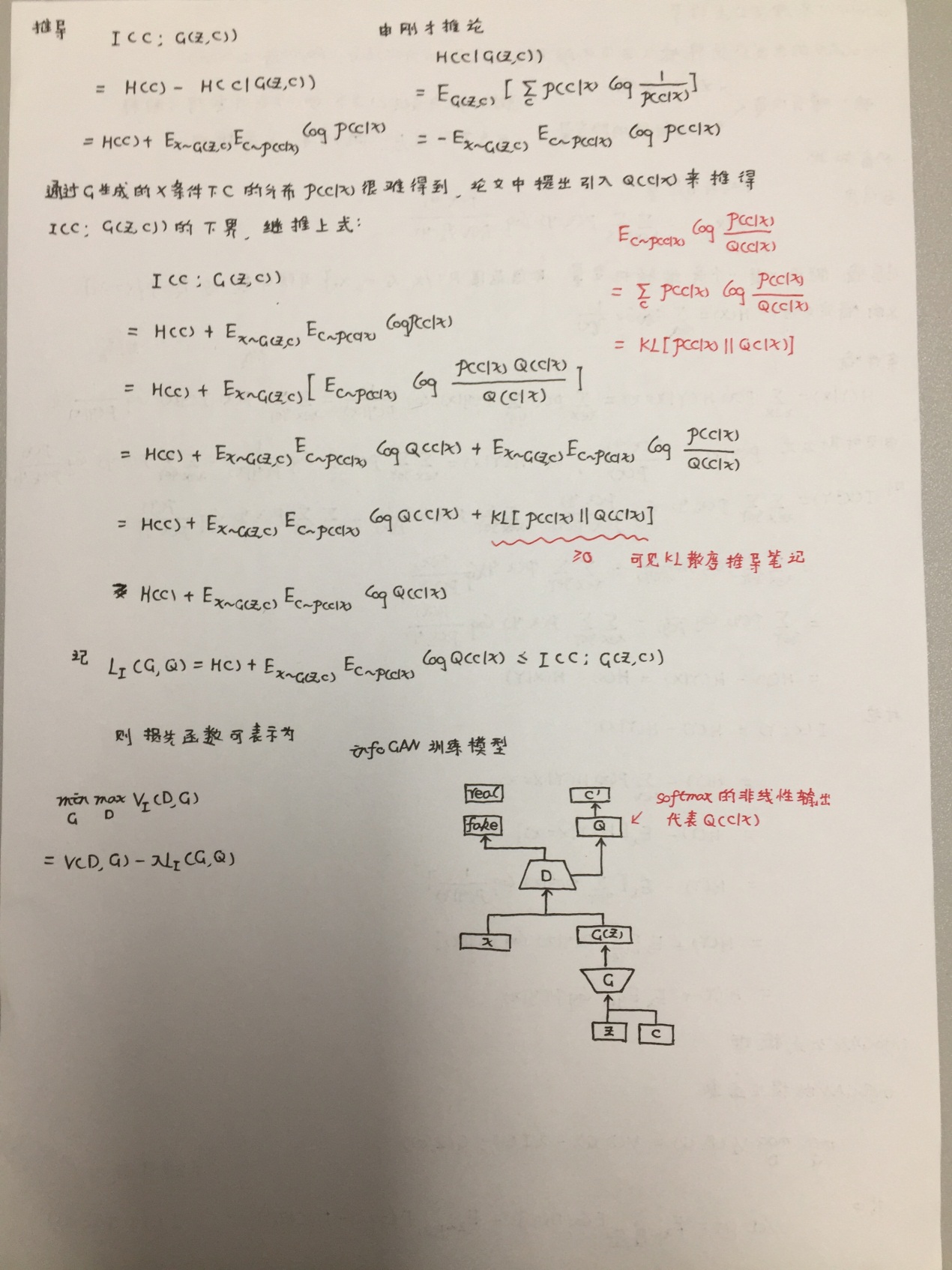
图2. 先分别将DCGAN训练1，20，25个epoch，然后固定生成器不动，判别器重新随机初始化从头开始训练，对于第一种形式的生成器loss产生的梯度可以打印出其尺度的变化曲线，可以看到随着判别器的训练，生成器的梯度均迅速衰减。注意y轴是对数坐标轴。

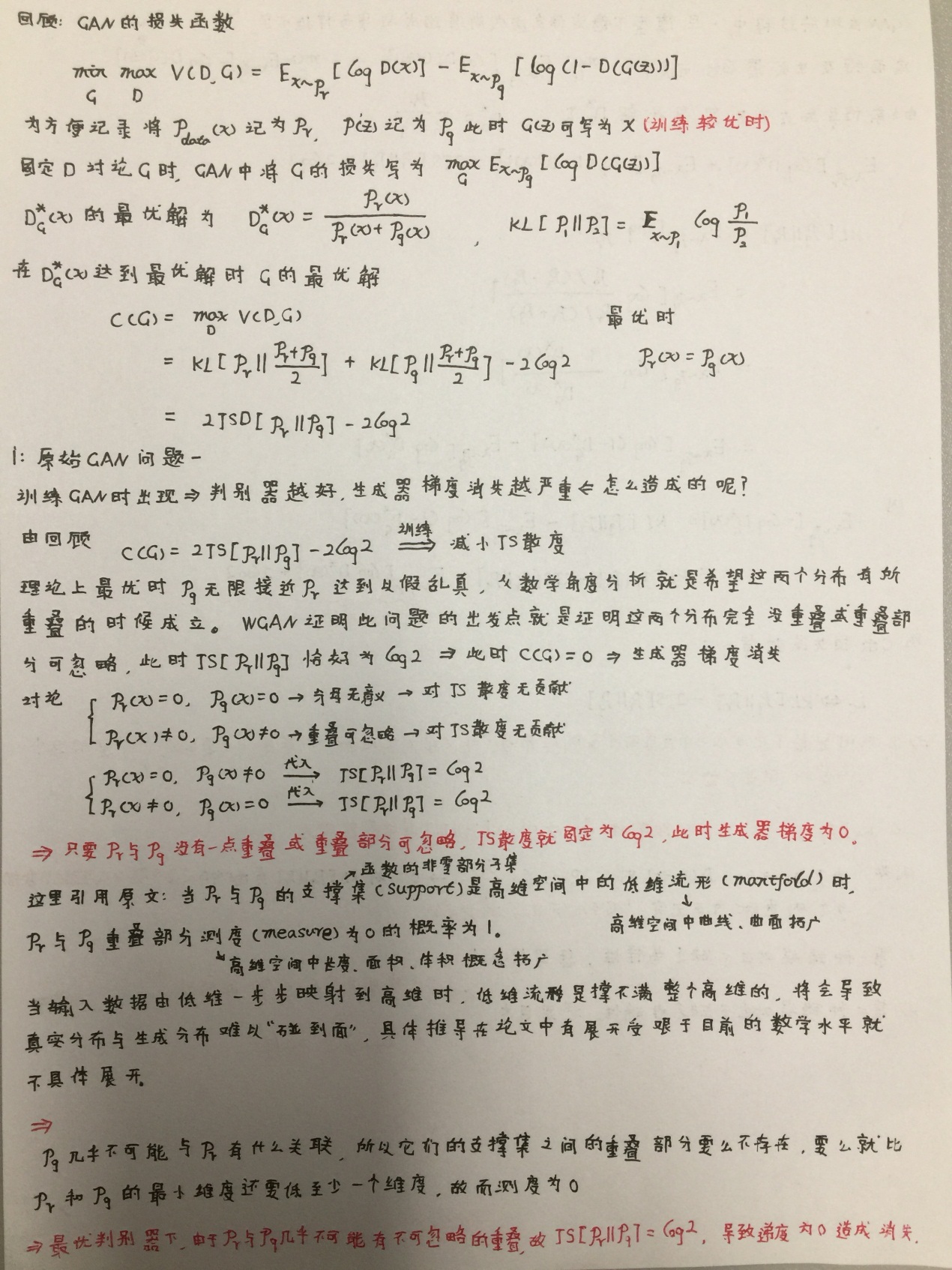
小结：本周在上周CVAE-GAN的基础上阅读了深层感知相似度量，对中间层特征在loss中的使用有了一定的了解。在已有GAN的知识基础上对两个杰出GAN做了研读。infoGAN在输入端将z拆分为两块，加入了可解释隐含量c用于对图像或者数据通过互信息达到对数据的可控制生成效果。WGAN前作通过繁琐的数学推证了GAN存在的两大问题的原因，我整理了一部分推导在附录中展示。

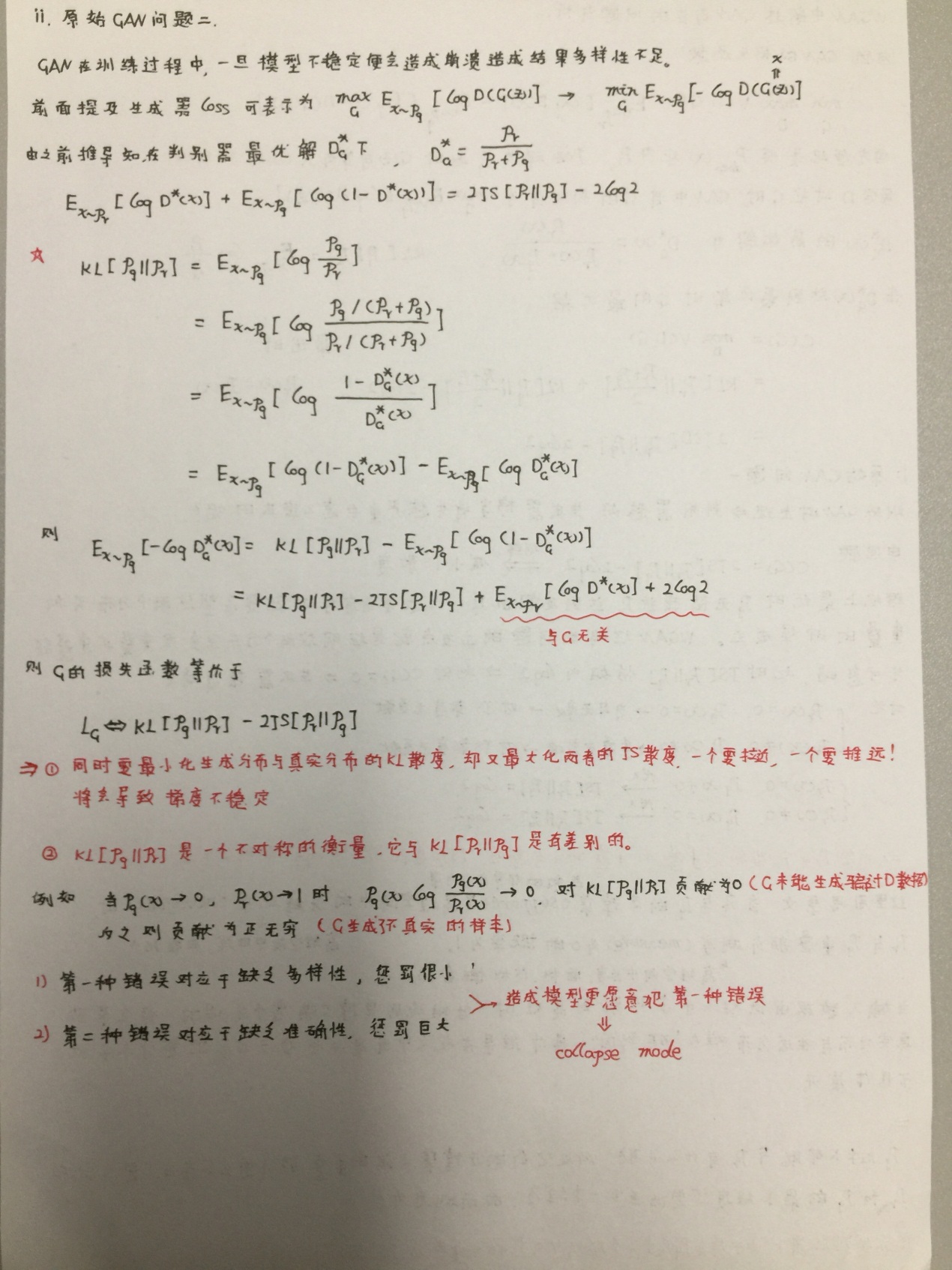
对WGAN前作的阅读让我认识了GAN存在的问题，在下一周把WGAN的后作研读完，在已有知识基础上再细致阅读LSGAN和其中的原理以及公式意义。我尝试着在笔记本上运行数据，发现效率太慢还会引起电脑卡机，暂时我以阅读论文和看代码框架为主。

**附录：**

**infoGAN推证一**

**infoGAN推证二**

**GAN存在问题一推证**

**GAN存在问题二推证**