Kantorovich-Rubinstein Duality

Instructor: John Thickstun

Discussion Board: Available on Ed

Zoom Link: Available on Canvas

Instructor Contact: thickstn@cs.washington.edu

Course Webpage: https://courses.cs.washington.edu/courses/cse599i/20au/

Wasserstein GAN

- Minimize Wasserstein distance between p and p_θ:

$$\theta_W = \arg\min_\theta W(p, p_\theta) = \arg\min_\theta \inf_{\pi \in \Pi(p, p_\theta)} \mathop{\mathbb{E}}_{(x,y)\sim\pi} \left[\|x - y\|_2\right].$$

- $\Pi(p, q)$ is the set of probability distributions on $\mathcal{X} \times \mathcal{X}$ with marginals p, q.

- We can't enforce these constraints on the marginals.

- Instead, minimize a dual characterization of the Wasserstein distance:

$$W(p, q) = \inf_{\pi \in \Pi(p,q)} \mathop{\mathbb{E}}_{(x,y)\sim\pi} \left[\|x - y\|_2\right] = \sup_{\|h\|_L \le 1} \left[\mathop{\mathbb{E}}_{x\sim p} h(x) - \mathop{\mathbb{E}}_{x\sim q} h(x)\right].$$

Kantorovich Rubinstein Duality

- Kantorovich-Rubinstein Duality:

$$W(p,q) = \inf_{\pi \in \Pi(p,q)} \mathop{\mathbb{E}}_{(x,y)\sim\pi} [\|x-y\|_2] = \sup_{\|h\|_L \leq 1} \left[\mathop{\mathbb{E}}_{x\sim p} h(x) - \mathop{\mathbb{E}}_{x\sim q} h(x) \right].$$

- Optimize over 1-Lipschitz functions instead of joint probability distributions.

- Only need access to samples from the marginals p, q.

- Where does this dual formulation of Wasserstein distance come from?

Outline of the Proof

- Kantorovich-Rubinstein Duality:

$$W(p,q) = \inf_{\pi \in \Pi(p,q)} \mathbb{E}_{(x,y)\sim\pi} [\|x-y\|_2] = \sup_{\|h\|_L \le 1} \left[\mathbb{E}_{x\sim p} h(x) - \mathbb{E}_{x\sim q} h(x) \right].$$

- Introduce Lagrange multipliers f, g to enforce the marginal constraints.

- Apply optimization duality: $W(p,q) = \inf_{\pi} \sup_{f,g} L(\pi, f, g) = \sup_{f,g} \inf_{\pi} L(\pi, f, g)$.

- Analytically compute the minimizer π^*, and plug it back in:

$$L(\pi^*, f, g) = \mathbb{E}_{x\sim p} [f(x)] + \mathbb{E}_{y\sim q} [g(y)].$$

- Cleverly replace f, g with a single optimization variable h.

Outline of the Proof

- Kantorovich-Rubinstein Duality:

$$W(p, q) = \inf_{\pi \in \Pi(p,q)} \mathop{\mathbb{E}}_{(x,y)\sim\pi} [\|x - y\|_2] = \sup_{\|h\|_L \le 1} \left[\mathop{\mathbb{E}}_{x\sim p} h(x) - \mathop{\mathbb{E}}_{x\sim q} h(x) \right].$$

- **Introduce Lagrange multipliers f, g to enforce the marginal constraints.**

- Apply optimization duality: $W(p, q) = \inf_{\pi} \sup_{f,g} L(\pi, f, g) = \sup_{f,g} \inf_{\pi} L(\pi, f, g)$.

- Analytically compute the minimizer π^*, and plug it back in:

$$L(\pi^*, f, g) = \mathop{\mathbb{E}}_{x\sim p} [f(x)] + \mathop{\mathbb{E}}_{y\sim q} [g(y)].$$

- Cleverly replace f, g with a single optimization variable h.

Introduce Lagrange Multipliers

- Wasserstein distance:

$$W(p,q) = \inf_{\pi \in \Pi(p,q)} \mathbb{E}_{(x,y)\sim\pi} [\|x-y\|_2] \,.$$

- The marginal constraints are: $p(x) = \int_{\mathcal{X}} \pi(x,y)\,dy, \text{ and } q(y) = \int_{\mathcal{X}} \pi(x,y)\,dx.$
- Need to constrain *functions* so multipliers will also be functions $f, g : \mathcal{X} \to \mathbb{R}$.

$$L(\pi, f, g) = \mathbb{E}_{(x,y)\sim\pi} [\|x-y\|_2] + \int_{\mathcal{X}} \left(p(x) - \int_{\mathcal{X}} \pi(x,y)\,dy \right) f(x)\,dx$$
$$+ \int_{\mathcal{X}} \left(q(y) - \int_{\mathcal{X}} \pi(x,y)\,dx \right) g(y)\,dy.$$

Compute the Lagrangian

$$
\begin{aligned}
L(\pi, f, g) = \underset{(x,y)\sim\pi}{\mathbb{E}}[\|x-y\|_2] &+ \int_{\mathcal{X}} \left(p(x) - \int_{\mathcal{X}} \pi(x,y)\,dy\right) f(x)\,dx \\
&+ \int_{\mathcal{X}} \left(q(y) - \int_{\mathcal{X}} \pi(x,y)\,dx\right) g(y)\,dy. \\
= \int_{\mathcal{X}\times\mathcal{X}} \|x-y\|^2 \pi(x,y)\,dy\,dx &+ \int_{\mathcal{X}} \left(p(x) - \int_{\mathcal{X}} \pi(x,y)\,dy\right) f(x)\,dx \\
&+ \int_{\mathcal{X}} \left(q(y) - \int_{\mathcal{X}} \pi(x,y)\,dx\right) g(y)\,dy. \\
= \underset{x\sim p}{\mathbb{E}}[f(x)] + \underset{y\sim q}{\mathbb{E}}[g(y)] &+ \int_{\mathcal{X}\times\mathcal{X}} \Big(\|x-y\|_2 - f(x) - g(y)\Big)\pi(x,y)\,dy\,dx.
\end{aligned}
$$

Outline of the Proof

- Kantorovich-Rubinstein Duality:

$$W(p,q) = \inf_{\pi \in \Pi(p,q)} \mathbb{E}_{(x,y)\sim\pi} [\|x-y\|_2] = \sup_{\|h\|_L \le 1} \left[\mathbb{E}_{x\sim p} h(x) - \mathbb{E}_{x\sim q} h(x) \right].$$

- Introduce Lagrange multipliers f, g to enforce the marginal constraints.

- **Apply optimization duality:** $W(p,q) = \inf_{\pi} \sup_{f,g} L(\pi, f, g) = \sup_{f,g} \inf_{\pi} L(\pi, f, g).$

- Analytically compute the minimizer π^*, and plug it back in:

$$L(\pi^*, f, g) = \mathbb{E}_{x\sim p} [f(x)] + \mathbb{E}_{y\sim q} [g(y)].$$

- Cleverly replace f, g with a single optimization variable h.

Apply Strong Duality

- Strong duality:

$$W(p, p_g) = \inf_{\pi} \sup_{f,g} L(\pi, f, g) = \sup_{f,g} \inf_{\pi} L(\pi, f, g).$$

- And recall our calculation that

$$L(\pi, f, g) = \mathop{\mathbb{E}}_{x \sim p}[f(x)] + \mathop{\mathbb{E}}_{y \sim q}[g(y)] + \int_{\mathcal{X} \times \mathcal{X}} \Big(\|x - y\|_2 - f(x) - g(y) \Big) \pi(x, y)\, dy\, dx.$$

- The first two terms are constant in π, so

$$\arg\min_{\pi} L(\pi, f, g) = \arg\min_{\pi} \int_{\mathcal{X} \times \mathcal{X}} \Big(\|x - y\|_2 - f(x) - g(y) \Big) \pi(x, y)\, dy\, dx.$$

Outline of the Proof

- Kantorovich-Rubinstein Duality:

$$W(p,q) = \inf_{\pi \in \Pi(p,q)} \mathbb{E}_{(x,y)\sim\pi} [\|x-y\|_2] = \sup_{\|h\|_L \leq 1} \left[\mathbb{E}_{x\sim p} h(x) - \mathbb{E}_{x\sim q} h(x) \right].$$

- Introduce Lagrange multipliers f, g to enforce the marginal constraints.

- Apply optimization duality: $W(p,q) = \inf_{\pi} \sup_{f,g} L(\pi, f, g) = \sup_{f,g} \inf_{\pi} L(\pi, f, g)$.

- **Analytically compute the minimizer** π^***, and plug it back in**:

$$L(\pi^*, f, g) = \mathbb{E}_{x\sim p} [f(x)] + \mathbb{E}_{y\sim q} [g(y)].$$

- Cleverly replace f, g with a single optimization variable h.

Compute the Minimizer

- We need to compute the minimizer

$$\arg\min_{\pi} L(\pi, f, g) = \arg\min_{\pi} \int_{\mathcal{X}\times\mathcal{X}} \Big(\|x-y\|_2 - f(x) - g(y)\Big)\pi(x,y)\,dy\,dx.$$

- What happens if $\|x-y\|_2 < f(x)+g(y)$?
- Concentrate mass of π at (x,y): this makes $L(\pi, f, g) = -\infty$.
- What happens if $\|x-y\|_2 \geq f(x)+g(y), \text{ for all pairs } (x,y)$?
- The best we can do is set $\pi(x,y) = 0$.

Plug in the Minimizer

- Strong duality: $W(p, p_g) = \inf_{\pi} \sup_{f,g} L(\pi, f, g) = \sup_{f,g} \inf_{\pi} L(\pi, f, g)$.
- Pick $f, g : \mathcal{X} \to \mathbb{R}$ such that $\|x - y\|_2 \geq f(x) + g(y)$:

$$W(p, p_g) = \sup_{\substack{f,g \\ f(x)+g(y) \leq \|x-y\|_2}} \inf_{\pi} L(\pi, f, g).$$

- In this case, $\pi(x, y) = 0$. Plugging in zero, we have

$$L(\pi, f, g) = \mathop{\mathbb{E}}_{x \sim p}[f(x)] + \mathop{\mathbb{E}}_{y \sim q}[g(y)] + \int_{\mathcal{X} \times \mathcal{X}} \Big(\|x - y\|_2 - f(x) - g(y) \Big) \pi(x, y) \, dy \, dx.$$

A Single Optimization Variable?

- What we've got so far:

$$W(p,q) = \sup_{f(x)+g(y)\leq \|x-y\|_2} \mathop{\mathbb{E}}_{x\sim p}[f(x)] + \mathop{\mathbb{E}}_{y\sim q}[g(y)]\,.$$

- What we're aiming for:

$$W(p,q) = \sup_{|h(x)-h(y)|\leq \|x-y\|_2} \mathop{\mathbb{E}}_{x\sim p}[h(x)] - \mathop{\mathbb{E}}_{y\sim q}[h(y)]\,.$$

- How to replace $f, g : \mathcal{X} \to \mathbb{R}$ with a single optimization variable $h : \mathcal{X} \to \mathbb{R}$?

5-Minute Break

Outline of the Proof

- Kantorovich-Rubinstein Duality:

$$W(p,q) = \inf_{\pi \in \Pi(p,q)} \mathop{\mathbb{E}}_{(x,y)\sim\pi} [\|x-y\|_2] = \sup_{\|h\|_L \leq 1} \left[\mathop{\mathbb{E}}_{x\sim p} h(x) - \mathop{\mathbb{E}}_{x\sim q} h(x) \right].$$

- Introduce Lagrange multipliers f, g to enforce the marginal constraints.
- Apply optimization duality: $W(p,q) = \inf_{\pi} \sup_{f,g} L(\pi, f, g) = \sup_{f,g} \inf_{\pi} L(\pi, f, g)$.
- Analytically compute the minimizer π^*, and plug it back in:

$$L(\pi^*, f, g) = \mathop{\mathbb{E}}_{x\sim p} [f(x)] + \mathop{\mathbb{E}}_{y\sim q} [g(y)].$$

- **Cleverly replace f, g with a single optimization variable h.**

A Lower Bound

- What we're aiming for:

$$W(p, q) = \sup_{|h(x)-h(y)| \leq \|x-y\|_2} \mathop{\mathbb{E}}_{x \sim p} [h(x)] - \mathop{\mathbb{E}}_{y \sim q} [h(y)] \,.$$

- A lower bound:

$$\begin{aligned} \mathop{\mathbb{E}}_{x \sim p} [h(x)] - \mathop{\mathbb{E}}_{y \sim q} [h(y)] &= \int_{\mathcal{X} \times \mathcal{X}} (h(x) - h(y)) \, \pi(x, y) \, dx \, dy \\ &\leq \int_{\mathcal{X} \times \mathcal{X}} \|x - y\|_2 \, \pi(x, y) \, dx \, dy = W(p, q). \end{aligned}$$

- If it's also an upper bound, then we are done!

An Upper Bound?

- What we've got from our Lagrangian analysis:

$$W(p,q) = \sup_{f(x)+g(y)\leq \|x-y\|_2} \mathop{\mathbb{E}}_{x\sim p}[f(x)] + \mathop{\mathbb{E}}_{y\sim q}[g(y)].$$

- What we need to show:

$$\begin{aligned} W(p,q) &= \sup_{f(x)+g(y)\leq \|x-y\|_2} \left[\mathop{\mathbb{E}}_{x\sim p}[f(x)] + \mathop{\mathbb{E}}_{y\sim q}[g(y)]\right] \\ &\leq \sup_{|h(x)-h(y)|\leq \|x-y\|_2} \left[\mathop{\mathbb{E}}_{x\sim p}[h(x)] - \mathop{\mathbb{E}}_{y\sim q}[h(y)]\right] \leq W(p,q). \end{aligned}$$

Infimal Convolution

- Define the "infimal convolution"

$$\kappa(x) = \inf_u \left[\|x - u\|_2 - g(u) \right].$$

- Lemma: κ is 1-Lipschitz. Proof

- $\kappa(x) \leq \|x - u\|_2 - g(u) \leq \|x - y\|_2 + \|y - u\|_2 - g(u)$ (triangle inequality)

- This holds for any value of u, so

$$\kappa(x) \leq \|x - y\|_2 + \inf_u \left[\|y - u\|_2 - g(u) \right] = \|x - y\|_2 + \kappa(y).$$

- Rearranged: $\kappa(x) - \kappa(y) \leq \|x - y\|_2$. Reverse x, y: $|\kappa(x) - \kappa(y)| \leq \|x - y\|_2$.

Finishing Up

- If $f(x) + g(y) \leq \|x - y\|_2$ (for all x, y) then for all v,

$$f(v) \leq \inf_u \left[\|v - u\|_2 - g(u) \right] = \kappa(v) \leq \|v - v\|_2 - g(v) = -g(v).$$

- This gives us the upper bound:

$$\mathop{\mathbb{E}}_{x \sim p} [f(x)] + \mathop{\mathbb{E}}_{y \sim q} [g(y)] \leq \mathop{\mathbb{E}}_{x \sim p} [\kappa(x)] - \mathop{\mathbb{E}}_{y \sim q} [\kappa(y)].$$

- I.e. $W(p, q) = \sup_{f(x)+g(y) \leq \|x-y\|_2} \left[\mathop{\mathbb{E}}_{x \sim p} [f(x)] + \mathop{\mathbb{E}}_{y \sim q} [g(y)] \right]$

$$\leq \sup_{|h(x)-h(y)| \leq \|x-y\|_2} \left[\mathop{\mathbb{E}}_{x \sim p} [h(x)] - \mathop{\mathbb{E}}_{y \sim q} [h(y)] \right] \leq W(p, q).$$