Wasserstein GAN

X: compact metric set

- Metric = Distance (Manhattan, Euclidean… ->Mincowski)

어떤 공간에 metric 개념이 중요한 이유는 **수렴**(convergence) 이란 개념을 정의내릴 수 있기 때문입니다

$$x_n \to x \quad \Leftrightarrow \quad \lim_{n\to\infty} d(x_n, x) = 0$$

... 수학에선 위상(topology)을 유도하다(induce)는 어려운 표현을 씁니다

Distance in function

함수 공간에서는 더욱 다양하게 정의내릴 수 있습니다

- L_1 거리

$$d_1(f,g) = \|f-g\|_1 = \int_{\mathcal{X}} |f(x)-g(x)|dx$$

- L_2 거리

$$d_2(f,g) = \|f-g\|_2 = \left(\int_{\mathcal{X}} |f(x)-g(x)|^2 dx\right)^{1/2}$$

여러분의 상상보다 **훨씬** 많습니다! 😨

- L_∞ 거리

$$d_\infty(f,g) = \|f-g\|_\infty = \sup_{x\in\mathcal{X}} |f(x)-g(x)|$$

- W_2^k 거리 (Sobolev norm)

$$\|f-g\|_{W_p^k} = \sum_{n=0}^{k} \|\partial_x^n (f-g)\|_2$$

sup, inf ??

$c=inf(A)$일 조건은 다음과 같다.

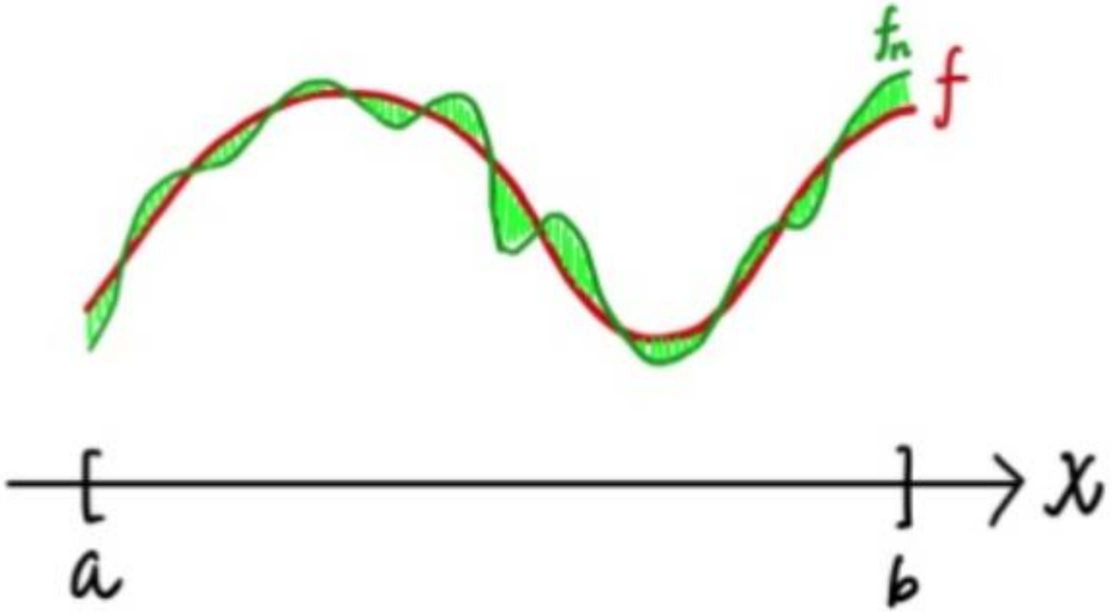
1. c is a lower bound of A
2. if c' is a lower bound of A then $c' \leq c$

$c=sup(A)$일 조건은 다음과 같다.

1. c is a upper bound of A
2. if c' is a upper bound of A then $c \leq c'$

수렴

f_n 과 f 의 차이를 제곱해서 적분한 값이 0 으로 수렴하게 만들 수 있으면
L_2-**수렴** 한다고 합니다



$$\|f_n - f\|_2 = \left(\int_a^b |f_n - f|^2 dx\right)^{1/2} \to 0$$

거리함수에 따라 수렴방식이 다름

참고로 **확률론** 에서 다루는 수렴은 여러가지 종류가 있습니다

- 균등수렴, 확률수렴, L_2-수렴, Weak-* convergence ...
- 참고로 WGAN 논문은 분포수렴 와 동등한 Wasserstein distance 를 다룹니다. 이 metric 은 **확률분포들의 공간** 에서 정의됩니다!

Heine-Borel property.

If $\mathcal{X}$ is compact, then it is closed and bounded

Heine-Borel 정리를 통해 해석하면 $\mathcal{X}$ 가 compact 란건 경계가 있고 (bounded) 동시에 경계를 포함한다(closed) 는 집합이란 겁니다

Σ : set of all the Borel subsets of $\mathcal{X}$

Borel 집합은 $\mathcal{X}$ 내에서 **측정가능**(measurable) 한 집합들을 말합니다.

- 여기서 **측정가능** 의 의미는 $\mathbb{P}_r$, $\mathbb{P}_g$ 같은 확률분포로 확률값이 계산될 수 있는 집합을 말합니다

Different distances

Total Variation (TV)

$$\delta(\mathbb{P}_r,\mathbb{P}_g) = \sup_{A\in\Sigma} |\mathbb{P}_r(A) - \mathbb{P}_g(A)|$$

Total Variation 은 두 확률측도의 측정값이 벌어질 수 있는 값 중 **가장 큰 값** (또는 앞에서 설명한 supremum) 을 말합니다

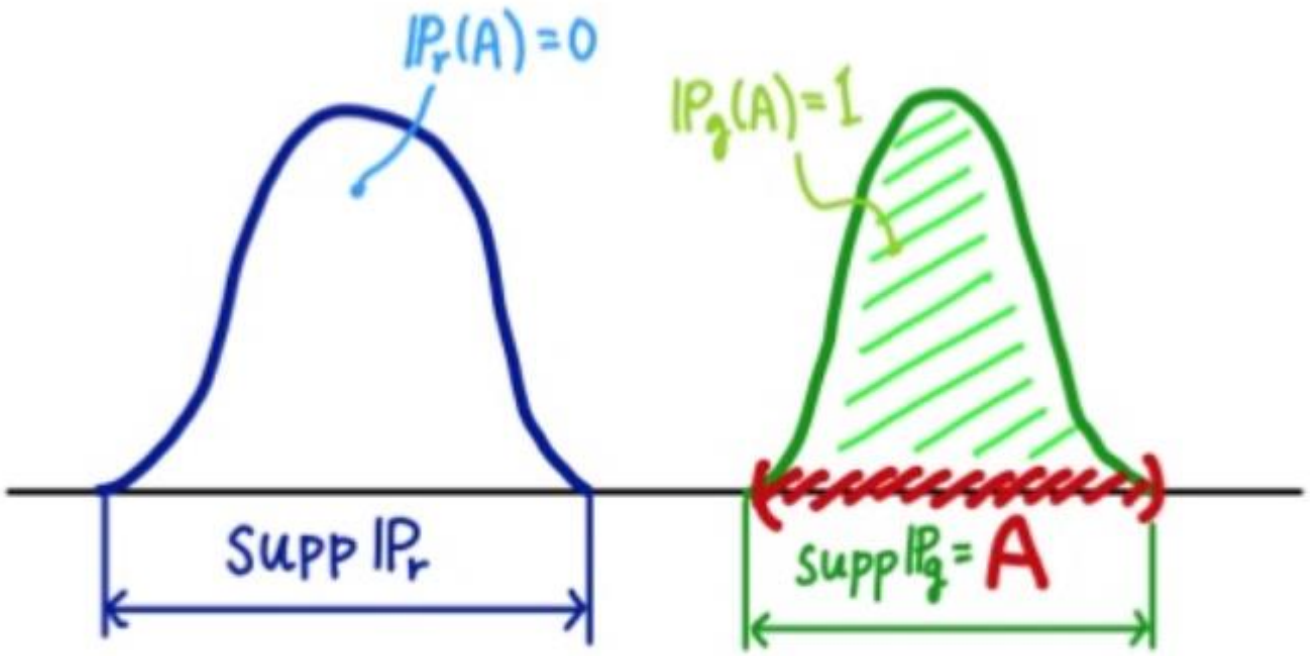
support

$$\operatorname{supp} f = \operatorname{cl}\{x \in X \colon f(x) \neq 0\}$$

극한점을 포함하는 부분집합, f(x) 는 0이 아님

만약 두 확률분포의 확률밀도함수가 서로 겹치지 않는다면, 다시 말해 확률분포의 support 의 교집합이 공집합이라면 TV 는 무조건 1입니다!

$$\delta(\mathbb{P}_r,\mathbb{P}_g) = |0-1| = 1$$



GAN학습에 맞는 distance가 필요

- ->wassertein distance

Wasserstein distance 의 정의는 이렇습니다

$$W(\mathbb{P},\mathbb{Q}) = \inf_{\gamma\in\Pi(\mathbb{P},\mathbb{Q})} \int d(x,y)\gamma(dxdy)$$

$$= \inf_{\gamma\in\Pi(\mathbb{P},\mathbb{Q})} \mathbb{E}^{\gamma}[d(X,Y)]$$

여기서 $\Pi(\mathbb{P},\mathbb{Q})$ 는 두 확률분포 $\mathbb{P},\mathbb{Q}$ 의 결합확률분포(joint distribution) 들을 모은 집합이고 γ 는 그 중 하나입니다. 즉 모든 결합확률분포 $\Pi(\mathbb{P},\mathbb{Q})$ 중에서 $d(X,Y)$ **의 기대값을 가장 작게 추정한 값** 을 의미합니다

출처

- https://www.slideshare.net/ssuser7e10e4/wasserstein-gan-i