GAN

GAN

数学原理：

原数据：

$$P_{data}$$

生成的概率分布(Generator生成)：

$$P_g(x:\theta_g)$$

噪声分布（比较简单的分布）：

$$P_z$$

过程简述：

从噪声分布中选取点，进入GNN学习（generator），输入z（分布中的点），通过学习输出

$$output = G(z:\theta_g)\text{中的样本}$$

ps: $P_g = G(z:\theta_g)$
再将输出的x（生成器的结果）和从P_{data} 中选取的x（样本--真品）送入另一个GNN（鉴别器）训练。

判断鉴别器的鉴别水平？

能够鉴别生成图片的真假： $D(x)$ 表示图片为真的概率 ，这个GNN的训练目标是：GNN使得假输入$D(x)$ 低，真输入高。
if x comes from P_{data} , $D(x)$ 大， $logD(x)$ 也大
if x comes from P_z , $D(x)$ 小，这里x 是来自generator产生分布的

样本，所以也就是可以用$output$ 代替，$D(output)$ 小，$D(G(z:\theta_g))$ 小，$log(G(z:\theta_g))$ 小，$1-log(G(z:\theta_g))$ 要大。
目标为：$\max_D E_{x\sim P_{data}}logD(x)+E_{z\sim P_z}log[1-G(z:\theta_g)]$

判断生成器生成水平?

if x comes from P_g ,$D(x)$ 大， （生成水平高让鉴别器误以为真实）
这里x 是来自generator，所以也就是可以用$output$ 代替，$D(output)$ 大，$D(G(z:\theta_g))$ 大，$log(D(z:\theta_g))$ 大，$log[1-G(z:\theta_g)]$ 小。
目标： $\min_G E_{z\sim P_z}log[1-G(z:\theta_g)]$
综上目标函数：
$\min_G \max_D E_{x\sim P_{data}}logD(x)+E_{z\sim P_z}log(1-G(z:\theta_g))$

GAN优化

discriminator：

伯努利分布（二项分布）

目标函数存在最优解吗?

存在最优解，要能证明$P_g == P_{data}$?

证明 存在最优解时，$P_g == P_{data}$(简做P_d)

假设 G不动。
$V(D,G)=E_{x\sim P_d}logD(x)+E_{x\sim P_g}(log(1-D(x))$
求 $\max_D V(G,D)$
$V(G,D)=\int P_d*logD(x)dx+\int P_g*log[1-D(x)]dx$
$=\int(P_d*logD(x)+P_g*log[1-D(x)])dx$
$=\int_0^1(P_d*logD+P_g*log[1-D])dD$

(注解：看上去D与x是相关的，输入一个x，输出一个D(x)，但是将x看作开区间，x可以取实数任何值，但是它们对D映射的值一定在0到1之间)

如果要求$V(G,D)$最大，那么它对D的导数就应该是0；

$\frac{\partial V(G,D)}{\partial D}=0$

$\frac{\partial V(G,D)}{\partial D}=\int\frac{\partial(P_d*logD(x)+P_g*log[1-D(x)])}{\partial D}=0$ 积分与微分可以交换计算

$P_d*\frac{1}{D}+P_g*\frac{-1}{1-D}=0$

$P_d*\frac{1}{D}=P_g*\frac{1}{1-D}$

$D_G^*=\frac{P_d}{P_d+P_g}$

现在D已经求出，回代到方程。

$\min_G V(G,D_G^*)=E_{x\sim P_d}log\frac{P_d}{P_d+P_g}+E_{x\sim P_g}log\frac{P_g}{P_g+P_d}$

观察上述算式，发现类似于kl divergrence

但是p_d+p_g不是分布，因为当p_d = 1,p_g = 1的时候，p_d+p_g 就等于2，明显不是分布。

凑出kl divergrence

$\min_G V(G,D_G^*)=E_{x\sim P_d}log[\frac{P_d}{\frac{P_d+P_g}{2}}*\frac{1}{2}]+E_{x\sim P_g}log[\frac{P_g}{\frac{P_d+P_g}{2}}*\frac{1}{2}]$

$=D_{KL}(P_d||\frac{P_d+P_g}{2})+D_{KL}(P_g||\frac{P_d+P_g}{2})-log4>=-log4$

当$P_d=P_g=\frac{P_d+P_g}{2}$时，“=”成立

目标函数变成最小，$D_G^*=\frac{1}{2}$

kl divergence 是啥(相对熵)

量化两种概率分布P和Q之间差异的方式

分布的熵

信息量：$-logp$ --p:一件事发生的概率

熵：信息量对概率分布的期望；

$E_{x\sim p(x)}[-logp]=\int-p(x)*logp(x)$

$H=-\sum_{i=1}^{N}p(x_i)*logp(x_i)$

KL divergrence定义：

$D_{KL}(p||q) = \sum_{i=1}^{N} p(x_i) * (logp(x_i) - logq(x_i))$

$D_{KL}(p||q) = \sum_{i=1}^{N} p(x_i) * log\frac{p(x_i)}{q(x_i)}$

***K-L散度其实是数据的原始分布p和近似分布q之间的对数差值的期望**

KL可以看出不同分布之间的差距（不是距离）用q近似p造成的误差