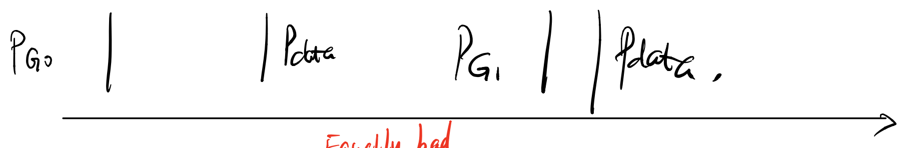
JSD 的问题，往往 P_G 和 P_{data} 没有重合

JSD is log2 if two distributions do not overlap



Equally bad

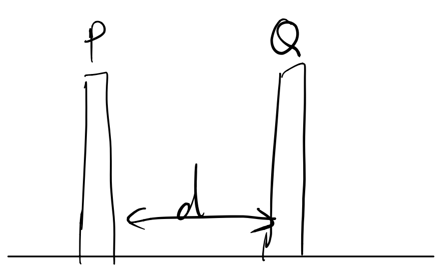
$JS(P_{G_0}, P_{data}) = \log 2$

$JS(P_{G_1}, P_{data}) = \log 2$.

⇒ Same objective value is obtained 不容易训练 G!

WGAN.

把 JSD 换成 Earth Mover's Distance



$W(P, Q) = d$

$P_{G_0} \| P_{data}$ $P_{G_{50}} \| P_{data}$

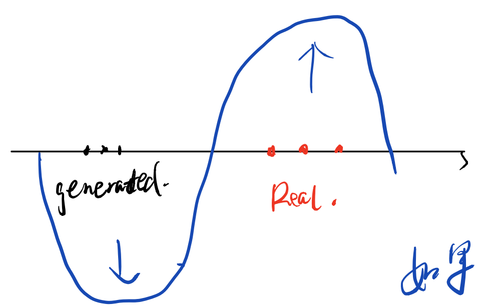
$W(P_{G_0}, P_{data}) = d_0$ $W(P_{G_{50}}, P_{data}) = d_{50}$

如何修改 discriminator ?

$$V(G,D) = \max_{D \in 1\text{-Lipschitz}} \{ E_{x\sim P_{data}}[D(x)] - E_{x\sim P_G}[D(x)] \}$$

(对于 JSD, E里面是有 log 的)

D has to be smooth enough



如果不加限制，会趋于 ∞，停不下来.

Lipschitz Function: $\|f(x_1) - f(x_2)\| \le K \|x_1 - x_2\|$

$K=1$ for 1-Lipschitz

实际解决方案 ① weight clipping

if $w > c$, $w = c$

if $w < -c$, $w = c$

② WGAN-GP

$D \in 1\text{-Lipschitz} \iff \|\nabla_x D(x)\| \le 1$ for all x

$$V(G, D) = \max_D \{ E_{x \sim P_{data}}[D(x)] - E_{x \sim P_G}[D(x)]$$

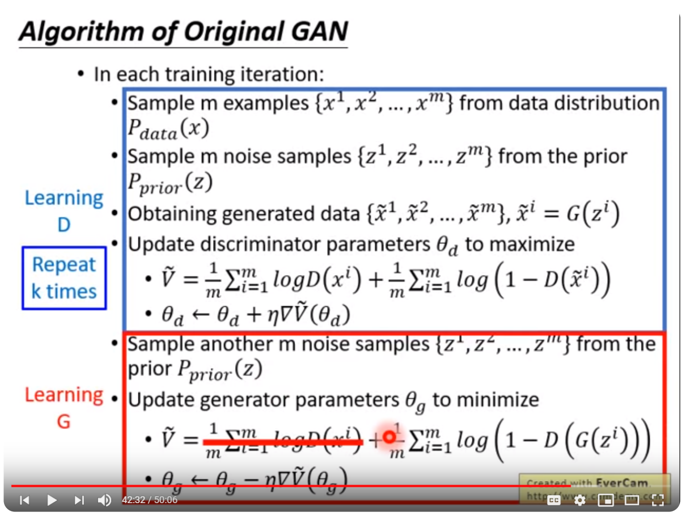
$$- \lambda \int_x \cancel{\max(0, \|\nabla_x D(x)\| - 1) dx} \}$$

$\downarrow$

$$E_{x \sim P_{penalty}}[\cancel{\max(0, \|\nabla_x D(x)\| - 1)}]\}$$

$\downarrow$

$$(\|\nabla_x D(x)\| - 1)^2$$



Algorithm of WGAN

- In each training iteration:

No sigmoid for the output of D

Learning D

Repeat k times

- Sample m examples $\{x^1, x^2, \dots, x^m\}$ from data distribution $P_{data}(x)$
- Sample m noise samples $\{z^1, z^2, \dots, z^m\}$ from the prior $P_{prior}(z)$
- Obtaining generated data $\{\tilde{x}^1, \tilde{x}^2, \dots, \tilde{x}^m\}$, $\tilde{x}^i = G(z^i)$
- Update discriminator parameters θ_d to maximize
  - $\tilde{V} = \frac{1}{m}\sum_{i=1}^{m} D(x^i) - \frac{1}{m}\sum_{i=1}^{m} D(\tilde{x}^i)$
  - $\theta_d \leftarrow \theta_d + \eta \nabla \tilde{V}(\theta_d)$

Weight clipping / Gradient Penalty ...

Learning G

Only Once

- Sample another m noise s } from the prior $P_{prior}(z)$
- Update generator parameters θ_g to minimize
  - $\tilde{V} = \cancel{\frac{1}{m}\sum_{i=1}^{m} logD(x^i}) - \frac{1}{m}\sum_{i=1}^{m} D\left(G(z^i)\right)$
  - $\theta_g \leftarrow \theta_g - \eta \nabla \tilde{V}(\theta_g)$