# Class 12

Shikhar Saxena

March 06, 2023

## Contents



## Recap

- Generative Modeling
- GANs
  - ⋆ KL and f divergence
  - ⋆ bring sup out of integral (variational approach)
  - ⋆ $E_p[f(x)] - E_q[f(y)]$
- WGANs
  - ⋆ Wasserstein distance W-1 $\inf_{\gamma \ \Pi(p,q)} E[\ \|x-y\|\ ]$
  - ⋆ KR Duality $\sup_{\|h\|\leq 1} E[h(x)] - E[h(y)]$
  - ⋆ Gradient Penalty in WGAN improves performance (by enforcing the 1-Lipschitz condition better). $\sup_{\|h\|_2\leq 1} E[h(x)] - E[h(y)] + (\|\nabla h\| - 1)^2$

*Gradient Penalty Proof.* There exists an $f^*$ that is a solution to the KR duality problem (optimal minimizer for W-1 distance) such that

$$\nabla f^*(x_t) = \frac{y - x_t}{\|y - x_t\|} \implies \|\nabla f^*(t)\| = 1$$

Here $v$ is the direction in which $\frac{\partial f}{\partial v} = 1$.

Since $f^*$ is 1-Lipschitz $\implies \|\nabla f^*(x)\| \leq 1$.

$$\|\nabla f^*(x_t)\|^2 = \langle v, \nabla f^*(x_t)\rangle^2 + \|\nabla f^*(x_t) - \langle v, \nabla f^*(x_t) v\rangle\|^2$$

Proof:

$$\begin{aligned}
1 &\geq \|\nabla f^*(x_t)\|^2 \\
&= \langle v, \nabla f^*(x_t)\rangle^2 + \|\nabla f^*(x_t) - \langle v, \nabla f^*(x_t) v\rangle\|^2 \\
&= \langle v, \nabla f^*(x_t)\rangle^2 + \|\nabla f^*(x_t)\|^2 + \|\langle v, \nabla f^*(x_t)\rangle v\|^2 - 2\langle \nabla f^*(x_t), \langle v, \nabla f^*(x_t)\rangle v\rangle
\end{aligned}$$

(Some terms cancel) and we get that expression. □

Thus, we get that $\nabla f^*(x_t) = v$ (since norm greater than and less than equal to 1).

# Optimal Transport

$(X, p)$ and $(Y, q)$ be some finite probability space with $|X| = n$ and $|Y| = m$.

$\Pi(p, q)$: Collection of distributions on the product $X \times Y$ with marginals $p$ and $q$

Consider a cost $C : X \times Y \to R_+$ then the optimal transport problem is:

$$d_c(p, q) = \min_{\pi \in \Pi(p,q)} \langle C, \pi\rangle = \min_{\pi} \sum_{x,y} C(x, y)\pi(x, y) \tag{1}$$

## Example

If $X, Y \subset R^n$, $C(x, y) = \|x - y\|_2$ then optimal transport between $p$ and $q$:

$$d_c(p, q) = \min_{\pi} \sum_{x,y} \|x - y\|_2 \pi(x, y) = W_1(p, q)$$

Marginals:

$p(x) = \sum_{i=1}^{m} \pi(x, y_i) = (\pi \ 1_m)_x$ where $1_m$ is vector of all ones.

Similarly, $q(y) = \sum_{i=1}^{n} \pi(x_i, y) = 1_n^T \pi$.

Hence (1) simplifies to

$$d_c(p, q) = \min_{\pi 1_m = p, \pi^T 1_n = q} \sum_{x,y} c(x, y)\pi(x, y) \tag{2}$$

**Remark.** *Discrete EM distance minibatch approximation of W-distance between two continuous distributions.*

Concretely, If $p$ and $q$ are continuous distribution.

$\{x_i\} \sim p$ and $\{y_i\} \sim q$

Then, $\tilde{p}(x) = 1/n \sum_{i=1}^{n} 1_{x_i = x}$, $\tilde{q}(y) = 1/m \sum_{i=1}^{m} 1_{y_i = y}$

Expect that $W_1(p, q) \approx W_1(\tilde{p}, \tilde{q})$

Recall LP and that (2) is a LP. We can use LP solvers.

# Sinkhorn Generative Modeling

- More tractable that WGAN

Key Idea: Sinkhorn Distance (*Cuturi, 2013, Sinkhorn Distance. Lightspeed computation of optimal transport.*)

He added a term to the optimization problem (2),

$$d_c^\lambda(p,q) = \min_{\pi \in \Pi} \langle C, \pi \rangle - \lambda H(\pi) \tag{3}$$

Clearly as $\lambda \to 0$, $d_c^*(p,q) \to d_c(p,q)$. Reframe (3) as a KL divergence.

- Define $K(x,y) = e^{-c(x,y)/\lambda}$.
- Let $Z_\lambda = \sum_{x,y} K(x,y)$.

Then $\frac{1}{Z_\lambda} K(x,y)$ defines a Gibb's distribution $p_k^\lambda$.

We have

$$D_{KL}(\pi \| p_k^\lambda) = \sum_{x,y} \pi(x,y) \log \left( \frac{\pi(x,y)}{K(x,y)} Z_\lambda \right)$$

$$\begin{aligned} D_{KL}(\pi \| p_k^\lambda) &= \sum_{x,y} \pi(x,y) \log \pi(x,y) - \sum_{x,y} \pi(x,y) \left( -\frac{c(x,y)}{\lambda} \right) + \sum_{x,y} \pi(x,y) \log Z_\lambda \\ &= -H(\pi) + \frac{1}{\lambda} \langle C, \pi \rangle + \log Z_\lambda \end{aligned}$$

Therfore,

$$arg \min_{\pi \in \Pi} \langle C, \pi \rangle - \lambda H(\pi) = arg \min_{\pi} D(\pi \| p_k^\lambda) \tag{4}$$

## Questions for solver for (4)

1. Existence: Does a solution exist?
2. Uniqueness: Is the solution unique?

Recall, from Metric Space Calculus and topology,

- A continuous function on a compact space attains extrama
- Claim 1 $\Pi(x,y)$ is a compact space
- Claim 2 $-H(\pi)$ is strongly convex

**Definition 1** ($-mu$ strongly convex)**.** *A function is strongly convex if*

$$\nabla^2 f(x) \geq \mu I$$

*(same as $\lambda_{min}(\nabla^2 f) \geq \mu > 0$). This implies $f$ is $\mu$-strongly convex.*

*For Claim 2.*

$$\nabla^2 H(\pi) = 1/\pi > 0$$

Hence strongly convex. Take any $\mu < 1$, then $H(\pi)$ will be $\mu$-strongly convex. $\square$

*For Claim 1.* If $\Pi$ is compact ($X$ and $Y$ is compact) then $\pi(p, q)$ is a **closed subset** of $\Pi$. This follows from closed subset of compact set is compact.

Also, Convexity implies continuity. So continuous functions on a compact space attain minima.

Strong Convexity implies uniqueness of solutions.

So, for some $f : X \to Y$ such that $f(X)$ is compact and is closed and bounded. $\square$

$\langle C, \pi \rangle$ being linear and $-\lambda H(\pi)$ being strongly convex, then $\langle C, \pi \rangle - \lambda H(\pi)$ is strongly convex.

## Conclusion

(3) has existence of solution and solution is unique.

$$arg \min_{\pi} D(\pi \| p_k^{\lambda})$$

Let $\Pi(p)$ and $\Pi(q)$ denote row and column marginal constraints. And another constraint that $\Pi(p, q) = \Pi(p) \cap \Pi(q)$. So we want to find intersection point for two convex sets here (which is a generic optimization problem).

This can be solved using alternating projection.

## Alternating Projection

Refer here