生成对抗网络

机器学习 · 生成模型专题

学习目标

没有配对目标图，生成器怎样获得学习信号?

两条路线共同目标

- 解释G/D与隐式生成分布，区分训练与生成
- 用原图区分特征张量、完整输出与训练更新
- 追踪交替更新，说明固定D参数为何仍保留输入梯度
- 解释最优D/JS结论及条件，比较两种G目标的梯度
- 解释覆盖、Wasserstein目标与约束，以及条件生成接口

完整路线进阶目标

- 推导最优D与JS联系，展开logit/Jacobian梯度链并复算数值例

1 生成器与判别学习

没有配对目标图，生成器怎样通过来源判别获得训练信号？

随机映射规定生成分布

生成器怎样把随机输入变成可采样的分布？先固定映射与分布的身份，再讨论训练信号从哪里来。

生成对抗网络（Generative Adversarial Network，GAN）通过可学习的判别信号训练生成映射。设噪声 $\boldsymbol{z} \in \mathbb{R}^k$ 来自固定先验 p_z，生成器

$$\tilde{\boldsymbol{x}} = G_\theta(\boldsymbol{z}) \in \mathbb{R}^d$$

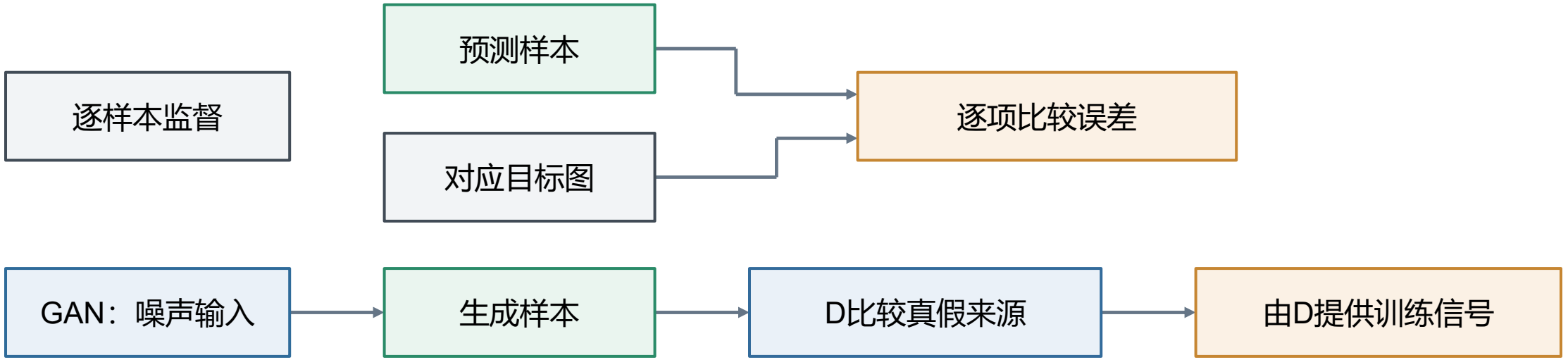
将它转换为样本。

生成分布 P_g 是输出样本的概率规律。

例如对区域 A，$P_g(A) = Pr(G_\theta(\boldsymbol{z}) \in A)$。$\theta$ 是训练改变、使用时固定的参数。

没有配对目标图，误差信号从哪里来?

逐项误差需要对应目标；GAN改用真实样本与生成样本的来源判别。

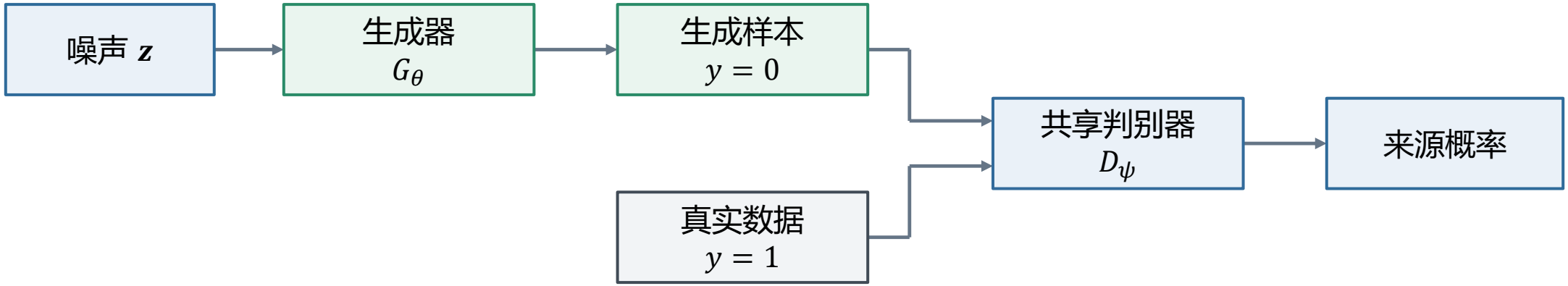


真实样本集提供分布参照；每个噪声没有唯一对应目标图

G改变输出，使D更倾向判为真实；下一页把两个来源送入同一个D。

两种来源进入共享判别器

判别器 $D_\psi(\boldsymbol{x}) \in (0,1)$ 输出在当前真假来源分类任务中属于真实来源的预测概率。训练构造真实标签 $y = 1$、生成标签 $y = 0$，两类来源在下述目标中等权；标签不表示图像语义类别。



生成映射 G_θ 提供直接采样方式，样本密度是否易于计算仍取决于具体模型。

生成阶段使用训练后的生成器

生成时固定 θ，依次执行：

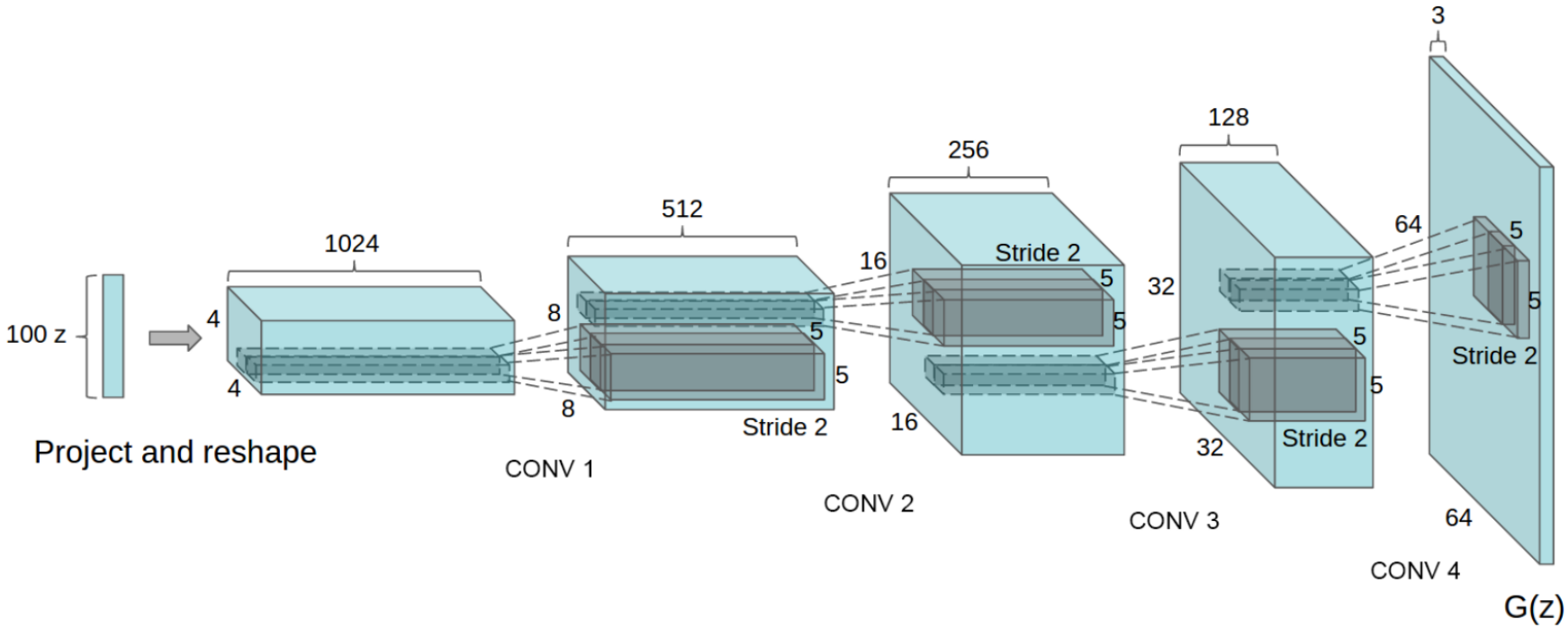
1. 按指定先验独立抽取 $\boldsymbol{z}$。
2. 前向计算 $G_\theta(\boldsymbol{z})$，得到新样本。
3. 需要多个样本时重新抽取噪声并重复映射。

判别器参与学习信号的形成，通常不属于这个基本采样算法。生成器也没有自动提供“观测图像→潜变量”的逆映射。

训练需要真实数据来建立判别任务；生成阶段无须提供一张逐项配对的真实目标图。生成是否覆盖目标分布、是否记忆训练数据，需要另行检查。

DCGAN把抽象生成映射落实为卷积网络

随机输入 → 特征空间逐级扩展 → 最终图像



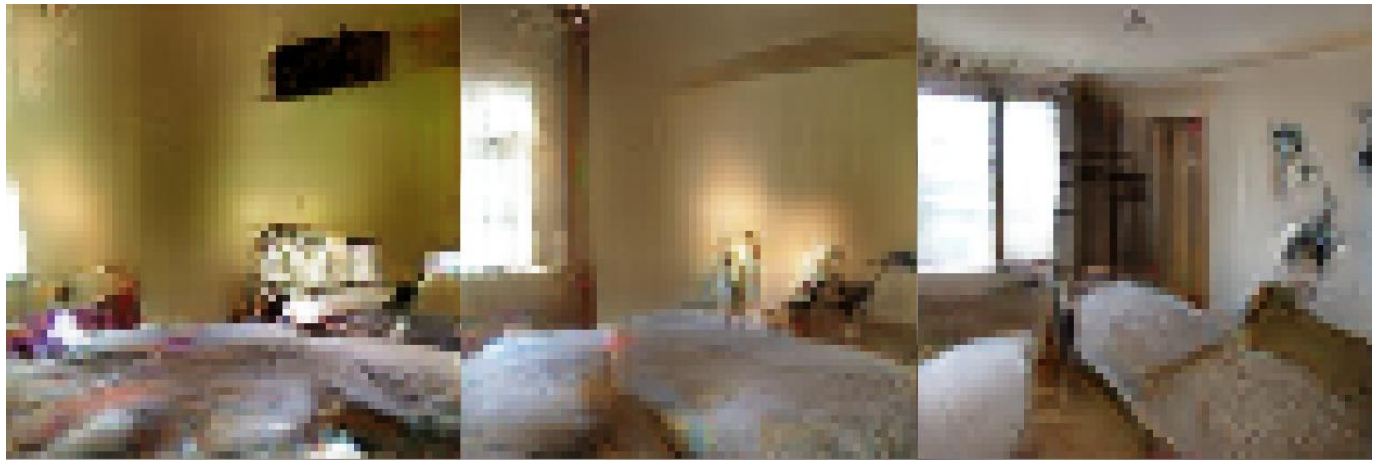
中间为特征张量；最右输出64×64×3 RGB图像。这里只展开G

Radford等 · 1511.06434v2 · Fig1 · PDF4

潜变量路径的起点、中段与终点

同一个固定G，三个不同输入，三张完整输出

z_0 → z_{mid} → z_1



三个潜变量位置与下方图像一一对应；中段仅标原第5列，不预设精确插值系数

固定 θ，改变 $\boldsymbol{z}$，分别计算 $G_\theta(\boldsymbol{z})$；每幅均为完整输出，相邻图像不作为下一幅的输入

怎样知道生成结果在接近真实数据？下一页定义来源二分类目标

Radford等 · 1511.06434v2 · Fig4第6行第1/5/10列 · PDF8

真假来源形成判别学习目标

生成结果如何获得判别误差？用来源二分类任务定义目标。

原始价值函数为

$$V(\psi,\theta)=\mathbb{E}_{\boldsymbol{x}\sim P_{\text{data}}}\ln D_{\psi}(\boldsymbol{x})+\mathbb{E}_{\boldsymbol{z}\sim p_z}\ln[1-D_{\psi}(G_{\theta}(\boldsymbol{z}))].$$

期望表示对指定分布平均；数据期望以真实样本平均估计，噪声期望以抽样平均估计。两项均按各自来源平均后相加。若两来源批大小相同，合并后统一求平均的BCE等于 J_D 额外乘 $\frac{1}{2}$；实现时应明确损失归一化约定，检查梯度尺度和优化器设置。

判别器最大化 V，等价于最小化二元交叉熵 $J_D=-V$。

真实样本希望概率升高，生成样本希望概率降低。

两种生成器目标

固定D后，G应让生成样本的来源概率向哪个方向改变?

原始minimax生成器最小化

$$J_G^{\text{sat}} = \mathbb{E}_{\boldsymbol{z}} \ln[1 - D_\psi(G_\theta(\boldsymbol{z}))].$$

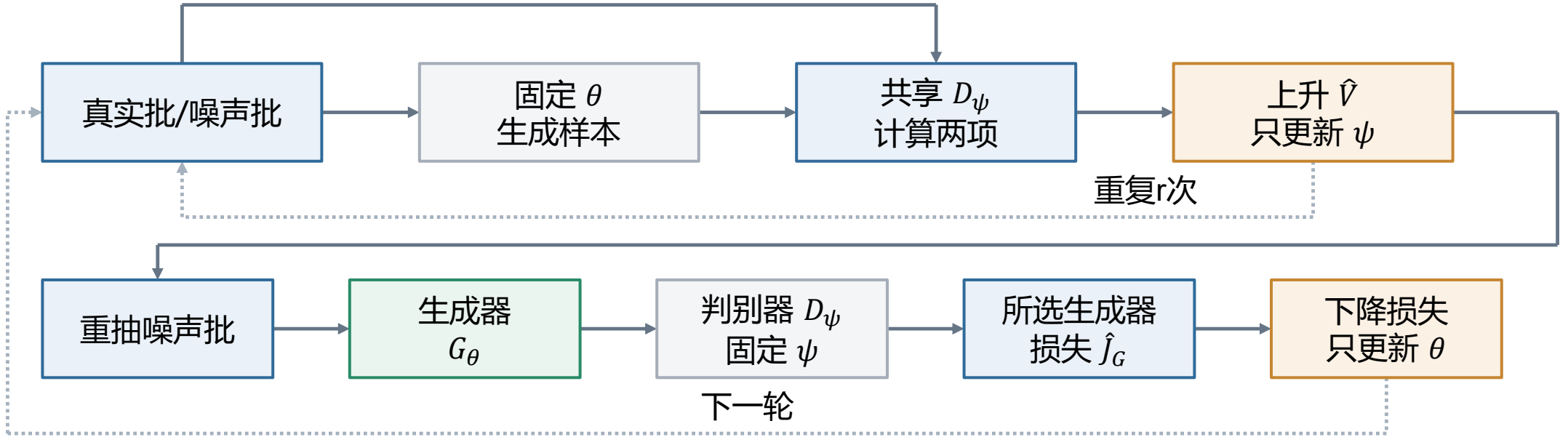
非饱和替代目标改为

$$J_G^{\text{ns}} = -\mathbb{E}_{\boldsymbol{z}} \ln D_\psi(G_\theta(\boldsymbol{z})).$$

两者都鼓励生成样本获得较高真实来源概率；替换的是G更新目标，D仍使用上述来源分类目标。

两段更新的计算与回传

先训练D，再利用固定D训练G；图中上排和下排分别只更新一组参数。



D更新固定生成参数；G更新固定判别参数，同时保留判别器对生成样本的输入梯度。

交替训练的算法步骤

怎样把两阶段更新落实为一次外层迭代?

据Goodfellow等（arXiv:1406.2661v1，Algorithm 1，PDF第4页）中文译写。输入为真实样本集、噪声先验与当前两组参数，输出为更新后的参数。取批大小 m、判别更新次数 $r \geq 1$、正学习率：

1. 重复 r 次：抽真实数据批与噪声批；固定 θ，计算生成样本及批价值 $\hat{V}$，更新 $\psi \leftarrow \psi + \eta_D \nabla_\psi \hat{V}$。
2. 重新抽噪声批；固定 ψ，原算法取 $\hat{J}_G = \hat{J}_G^{\mathrm{sat}}$，更新 $\theta \leftarrow \theta - \eta_G \nabla_\theta \hat{J}_G$。
3. 使用更新后的两组参数进入下一轮，按预设训练轮数重复。r 是训练设置，没有通用最佳比例。

原算法的 k, θ_d, θ_g 分别对应这里的 r, ψ, θ。采用前文非饱和替代时，仅把第2步损失改为 $\hat{J}_G^{\mathrm{ns}}$；这是原文§3的替代准则，不冒称Algorithm 1原式。

一对样本的判别损失

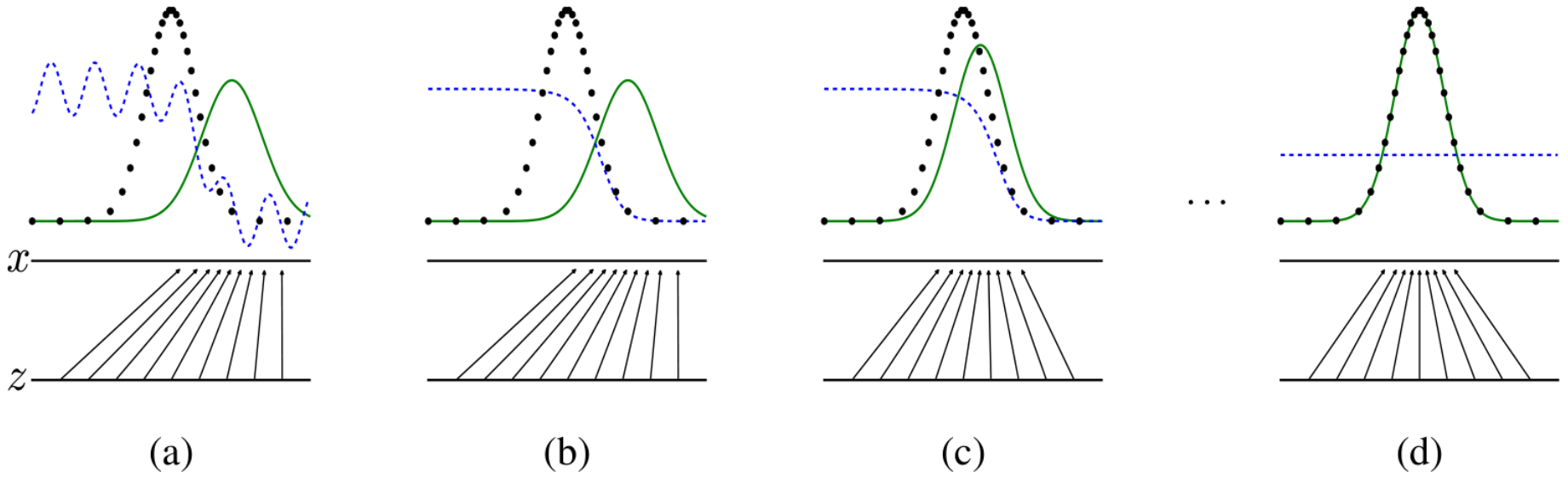
来源与标签	概率与本项损失	对本项logit的导数
真实，y=1	真实概率0.8；损失−ln0.8	-0.2
生成，y=0	真实概率0.2；损失−ln(1−0.2)	0.2

本例 $D(\boldsymbol{x}) = 0.8, D(G(\boldsymbol{z})) = 0.2$；两来源求和 $J_D = -\ln 0.8 - \ln 0.8 \approx 0.446287$

合批BCE均值：0.223144，两个导数也减半；参数梯度仍需网络链式法则；不能据此判断生成质量

原论文训练机制中哪些对象在改变

怎样从两类更新走向分布匹配？先辨三条曲线，再按(b)更新D、(c)更新G、(d)理想匹配读取。



黑点线：真实分布；绿实线：生成分布；蓝虚线：D。下方箭头表示z到G(z)的映射。

(a)为近收敛参照，非初始化。Goodfellow等，1406.2661v1，Fig1，PDF3；理论示意，非实验轨迹。

从(a)到(b)：固定G，判别器学习什么?

先认曲线

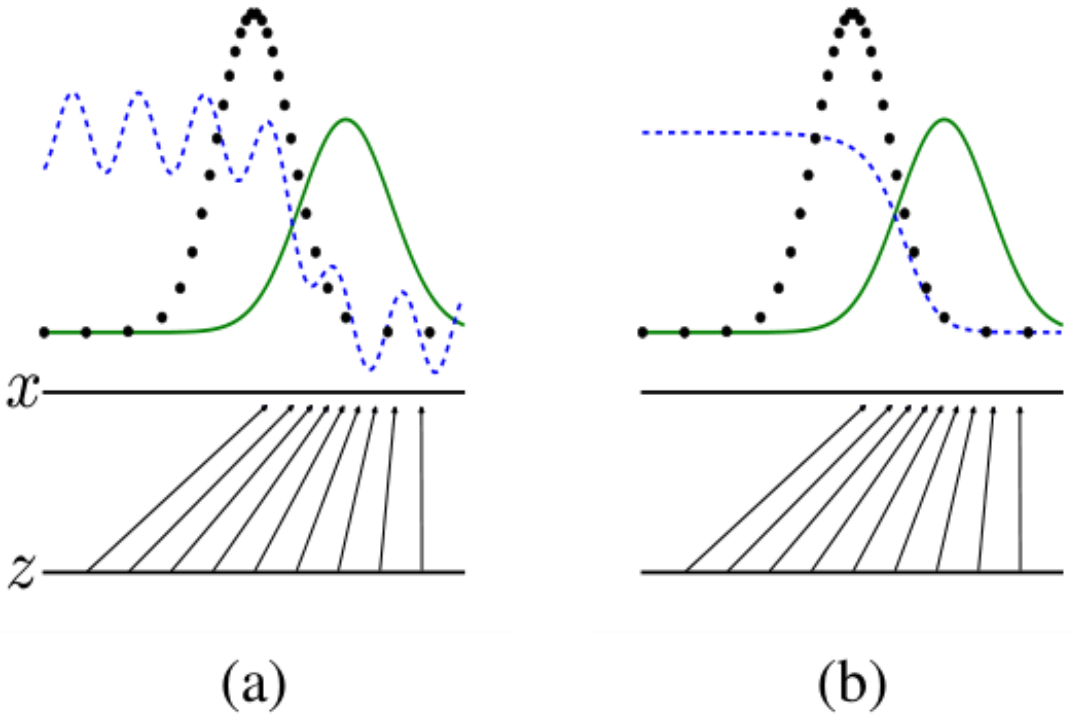
- 黑点线：真实分布
- 绿实线：生成分布
- 蓝虚线：D

固定：真实分布、G参数及下方映射

更新：只改变D；真实占优处输出较大，生成占优处较小

(a)为近收敛参照，(b)突出D更新；理论示意

Goodfellow等 · 1406.2661v1 · Fig1(a,b) · PDF3



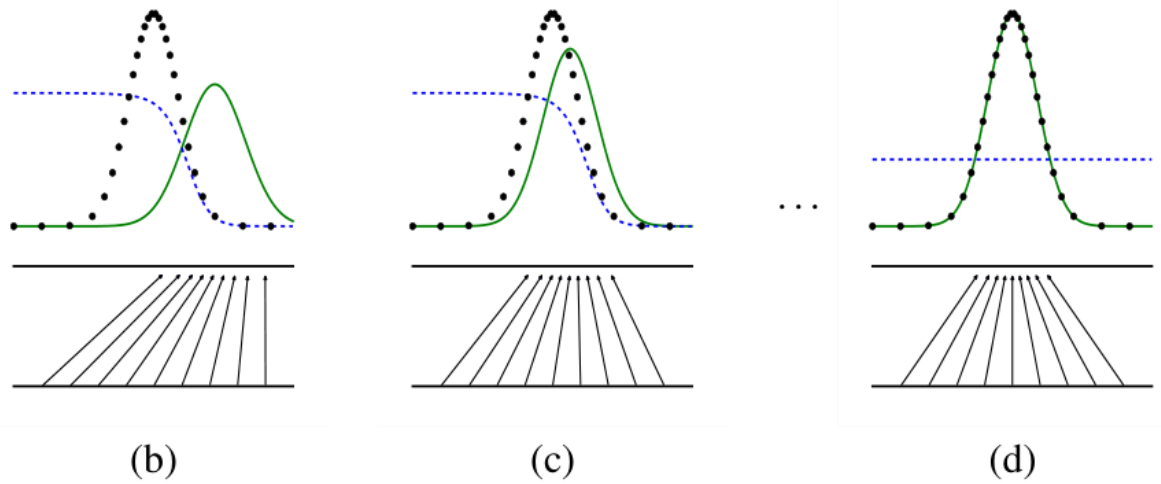
从(b)到(c)：固定D，生成映射怎样改变?

先看(b)→(c)：固定蓝虚线D；利用其输入梯度更新G。

再看映射箭头：噪声到样本的位置改变，绿生成分布随之改变。黑真实分布保持固定。

最后回到(d)：理想匹配时，绿黑重合，最优D在有质量处为1/2。

下方箭头表示映射，不是反向梯度。来源：Goodfellow等，1406.2661v1，Fig1，PDF3，裁取(b,c,d)。

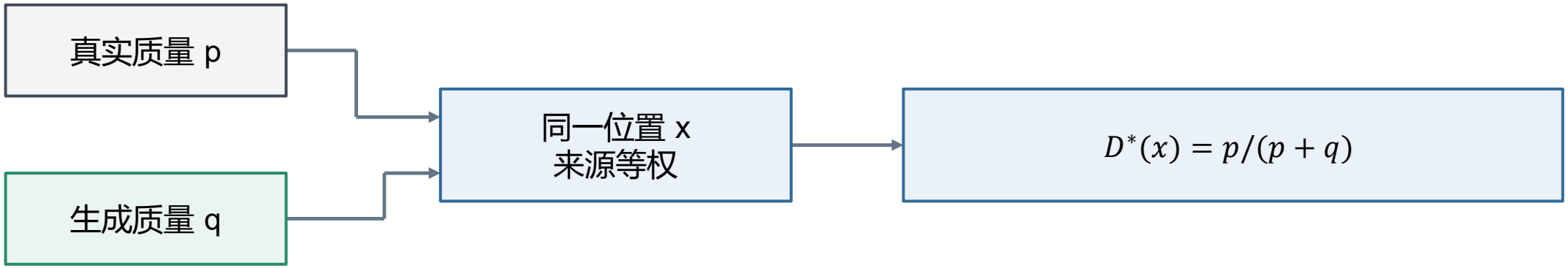


2 对抗目标与梯度

固定当前G后，理想D输出什么？当前G又得到怎样的梯度？

理想判别器读出当前位置的来源比例

固定G后，D训练到理想状态会输出什么？比较同一位置的两来源质量。



A：p=0.75，q=0.25 → D*=0.75

B：p=0.25，q=0.75 → D*=0.25

真实占比越高，最优来源概率越高；两者相等且非零时为1/2。该比例来自固定G的理想判别优化；再看最优价值比较什么分布量。

固定生成分布后逐点优化判别器

固定当前G后，理想D的输出是什么？先把目标化为逐点优化。

以下分析固定生成分布。令 p, q 分别为 P_{data}, P_g 相对于同一支配测度 μ 的密度；离散情形为概率质量。可选 $\mu = P_{\text{data}} + P_g$，不要求两者都具有Lebesgue密度。两者各自积分为1。

理想判别器可在所有可测函数 $D: \mathcal{X} \to [0,1]$ 中选择。价值函数改写为

$$V(D,G) = \int [p(\boldsymbol{x})\ln D(\boldsymbol{x}) + q(\boldsymbol{x})\ln(1 - D(\boldsymbol{x}))]d\mu.$$

第二项通过生成分布的定义把噪声上的平均改成样本上的平均。

在这个不受网络参数共享限制的函数空间中，各位置的 $D(\boldsymbol{x})$ 可以分别选择。因此固定一点，记 $a = p(\boldsymbol{x}), b = q(\boldsymbol{x})$，最大化

$$h(u) = a\ln u + b\ln(1-u), 0 \le u \le 1.$$

零权重项按 $0\ln 0 = 0$ 的连续约定处理。

最优D为什么等于来源密度比?

若 $a > 0, b > 0$，在 $0 < u < 1$:

$$h'(u) = \frac{a}{u} - \frac{b}{1-u}, h''(u) = -\frac{a}{u^2} - \frac{b}{(1-u)^2} < 0.$$

由 $h'(u) = 0$ 得 $a(1-u) = bu$，所以 $u^* = a/(a+b)$。严格凹性保证这是唯一最大值。

零密度情况分别处理:

- 若 $a > 0, b = 0$，目标随 u 增大，最大值在 $u = 1$。
- 若 $a = 0, b > 0$，最大值在 $u = 0$。
- 若 $a = b = 0$，该点不影响积分，任意输出均可。

因此在 $p + q > 0$ 处

$$D^*(\boldsymbol{x}) = \frac{p(\boldsymbol{x})}{p(\boldsymbol{x}) + q(\boldsymbol{x})}.$$

有限logit的sigmoid不能精确到0或1。端点结论属于理想函数空间或极限。

公式依赖本讲真假来源等权。改变来源先验会改变后验概率比例。

代回最优判别器并引入混合分布

把最优D代回后，G比较的是什么?

令混合密度 $m = (p+q)/2$。代回最优判别器并使用 $p+q=2m$:

$$
\begin{aligned}
C(G) &= max_D V(D,G) \\
&= \int p \ln \frac{p}{p+q} d\mu + \int q \ln \frac{q}{p+q} d\mu \\
&= \int p \ln \frac{p}{m} d\mu + \int q \ln \frac{q}{m} d\mu - 2\ln 2.
\end{aligned}
$$

最后一步中两项各减去 $\ln 2$，因为 $\int p d\mu = \int q d\mu = 1$。

两项KL给出JS关系

两个相对混合分布的KL，怎样合成JS?

KL散度为 $D_{\mathrm{KL}}(p\|r) = \int p\ln(p/r)d\mu$。**Jensen–Shannon散度**采用半权定义：

$$JS(p,q) = \frac{1}{2}D_{\mathrm{KL}}(p\|m) + \frac{1}{2}D_{\mathrm{KL}}(q\|m).$$

于是

$$C(G) = -\ln 4 + 2JS(p,q).$$

混合分布覆盖两者的支持，相关KL在此定义下有限。

自然对数下JS介于0与 $\ln 2$ 之间。

一个离散例核对目标与分布最优

同一组A/B概率，能否用两种方法算出相同价值?

在A、B两点上构造教学例，取 $p=(0.75,0.25)$、$q=(0.25,0.75)$，则 $m=(0.5,0.5)$，$D^*(A)=0.75, D^*(B)=0.25$。

$$D_{\mathrm{KL}}(p\|m)=0.75\ln 1.5+0.25\ln 0.5\approx 0.130812.$$

两向量对称交换，因此两项KL相同。

由此得到 $JS=0.130812$，并有

$$C=-\ln 4+2(0.130812)\approx -1.124670.$$

也可直接计算 $C=2(0.75\ln 0.75+0.25\ln 0.25)$，结果相同。

KL非负，且JS仅在两分布相同时为零。因此理想分布空间中的最小值为 $-\ln 4\approx -1.386294$，当 $p=q$ 时在有质量处 $D^*=1/2$。

这个结论分析最优判别函数与原始minimax价值。训练中的 D_ψ 通常未达到 D^*；有限生成器参数更新也不等于自由地优化全部分布。

最优判别器与目标的结论接口

理想D读出当前位置的真实来源占比；把它代回原始目标，就得到两分布的JS比较。

在固定 G、真假来源等权、分布使用同一支配测度的密度，且D可在理想可测函数空间中选择时：

$$D^* = \frac{p}{p+q}(p+q>0), max_D V(D,G) = -\ln 4 + 2JS(p,q).$$

这里 $p = P_{\text{data}}$ 的密度，$q = P_g$ 的密度，$m = (p+q)/2$，

$$D_{\text{KL}}(p\|r) = \mathbb{E}_p \ln(p/r), JS(p,q) = \frac{1}{2} D_{\text{KL}}(p\|m) + \frac{1}{2} D_{\text{KL}}(q\|m).$$

双方密度均零的位置任意；只一方非零时取对应端点。理想最小值在两分布相等时达到，值为 $-\ln 4$，有质量处判别输出为 $1/2$。

这个JS联系对应原始minimax的最优D价值。非饱和目标是另一生成器更新准则；有限网络、有限数据及交替梯度训练需要分别分析。

两种损失对判别概率的导数不同

损失对概率、logit和参数的导数，为什么要分开?

对一个噪声样本，令判别器logit为 $a = a_\psi(G_\theta(\boldsymbol{z}))$，输出 $d = \sigma(a) = 1/(1+e^{-a})$。生成器的两种单样本损失为

$$\ell_{\text{sat}}(d) = \ln(1-d), \ell_{\text{ns}}(d) = -\ln d.$$

二者均最小化。先对概率 d 求导:

$$\frac{d\ell_{\text{sat}}}{dd} = -\frac{1}{1-d}, \frac{d\ell_{\text{ns}}}{dd} = -\frac{1}{d}.$$

当 d 很小时，前者接近 -1，后者绝对值很大。但网络通常通过logit计算概率，完整梯度还包含sigmoid导数；只比较这两个概率导数不足以解释参数更新。

链式法则给出logit上的饱和差异

再乘sigmoid导数后，局部信号会怎样改变?

由 $\sigma'(a) = d(1-d)$，逐项应用链式法则：

$$\frac{d\ell_{\text{sat}}}{da} = -\frac{1}{1-d}\,d(1-d) = -d,$$

$$\frac{d\ell_{\text{ns}}}{da} = -\frac{1}{d}\,d(1-d) = d - 1.$$

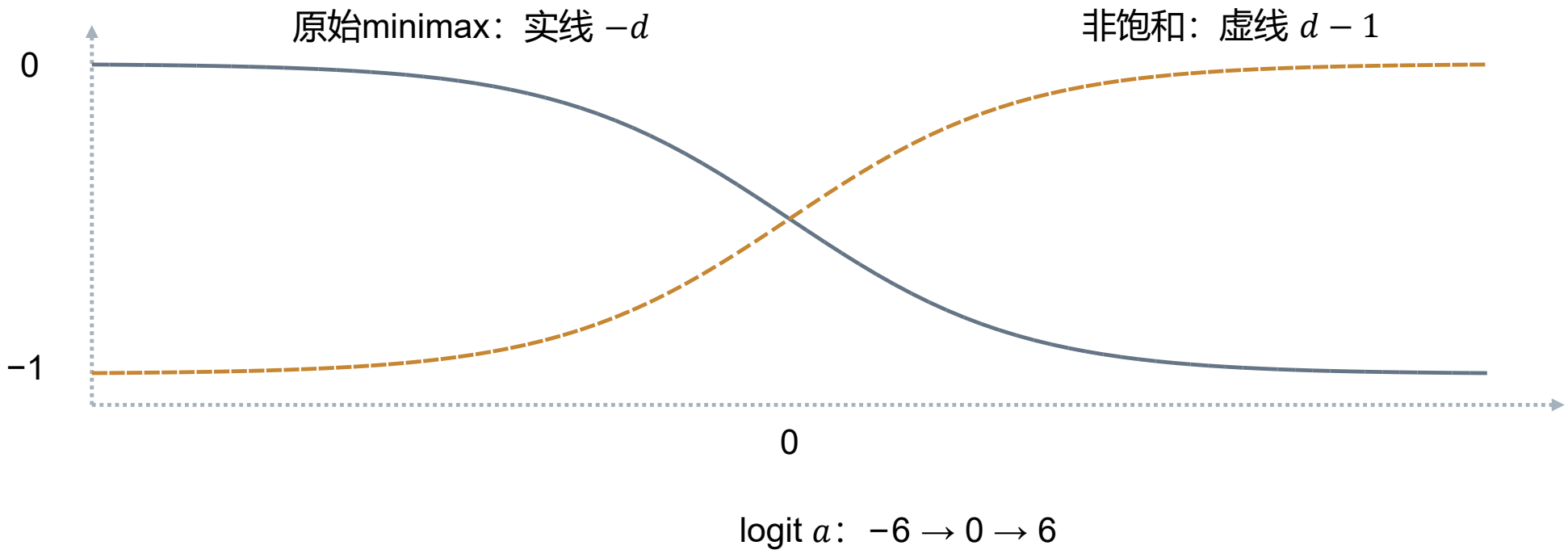
当判别器很确信生成样本为假，$d \to 0$，原始损失的logit导数趋于0，非饱和损失的logit导数趋于 -1。两者梯度下降都倾向提高该样本的logit，局部强度却不同。

取 $a = -4$，有 $d \approx 0.017986$。两种logit导数分别约为 -0.017986 与 -0.982014。非饱和目标改善的是这一标量环节，尚未评价网络其余Jacobian。

判为假时，非饱和目标保留较强logit信号

固定D，a为sigmoid前分数，$d = \sigma(a)$；纵轴为损失对 a 的导数

两种待最小化损失：$\ell_{\text{sat}} = \ln(1-d)$、$\ell_{\text{ns}} = -\ln d$



左端：两者都促使logit升高，但原始目标导数趋0，非饱和导数趋−1。完整参数梯度还要经过两个网络的导数。

完整参数梯度的链式结构

标量logit信号还要经过什么，才能到达生成器参数?

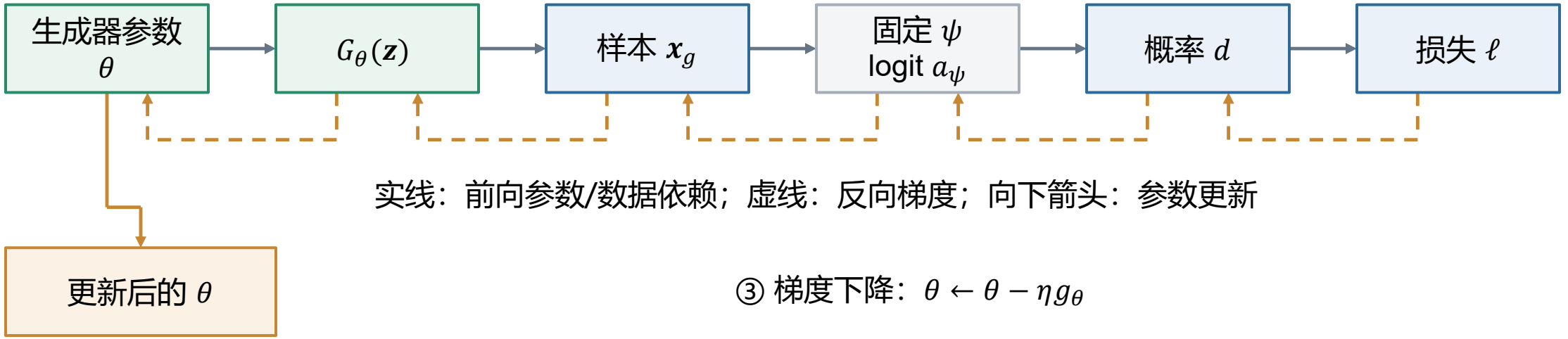
记 $\boldsymbol{x}_g = G_\theta(\boldsymbol{z})$。在可微处，链式法则给出

$$\nabla_\theta \ell = \left(\frac{\partial G_\theta}{\partial \theta}\right)^{\top} \nabla_{\boldsymbol{x}_g} a_\psi(\boldsymbol{x}_g) \frac{d\ell}{da}.$$

生成器Jacobian的形状为 $\dim(\boldsymbol{x}_g)\times\dim(\theta)$，logit对输入的梯度为 $\dim(\boldsymbol{x}_g)$ 维列向量，因此结果与参数形状一致。

固定判别参数仍保留输入梯度

D参数固定，梯度还能通过D吗？沿三步路径追踪。



图中记 $g_\theta = \nabla_\theta \ell$。反向依次经过D输入导数与G参数导数，再只更新θ。非饱和标量因子较大，仍可能遇到输入梯度为零、生成器Jacobian退化、方向不利或优化波动。固定D参数意味着不更新 ψ，并不取消 $\nabla_{\boldsymbol{x}_g} a_\psi$。

三层演算：从局部导数到参数更新

第1层：标量导数为−d与d−1。采用标量教学构造 $G_\theta(z)=\theta+z$、$a_\psi(x)=2x-4$。固定 $z=0,\theta=0$，因此 $a=-4,d=0.017986$，两网络的导数为 $dG/d\theta=1,da/dx=2$。

第2层：乘输入导数2与生成器导数1，得到参数梯度。

$$\frac{d\ell_{\text{sat}}}{d\theta}=-2d\approx-0.035972,\frac{d\ell_{\text{ns}}}{d\theta}=2(d-1)\approx-1.964028.$$

第3层：分别从同一初值执行学习率0.1的一步梯度下降，参数成为 0.003597 或 0.196403。这是固定判别器的一步演算，不是训练收敛实验。

结论：对于sigmoid判别器，两种G损失的logit导数分别为 $-d$、$d-1$；它们还需乘判别器输入导数与生成器参数Jacobian。非饱和目标在 $d\approx 0$ 时保留较强logit信号，完整梯度有效性仍有条件。两者的目标定义见前文，不能用原始JS恒等式替代非饱和梯度分析。

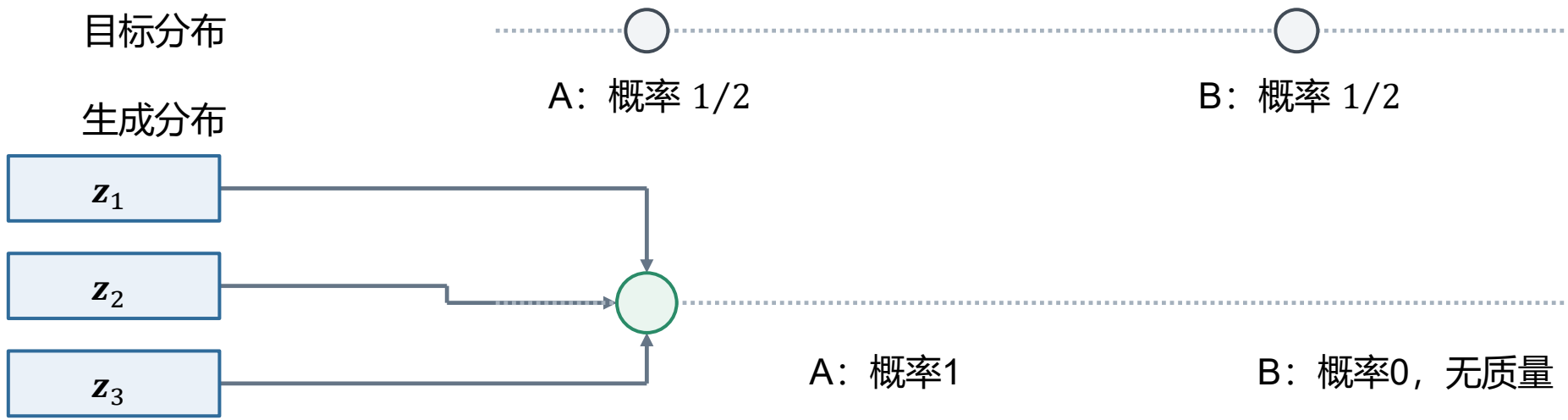
3 训练动态与分布覆盖

有了更新信号，为什么仍需检查质量、覆盖与判别能力？

样本清晰与分布覆盖分别判断

模式坍塌指生成分布集中到较少结果或模式，不能充分表达目标分布的多样性。许多不同噪声映射到相近输出，是一种直观表现。

构造一个双模式教学例，目标在 $A = (-1,-1)$、$B = (1,1)$ 各有概率 $1/2$。生成器却对全部噪声输出A。生成的每个样本都位于真实模式上，但B始终缺失。



质量、覆盖和记忆是不同问题。训练样本上重现清晰图像不能单独说明多样性；判别训练本身也不提供未记忆数据的保证。具体指标和数据检查交由GM08。

移动目标使训练损失需要结合条件解释

看到D输出接近1/2，能否断定训练成功?

每次更新G都会改变判别训练中的生成分布，每次更新D也会改变G接收到的函数。损失沿两个不断变化的目标计算，不能按单一固定目标的下降曲线解释整个过程。

若观察到 $D_\psi(\boldsymbol{x}) \approx 1/2$，至少有两种可能：

- 两分布已匹配，且D有足够能力并已学到相应判别函数。
- D没有学会区分。例如常数判别器 $D \equiv 1/2$ 对任意两分布都给相同输出。

原始理想分析需要固定G时D达到最优，并在分布空间按适当条件改进目标。有限参数网络、随机批、交替次数和优化器可能改变训练动态。损失数值、判别准确率、样本图与覆盖证据应联合解释；本讲不据其中单项宣称训练成功。

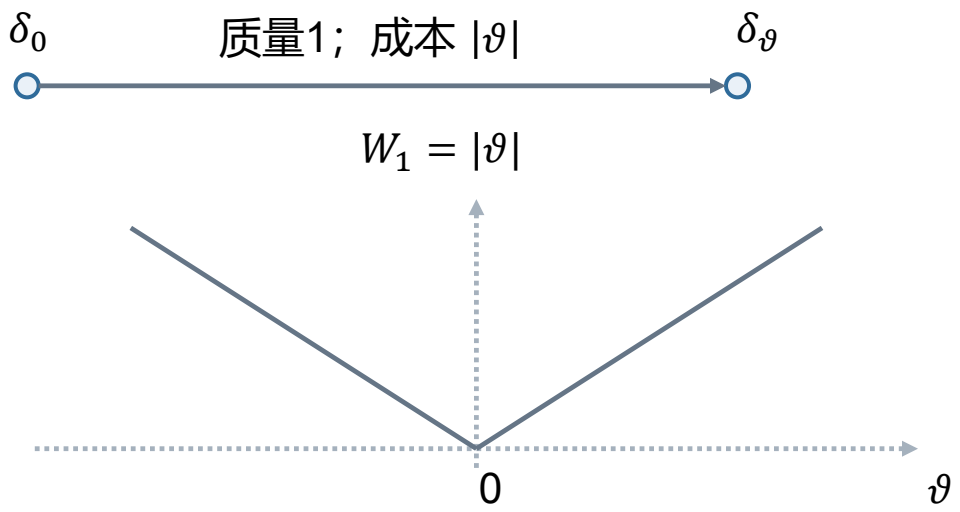
接下来问：当两分布的支持不重叠，JS能否反映它们在空间中正在接近？先看移动点质量，再引入几何运输。

4 Wasserstein目标与约束

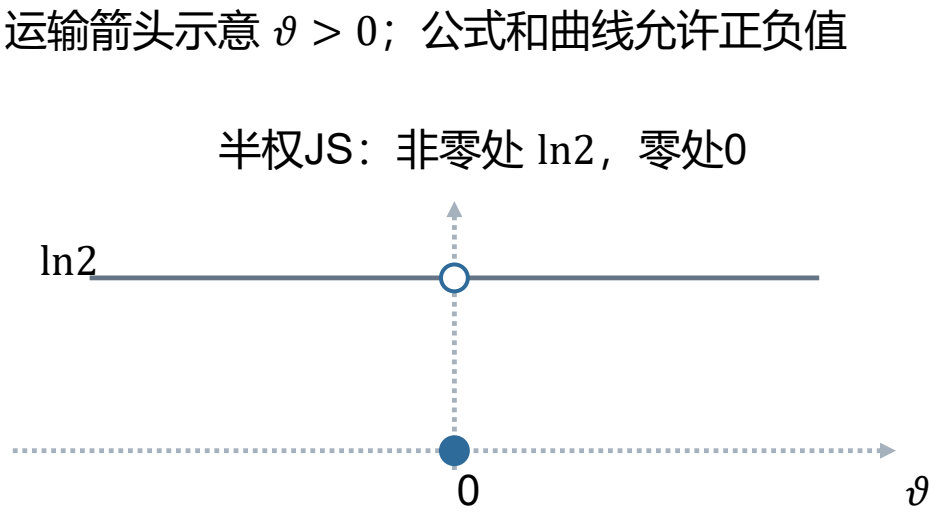
当支持不重叠时，怎样让分布比较反映空间接近？

点质量显示几何间距

距离正在缩短时，比较量是否随之改变？构造点质量教学例，取 $P = \delta_0, Q_\vartheta = \delta_\vartheta$，分别集中在0和 ϑ。唯一运输方案把全部质量从0移到 ϑ，故 $W_1 = |\vartheta|$。若 $\vartheta \neq 0$，按本讲半权JS定义，两点支持分离，JS恒为 $\ln 2$；在 $\vartheta = 0$ 两者均为0。



运输箭头示意 $\vartheta > 0$；公式和曲线允许正负值



分布间几何距离趋近时Wasserstein可随之变小。该例只说明两种比较量的差别，不证明任意GAN训练具有相同梯度行为。

耦合与最小运输成本

相同的边缘分布，为什么仍要选择运输方案?

教学构造：$P = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_2$，$Q = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_3$

方案A：相邻运输

	$y=1$	$y=3$	行和
$x=0$	0.5	0	0.5
$x=2$	0	0.5	0.5
列和	0.5	0.5	总量1

$C_A = 0.5|0-1| + 0.5|2-3| = 1$

方案B：交叉运输

	$y=1$	$y=3$	行和
$x=0$	0	0.5	0.5
$x=2$	0.5	0	0.5
列和	0.5	0.5	总量1

$C_B = 0.5|0-3| + 0.5|2-1| = 2$

每格是从x运到y的质量；行和恢复P、列和恢复Q。**耦合** $\gamma \in \Pi(P,Q)$ 正是边缘为 P, Q 的联合分布；运输方案不同，成本可以不同

Wasserstein-1距离在所有可行耦合中选择最小平均成本：

$W_1(P,Q) = inf_{\gamma \in \Pi(P,Q)} \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim\gamma} \|\boldsymbol{x}-\boldsymbol{y}\|$.

对偶形式由受约束的critic提供分数

怎样把运输比较转成可以训练的函数目标?

在紧致度量空间上的概率分布情形，可引用Kantorovich–Rubinstein对偶关系：

$$W_1(P_{\text{data}}, P_g) = sup_{f:Lip(f)\leq 1}[\mathbb{E}_{\text{data}} f(\boldsymbol{x}) - \mathbb{E}_{\boldsymbol{z}} f(G_\theta(\boldsymbol{z}))].$$

1-Lipschitz条件指任意两点满足

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq \|\boldsymbol{x} - \boldsymbol{y}\|.$$

它限制函数变化速度。

critic f_w 输出实数分数，无需sigmoid，也不是来源概率。

本讲引用对偶定理，不证明完整最优传输理论。

实际critic在受限网络族中最大化分数差；G固定critic后最小化

$$J_G^W = -\mathbb{E}_{\boldsymbol{z}} f_w(G_\theta(\boldsymbol{z})).$$

这相当于提高生成样本分数。对偶把搜索运输计划转为训练受约束函数，便于用网络与批平均实现。

有限函数族、约束近似与未充分优化使训练分数差通常只是距离的近似。

若精确满足1-Lipschitz，则任意可行函数的真实分数差不超过 W_1。

权重裁剪近似实施critic约束

Lipschitz约束在训练的哪一步近似实施?

原始WGAN：训练循环

1. 固定G，多次抽真实批与噪声批，增大critic批分数差
2. 每次critic更新后，把权重裁剪到给定区间
3. 固定critic，重抽噪声，按 $-\mathbb{E} f_w(G_\theta(\boldsymbol{z}))$ 更新G；重复

约束近似：三个限制

- 给定网络/激活条件下，有界权重可限制某个全局Lipschitz常数
- 裁剪不精确覆盖全部1-Lipschitz函数，也不自动校准到1
- 裁剪可能限制表达能力并损害梯度

后续方向：WGAN-GP

以critic输入梯度范数软惩罚替代裁剪；不等于全空间精确1-Lipschitz

Gulrajani等 · 1704.00028v3 · §3—4

可行critic与最优critic给出不同梯度

只满足约束，是否足以给G有效梯度？先比较两个可行critic。

取 $\vartheta > 0$ 的同一点质量例，比较两个均满足1-Lipschitz的critic：

- $f(x) = -x$ 可行且最优，真实分数为0，生成分数为 $-\vartheta$，差为 $\vartheta = W_1$。G损失为 ϑ，其导数为1，梯度下降可向0移动。
- 常数函数 $f(x) = 0$ 同样可行，却给出零梯度。

因此Lipschitz仅约束变化上界。理想最优critic的梯度关系需要最优性、可微性及期望求导等正则条件。

近似训练不能由此获得必然收敛、消除坍塌或处处非零梯度的保证。

5 条件生成

前面改变无条件训练方式；现在改变任务要求：按指定条件生成

条件目标比较配对数据与生成对

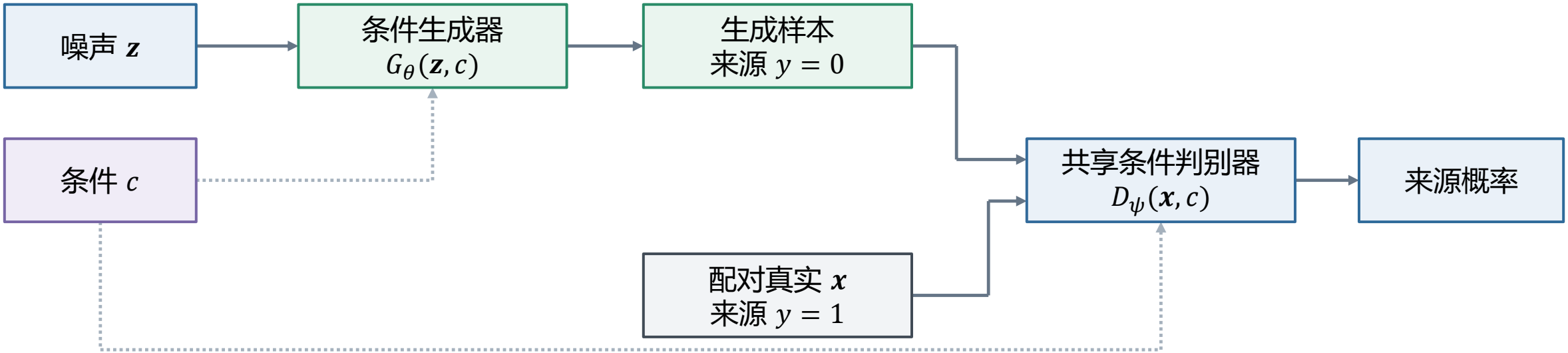
前面改变训练准则；现在改变任务要求：给定c生成什么？

条件GAN（cGAN）学习给定条件 c 的生成分布。条件可为类别或其他已知信息，噪声仍提供条件内的随机性。真实训练对来自 $P_{\text{data}}(\boldsymbol{x}, c)$。构造生成对时，从真实数据的条件边缘 $P_{\text{data}}(c)$ 抽取 c，再独立抽取噪声：

$$V_c = \mathbb{E}_{(\boldsymbol{x},c)\sim P_{\text{data}}} \ln D_\psi(\boldsymbol{x}, c) + \mathbb{E}_{c\sim P_{\text{data}}(c),\boldsymbol{z}\sim p_z} \ln[1 - D_\psi(G_\theta(\boldsymbol{z}, c), c)].$$

条件同时进入生成器和判别器

同一个条件c为什么需要同时进入G和D?



反例：c等概率；真实 $x=c$，错误 $G(z,c)=1-c$（此构造忽略噪声z）

$c=0$：真实 $x=0$，错误输出 $x=1$

$c=1$：真实 $x=1$，错误输出 $x=0$

两来源的x边缘均为0、1各半：D只看x无法发现条件颠倒；D看(x,c)才能按配对关系区分

条件输入不保证实际训练成功；普通目标也未自动加入专门的错配负样本

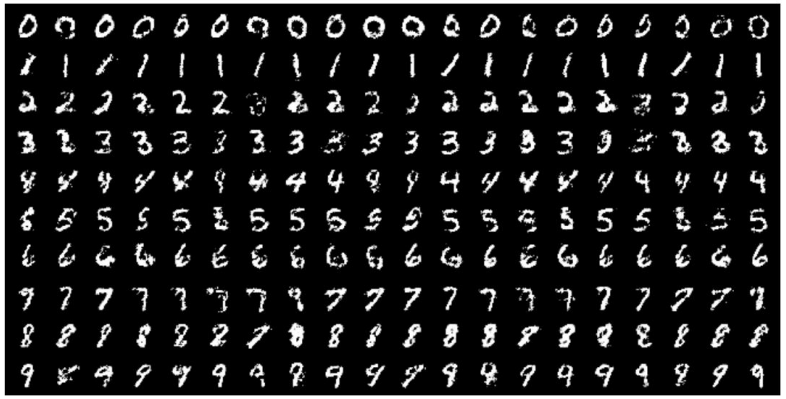
固定类别条件，观察不同生成样本

给定类别后，生成结果怎样变化？先看全图，再看局部

右：论文完整结果，每行固定一个条件

下：原图前三行放大，沿行比较笔画

同一类别，多个生成样本





c指定类别，z提供随机输入，y标记真假来源；样本图不能单独认证覆盖

Mirza与Osindero · 1411.1784v1 · Fig2 · PDF4；局部为原图前3行

条件采样与架构应用的职责

固定训练后的G，如何按指定条件反复采样?

条件生成依次执行:

1. 给定所需条件 c，并确认模型训练支持该条件范围。
2. 抽取 $\boldsymbol{z}\sim p_z$。
3. 输出 $G_\theta(\boldsymbol{z},c)$；重复抽噪声可探索同一条件下的结果差异。

类别条件例中，固定c、重抽噪声可得到同类别的不同输出；改变c则提出新的生成要求。cGAN增加的是条件输入接口，具体网络仍需相应训练。

这里形成的是条件变量的接口。条件扩散、分类器或无分类器引导和潜在生成系统由GM06建立；条件GAN本身不提供这些算法。样本质量、覆盖、条件符合度与记忆的完整评价交给GM08。

本讲总结

学习信号怎样连接训练、理论、梯度与扩展?

来源判别信号
G生成样本，D区分来源

交替更新
固定D参数，保留输入梯度

训练后使用
固定G，抽噪声得完整输出

理想D与原始价值：固定G、来源等权
D*：$p/(p+q)$
C(G)：$-\ln 4 + 2JS$

sigmoid判别器的logit导数
原始：$-d$，非饱和：$d-1$
还需D输入导数与G Jacobian

从核心机制区分：结果证据、训练比较、任务要求

结果评价
质量 / 覆盖 / 记忆
分别需要证据

训练比较扩展：WGAN
受约束critic近似运输
约束不保证有效梯度

任务要求扩展：cGAN
c同时进入G与D
c / z / 真假标签各司其职

理想最优、当前参数梯度和实际结果分别判断；Lipschitz只给函数变化上界