An alternative view of GAN

Consider the following process:

① the machine flips a (possibly bent) coin

② depending on the outcome:

1) if heads, the machine shows a <u>real</u> image

2) if tails, the machine shows a <u>generated</u> image

This process is repeated several times.

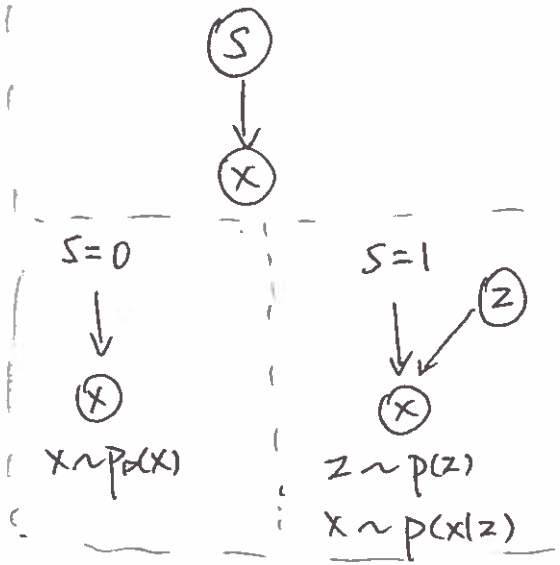
You're asked to guess the probability that the coin flip outcome is tail.

What would the generative model do to fool you on this task?

This process is equivalent to the following <u>augmented generative model</u>

$$s \sim \tilde{p}(s) = \pi^{s}(1-\pi)^{1-s}$$

$$x|s \sim \tilde{p}(x|s) = \begin{cases} p(x), & s=1 \\ p_0(x), & s=0 \end{cases}$$



We train this model by

$$\min_{p(x)} I[s;x] = KL[\tilde{p}(s,x) \| \tilde{p}(s)\tilde{p}(x)]$$

when $I[s;x]=0$:

observation x tells us nothing about the coin flip outcome s (so we can only do random guess).

Notes: ① $I[s;x]=0 \iff p(x)=p_0(x)$

② when $\pi=\frac{1}{2}$, $I[s;x] = JS[p(x)\|p_0(x)]$

$$\min_{p(x)} I[s;x] \qquad \text{(intractable)}$$

Bringing in the idea of variational approximation:

(similar to variational information maximization)

$$I[s;x] \geq I[s;x] - \mathbb{E}_{\hat{p}(x)}\left[KL[\hat{p}(s|x) \| q(s|x)]\right]$$

$$= H[s] - H[s|x] + H[s|x] + \mathbb{E}_{\hat{p}(x,s)}[\log q(s|x)]$$

$$= H[s] + \mathbb{E}_{\hat{p}(x,s)}[\log q(s|x)]$$

Now the objective becomes

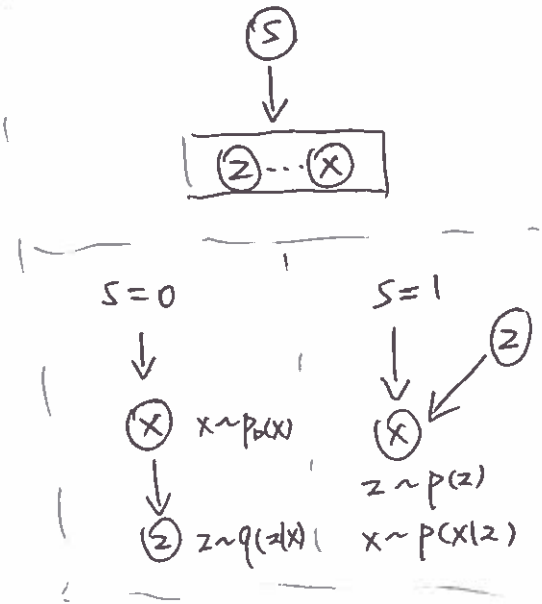
$$\min_{p(x)} \max_{q(s|x)} H[s] + \mathbb{E}_{\hat{p}(x,s)}[\log q(s|x)] \qquad \text{(tractable)}$$

$$\mathbb{E}_{\hat{p}(x,s)}[\log q(s|x)] = \underbrace{(1-\pi)\,\mathbb{E}_{p_0(x)}[\log q(s=0|x)] + \pi\,\mathbb{E}_{p(x)}[\log(1-q(s=0|x))]}_{= \text{GAN's objective (when } \pi = \frac{1}{2})}$$

ALI as an augmented generative model

$$s \sim \hat{p}(s) = \pi^{s}(1-\pi)^{1-s}$$

$$x, z \mid s \sim \hat{p}(x,z|s) = \begin{cases} p(z)p(x|z), & s=1 \\ p_0(x)\,q(z|x), & s=0 \end{cases}$$



we train this model by

$$\min_{p(x,z),\,q(z|x)} I[s;\{x,z\}]$$

Notes: under this model:

① $I[s;\{x,z\}] = I[s,x] + I[s,z|x]$

$$I[s,z|x] = \mathbb{E}_{\hat{p}(x)}\left[KL\left[\hat{p}(s,z|x) \,\|\, \hat{p}(s|x)\hat{p}(z|x)\right]\right]$$

So $I[s,z|x] = 0 \iff p(z|x) = q(z|x)$

so min. over $q(z|x)$ make sense!

To see why $I[s,z|x]=0 \iff p(z|x)=q(z|x)$:

$$I[s,z|x]=0 \iff \hat{p}(s,z|x)=\hat{p}(s|x)\hat{p}(z|x)$$

$$\hat{p}(s,z|x)=\hat{p}(s|x)\boxed{\hat{p}(z|s,x) = \begin{cases} q(z|x), & s=0 \\ p(z|x), & s=1 \end{cases}}$$

$$\hat{p}(z|x)=\hat{p}(s=0|x)q(z|x)+\hat{p}(s=1|x)p(z|x)$$

$$\hat{p}(z|x)=\hat{p}(z|s,x),\ s=0,1 \iff p(z|x)=q(z|x)$$

② ALI's objective function:

$$\min_{p(x|z),q(z|x)} \max_{q(s|x,z)} I[s;\{x,z\}]-\mathbb{E}_{\hat{p}(x,z)}\left[KL[\hat{p}(s|x,z)\|q(s|x,z)]\right]$$

$$= H[s]+\mathbb{E}_{\hat{p}(s,x,z)}[\log q(s|x,z)]$$