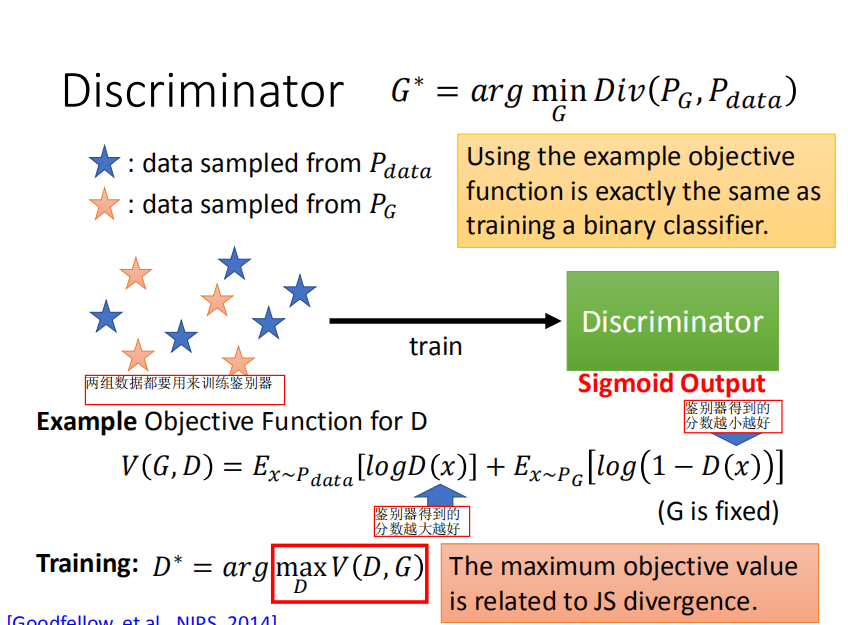
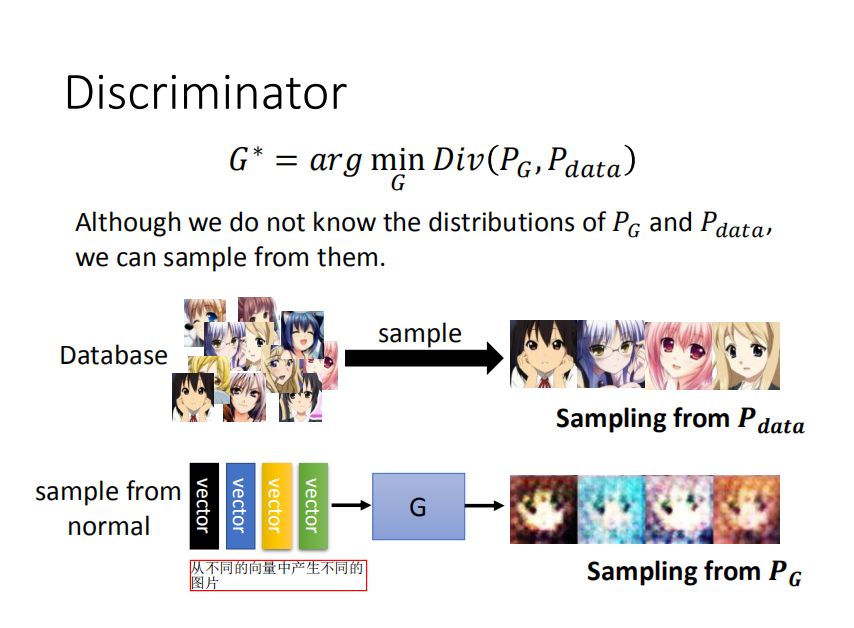
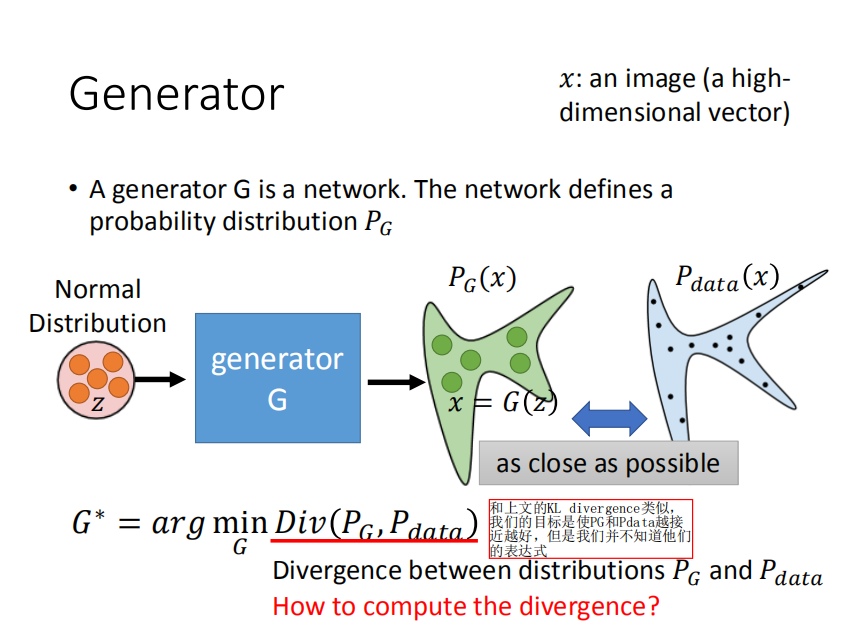
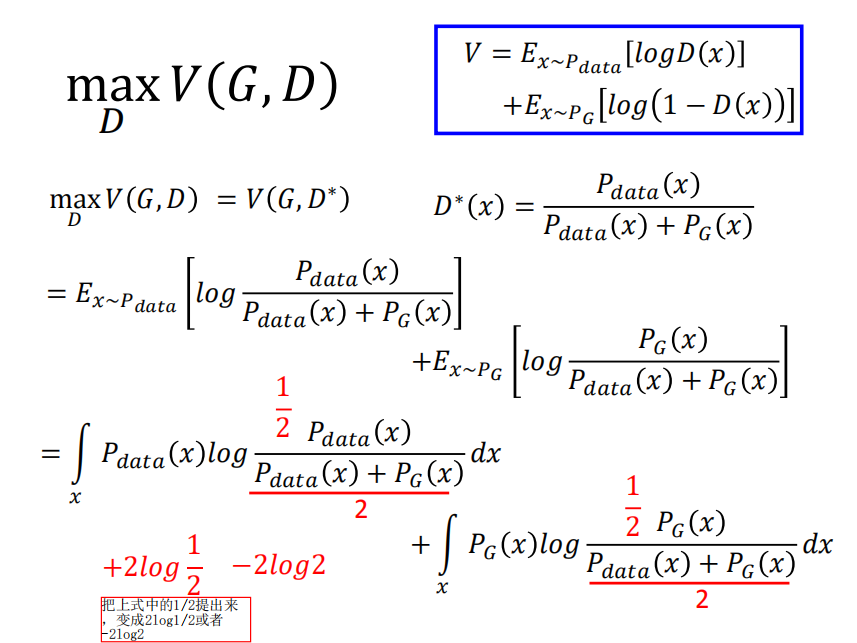
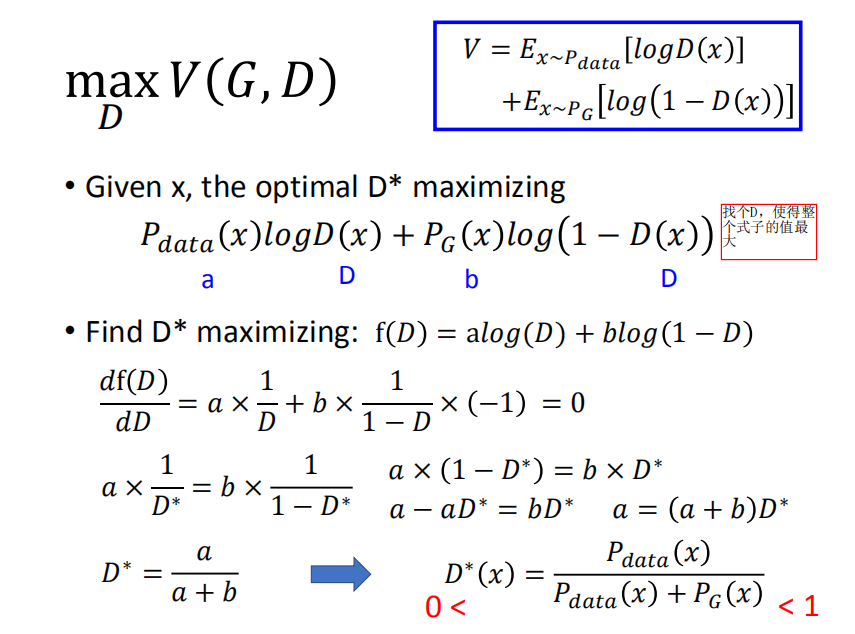
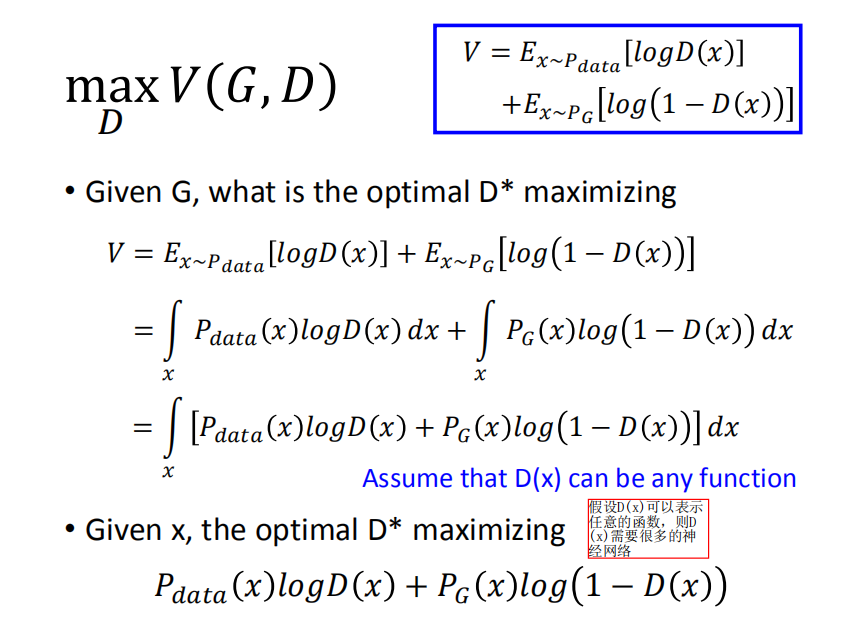
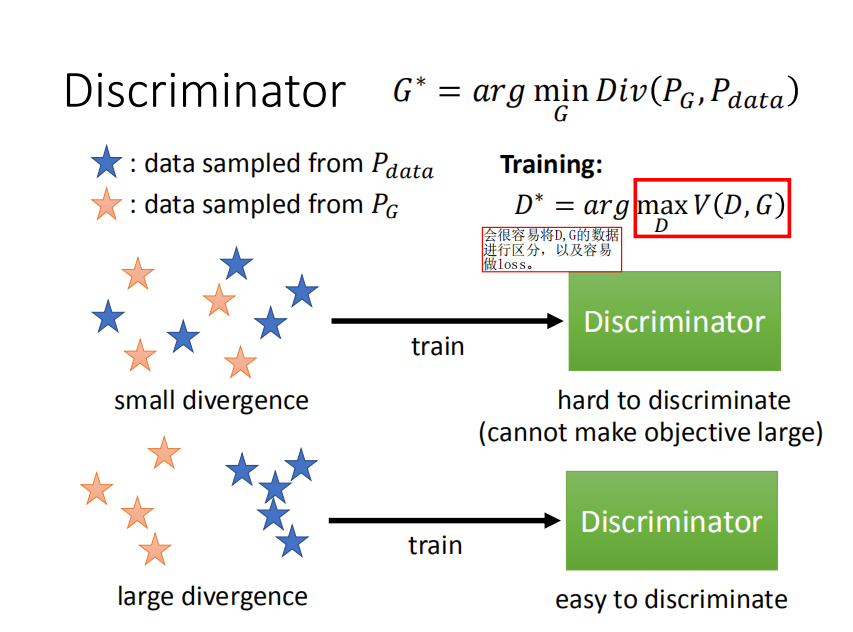


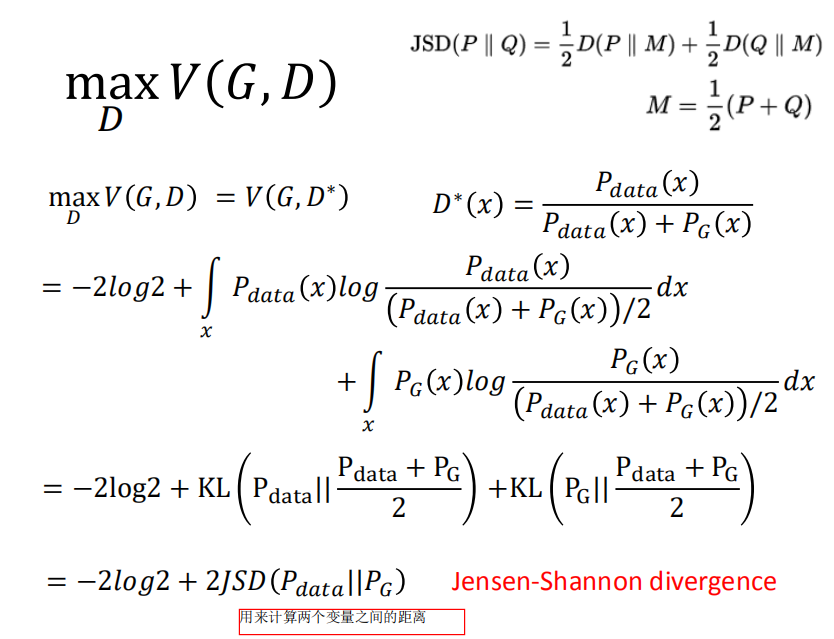
在后减去为了让更加的接近，所以才会使用KL最小化散度



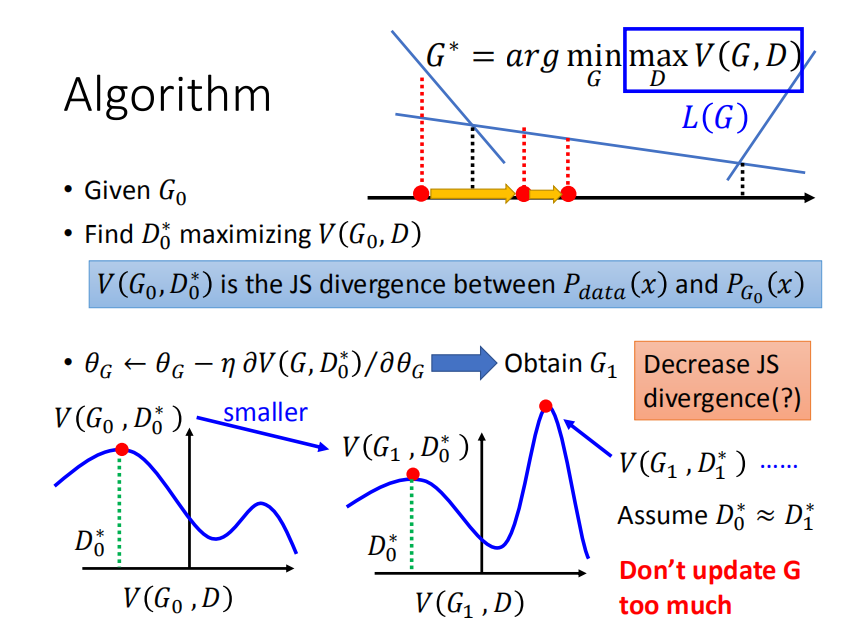
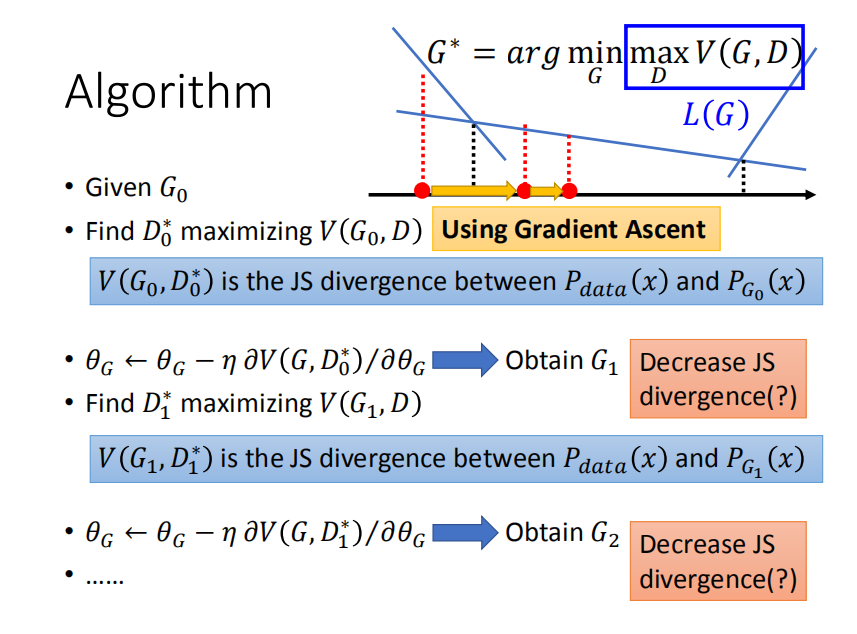
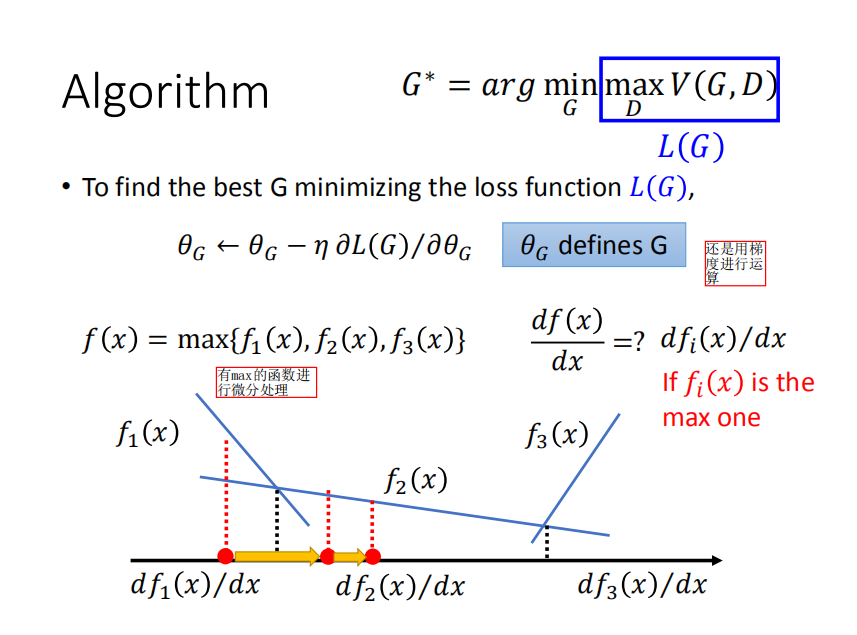
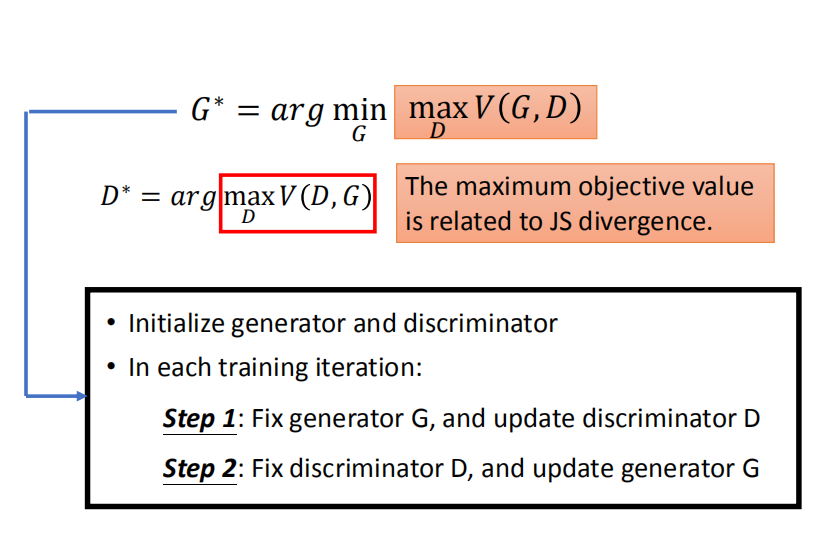
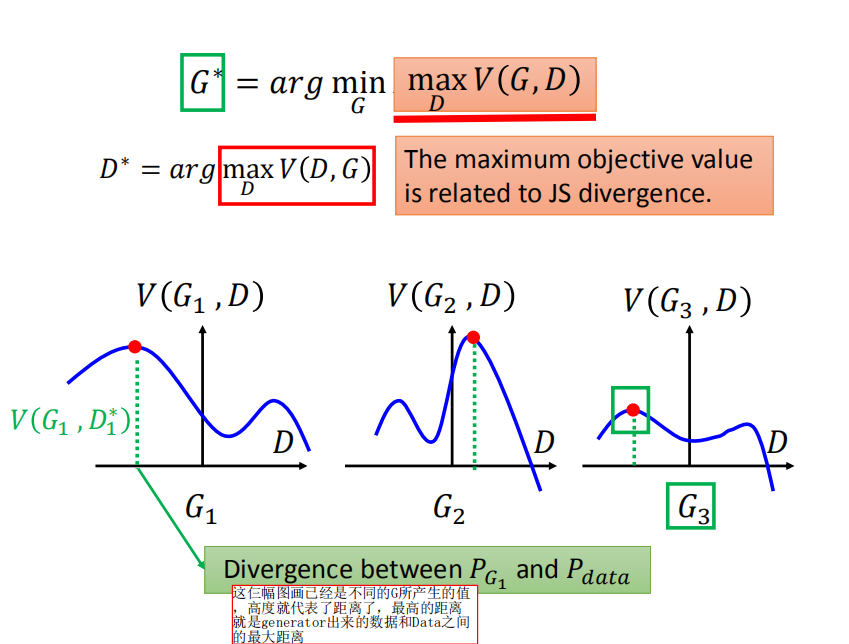
对于鉴别器D来说，它需要的是两种样本有足够大的差距，通过这种差距可以很好的训练出来鉴别器。所以，我们希望从从data中sample出来的样本可以取得较高的成绩，从生成器中得出的样本会取得较低的成绩。



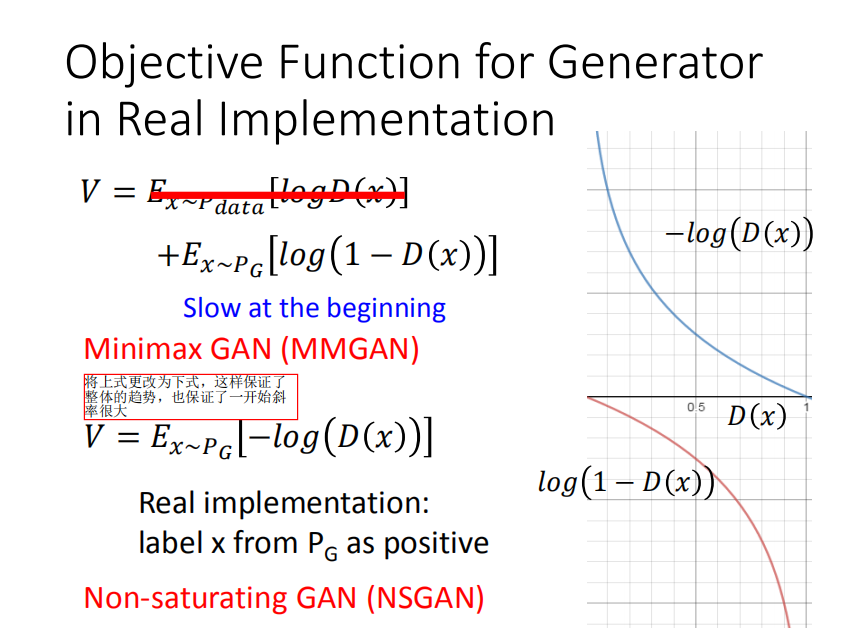
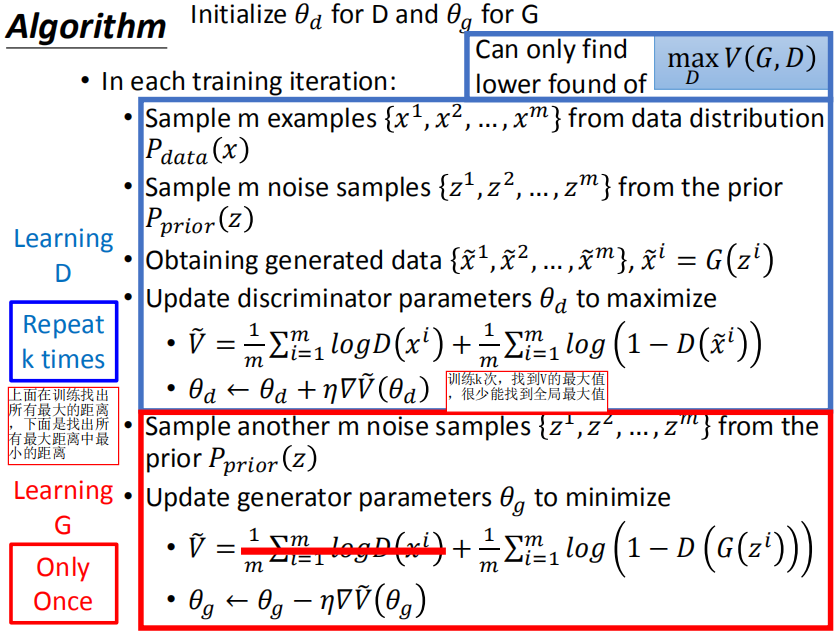
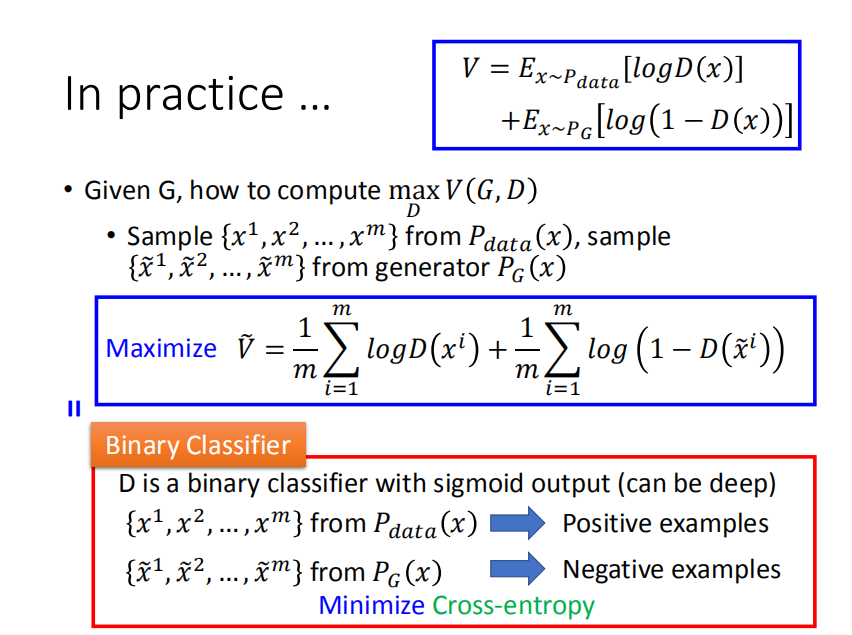
和KL-Divergence的公式进行类比，得出log2是因为对概率积分之后，得到的值是1。



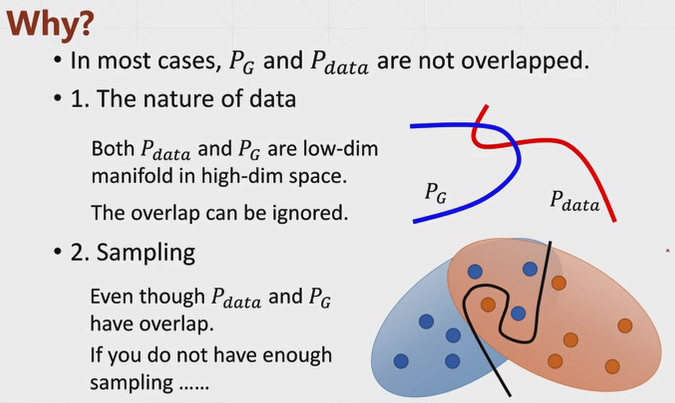
最小值的情况为2JSD(Pdata||PG)的值最小0，此时Pdata==PG,此时D 的最优值为0.5，也表示鉴别器D最优的值时0.5



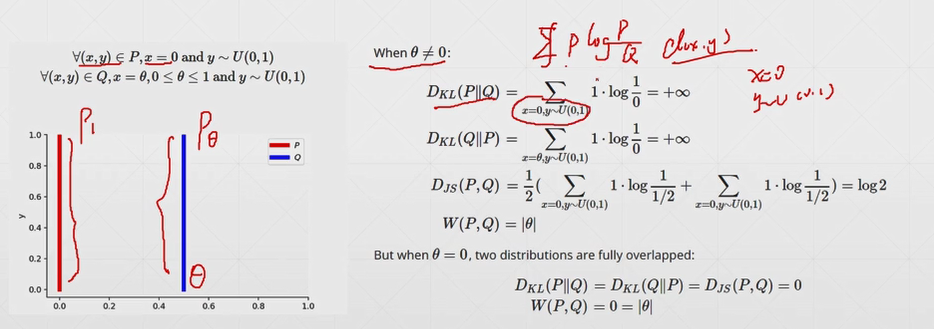
假设在D0的位置上V(G0,D)的值大于V(G1,D)的值，那么我们在最小化的时候会选择V(G1,D)，当在V(G1,D1)时，V(G0,D1)的值又会比较小，如何处理呢？不要更新G太多。



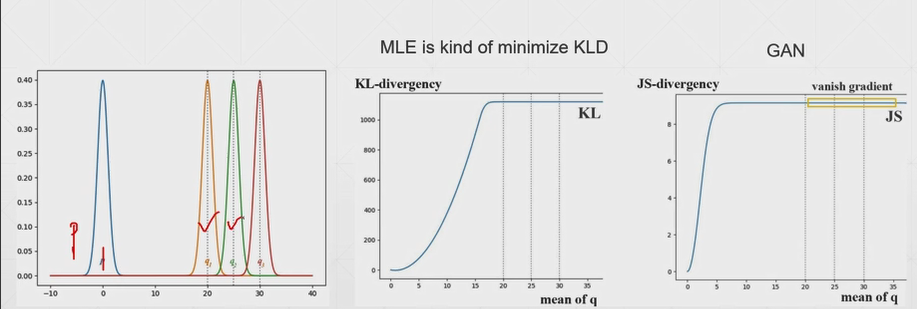
**GAN在train时的缺陷？**



在GAN刚刚开始训练的时候PG使一些噪声数据和Pdata的真实数据肯定没有太多重合的数据部分。



我们可以将没有重合的情况理解为两个完全不同的分布，通过Dkl的公式可以发现，当x=0，y~U(0,1)的情况时，P1的概率为1而Pθ的概率为0此时Dkl为正无穷，Djs也为一个固定的值，并不利于做梯度运算。



p是最左端的分布曲线，q1、q2、q3是三个参照的分布，当q的均值是0的时候，此时的p、q已经重合，此时KL-divergence的值也为0，但是当q的均值大于20的时候，KL的值就恒定不变了，没办法再反向做梯度运算了。

