WGAN & WGAN-GP

GAN: ONE MORE STEP FURTHER

BEFORE USING 1-WASSERSTEIN

- EMD에서 $\Pi(p_r, p_g)$ 의 모든 결합 분포를 추적하기는 불가능
- Kantorovich-Rubinstein duality를 사용:
- $$W(p_r, p_g) = \inf_{\gamma \sim \Pi(p_r, p_g)} \mathbb{E}_{(x,y)\sim\gamma} [\|x - y\|]$$ 대신

$$W(p_r, p_g) = \frac{1}{K} \sup_{\|f\|_L \leq K} \mathbb{E}_{x\sim p_r}[f(x)] - \mathbb{E}_{x\sim p_g}[f(x)]$$

  - Where $K = 1$

BEFORE USING 1-WASSERSTEIN

- $W(p_r, p_g) = \frac{1}{K} \sup_{\|f\|_L \leq K} \mathbb{E}_{x \sim p_r}[f(x)] - \mathbb{E}_{x \sim p_g}[f(x)]$

- f는 $||f||_L \leq K$를 만족해야 함
  - K-lipschitz continuous를 만족한다는 의미
  - 모든 $(x_1, x_2) \in R^2$와 Lipschitz 상수 K에 대해
  - $|f(x_1) - f(x_2)| \leq K|x_1 - x_2|$를 만족
  - 거의 모든 점에서 연속적으로 미분 가능

BEFORE USING 1-WASSERSTEIN

- $W(p_r, p_g) = \frac{1}{K} \sup_{\|f\|_L \leq K} \mathbb{E}_{x \sim p_r}[f(x)] - \mathbb{E}_{x \sim p_g}[f(x)]$

- $K = 1$ 이므로 f는 $||f||_L \leq 1$ 를 만족해야 함
  - $|f(x_1) - f(x_2)| \leq |x_1 - x_2|$
  - $\frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} \leq 1$
  - f는 임의의 두 점 사이 변화율이 1을 넘지 않는 함수

BEFORE USING 1-WASSERSTEIN

- $W(p_r, p_g) = \frac{1}{K} \sup_{\|f\|_L \leq K} \mathbb{E}_{x \sim p_r}[f(x)] - \mathbb{E}_{x \sim p_g}[f(x)]$

- $KR(p, q) \leq W(p, q)$ 이므로
- sup는 inf의 반대
  - 상한(upper bound)

BEFORE USING 1-WASSERSTEIN

- $$W(p_r, p_g) = \frac{1}{K} \sup_{\|f\|_L \leq K} \mathbb{E}_{x \sim p_r}[f(x)] - \mathbb{E}_{x \sim p_g}[f(x)]$$

- f를 구하면 EMD를 계산할 수 있음
  - f를 구하기도 불가능
  - 그러나 근사는 훨씬 쉬움 … NN으로 근사

$$\max_{w \in W} \mathbb{E}_{x \sim p_r}[f_w(x)] - \mathbb{E}_{z \sim p_r(z)}[f_w(g_\theta(z))]$$

BEFORE USING 1-WASSERSTEIN

- $L(p_r, p_g) = W(p_r, p_g) = \max_{w \in W} \mathbb{E}_{x \sim p_r}[f_w(x)] - \mathbb{E}_{z \sim p_r(z)}[f_w(g_\theta(z))]$
  - 파라미터(판별기의 가중치) w에 대하여
  - NN인 f_w를 업데이트하며 근사

BEFORE USING 1-WASSERSTEIN

- $L(p_r, p_g) = W(p_r, p_g) = \max_{w \in W} \mathbb{E}_{x \sim p_r}[f_w(x)] - \mathbb{E}_{z \sim p_r(z)}[f_w(g_\theta(z))]$
  - $||f||_L \leq 1$ 제약이 남음
  - 가중치 w를 [-0.01, 0.01]로 제한
  - 공간 W 가 compact parameter 공간이 됨
  - f_w에 상한과 하한이 생기면서 제약을 만족
- Terrible way to enforce a Lipschitz constraint

BEFORE USING 1-WASSERSTEIN

- Weight clipping
  - 가중치 w를 [-0.01, 0.01]로 제한
  - 공간 W 가 compact parameter 공간이 됨
  - f_w에 상한과 하한이 생기면서 제약을 만족
- Terrible way to enforce a Lipschitz constraint

WGAN

LOSS FUNCTION

- $\max_{w \in W} \mathbb{E}_{x \sim p_r}[f_w(x)] - \mathbb{E}_{z \sim p_r(z)}[f_w(g_\theta(z))]$

  을 손실 함수의 기반으로 사용

  - 판별기 손실 함수
  - 생성기 손실 함수

LOSS FUNCTION

- $\max_{w \in W} \mathbb{E}_{x \sim p_r}[f_w(x)] - \mathbb{E}_{z \sim p_r(z)}[f_w(g_\theta(z))]$
- 판별기 손실 함수
  - $-W(p_{data}, p_g)$를 최소화
  - $L^{(D)} = -\mathbb{E}_{x \sim p_{data}} D_w(x) + \mathbb{E}_z D_w(G(z))$

LOSS FUNCTION

- $\max_{w \in W} \mathbb{E}_{x \sim p_r}[f_w(x)] - \mathbb{E}_{z \sim p_r(z)}[f_w(g_\theta(z))]$
- 생성기 손실 함수
  - $L^{(G)} = -\mathbb{E}_z D_w(G(z))$
  - 실제 데이터 관점과 상관 없음

ISSUES & FEATURES

- 생성기 가중치 θ 의 1회 훈련 전에
- 판별기 가중치 w를 n_{critic} 회 훈련
  - 원 논문에서는 $n_{critic} = 5$
- WGAN에서는 판별기가 먼저 최적화되어도
  유의미한 gradient를 생성하기 때문

ISSUES & FEATURES

- Momentum-based optimizer는 불안정
  - RMSProp을 사용

ISSUES & FEATURES

- 판별기 경사 계산

$$g_w \leftarrow \nabla_w \left[\frac{1}{m} \sum_{i=1}^{m} f_w(x^{(i)}) - \frac{1}{m} \sum_{i=1}^{m} f_w(g_\theta(z^{(i)})) \right]$$

  - 를 ascending 하거나
  - (−)를 곱해 descending 하거나
  - Labeling을 다음과 같이 주고 descending
    - 진짜 데이터: -1.0
    - 가짜 데이터: 1.0

ISSUES & FEATURES

- 판별기 경사 계산

$$g_w \leftarrow \nabla_w \left[\frac{1}{m} \sum_{i=1}^{m} f_w(x^{(i)}) - \frac{1}{m} \sum_{i=1}^{m} f_w(g_\theta(z^{(i)})) \right]$$

  - (–)를 곱해 descending

- $$L = - y_{label} \frac{1}{m} \sum_{i=1}^{m} y_{prediction}$$

  - y_{label} 이 부호의 역할을 수행
    - 진짜 데이터: 1.0
    - 가짜 데이터: -1.0

ISSUES & FEATURES

- 판별기 경사 계산
- $L = - y_{label} \frac{1}{m} \sum_{i=1}^{m} y_{prediction}$
- $-\frac{1}{n} \sum_{i=1}^{n} (y_i p_i)$
-
```
def wasserstein_loss(y_label, y_pred):
    return -K.mean(y_label * y_pred)
```

WGAN-GP

WEIGHT CLIPPING IS TERRIBLE WAY

- Ref. Gulrajani, Ishaan, et al. "Improved training of wasserstein gans." Advances in neural information processing systems. 2017.

WEIGHT CLIPPING IS TERRIBLE WAY

- WGAN의 가중치는 clipping boundary 근처로 몰림
  - Weight clipping 때문

WEIGHT CLIPPING IS TERRIBLE WAY

- WGAN-GP는 Weight clipping 방법의 해결책을 제시
  - Gradient Penalty (GP)

GRADIENT PENALTY

- WGAN 손실 함수:

$$\max_{w \in W} \mathbb{E}_{x \sim p_r}[f_w(x)] - \mathbb{E}_{z \sim p_r(z)}[f_w(g_\theta(z))]$$

  - max로 만드는 f 를 f^* 라 하면
  - $f^* = arg\max_{\|f\|_L \leq 1} \mathbb{E}_{y \sim \mathbb{P}_r}[f(y)] - \mathbb{E}_{x \sim \mathbb{P}_g}[f(x)]$
    - $x \sim P_{g'}$ $y \sim P_r$ 로 샘플링

GRADIENT PENALTY

- $f^* = arg\max_{\|f\|_L \leq 1} \mathbb{E}_{y\sim\mathbb{P}_r}[f(y)] - \mathbb{E}_{x\sim\mathbb{P}_g}[f(x)]$
  - x~$P_{g'}$ y~P_r 로 샘플링
- x와 y를 보간한 직선 중 x와 y 사이의 점 x_t
  - $x_t = tx + (1-t)y$
  - $0 \leq t \leq 1$
  - 에서는 norm $\left[\nabla f^*(x_t) = \frac{y-x_t}{\|y-x_t\|}\right] = 1$ 을 만족함이 증명

GRADIENT PENALTY

- x_t는 $x \sim P_{g'}$ $y \sim P_r$ 로 샘플링한 점을 보간한 직선에서 x와 y 사이의 점 중 하나를 샘플링한 것
- Norm $\left[\nabla f^*(x_t) = \frac{y - x_t}{\|y - x_t\|}\right] = 1$
  - 최적해 f^* 의 특성

GRADIENT PENALTY

- f 가 최적해 f^* 의 특성을 가지도록 근사
  - $||\nabla f(x_t)|| = 1$ 이 되도록 근사
  - 손실 함수를 수정

GRADIENT PENALTY

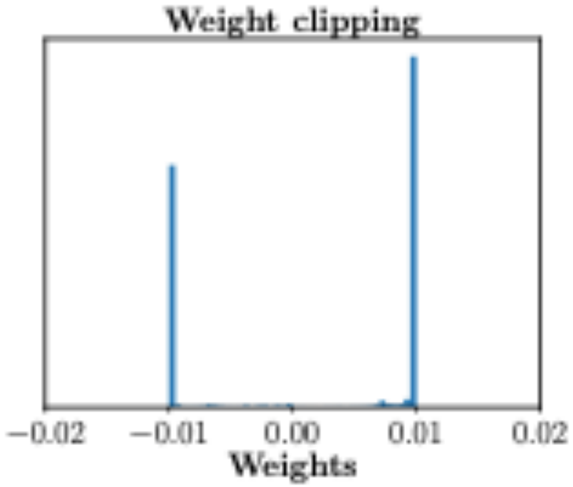
- $$L = \underbrace{\underset{\tilde{\boldsymbol{x}}\sim\mathbb{P}_g}{\mathbb{E}}\left[D(\tilde{\boldsymbol{x}})\right] - \underset{\boldsymbol{x}\sim\mathbb{P}_r}{\mathbb{E}}\left[D(\boldsymbol{x})\right]}_{\text{Original critic loss}} + \underbrace{\lambda \underset{\hat{\boldsymbol{x}}\sim\mathbb{P}_{\hat{\boldsymbol{x}}}}{\mathbb{E}}\left[\left(\|\nabla_{\hat{\boldsymbol{x}}} D(\hat{\boldsymbol{x}})\|_2 - 1\right)^2\right]}_{\text{Our gradient penalty}}.$$

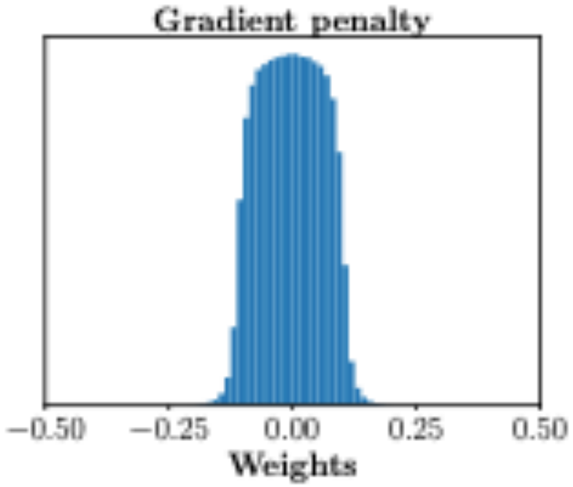
  - $||\nabla f(x_t)|| = 1$ 이 되도록 근사
  - $\hat{\boldsymbol{x}} \sim \mathbb{P}_{\hat{\boldsymbol{x}}}$는 uniform하게 샘플링한 x_t
    - $x \sim P_{g'}$ $y \sim P_r$ 로 샘플링한 점을 보간한 직선에서 x와 y 사이의 점 중 하나를 uniform 하게 샘플링

ISSUES & FEATURES

- Adam을 사용할 수 있음
  - 더 안정적인 학습

ISSUES & FEATURES





- Weight clipping과 GP의 비교
  - 가중치들이 의미있는 값을 가짐

WGAN & WGAN-GP

GAN: ONE MORE STEP FURTHER