Deduction of Wasserstein Generative Adversarial Networks

hello

Abstract:

we propose Primal Wasserstein GAN(P-WGAN),a variant of generative adversarial networks to minimize the wasserstein distance between the generator distribution and the data distribution directly instead of the dual form.Experimentally we show P-WGAN to be stable and avoid mode collapse.and we present results on serveral popular benchmark probles for image generation,and converge fastly. when train with large mini-batches.

Advantages:

1. Direct optimize the primal problems of Wasserstein Distance,ensure the stability of the model.
2. Propose the One-to-one matching method,Construct a new linear optimization objectives,and avoid mode collapse.
3. converge fastly.
4. Introduce isometric feature mapping,to solve the poor statistical efficiency of simple fixed cost function c like Euclidean distance in high dimensions.

Proof:

1. proof the linear combination of different metrics is also a metric
2. proof lipschitz-free space is isometric space.

Experiments:

1. mixture of gaussian dataset
2. mnist and fashion mnist
3. cifar10 inception score

Proofs for Theoretical Results

Assume $||f||_L \leq 1$,

$$||f(x_i) - f(y_j)|| \leq ||x_i - y_j||$$

because $f(x) \in R^1$ therefore

$$\begin{aligned} ||x_i - y_j|| &\geq ||f(x_i) - f(y_j)|| \\ &= |f(x_i) - f(y_j)| \\ &\geq f(x_i) - f(y_j) \end{aligned}$$

Optimal distribution π satisfy the condition as follow

$$s.t. \begin{cases} \sum_j \pi_{ij} = p_r(x_i) \\ \sum_i \pi_{ij} = p_\theta(y_j) \\ \pi_{ij} \geq 0 \end{cases}$$

Then the Wasserstein distance is defined as

$$\begin{aligned} W(p_\theta, p_r) &= \inf_{\gamma \in \pi} \sum_{x_i \sim p_r, y_j \sim p_\theta} c(x_i, y_j)\gamma_{ij} \\ &= \sup_{\gamma \in \pi} \sum_{ij} (f(x_i) - f(y_j))\gamma_{ij} \\ &= \sup(\sum_i f(x_i) - \sum_j f(y_j)) \end{aligned}$$

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

$$W(x, G(z)) = \inf_{\gamma \in \pi} \sum_{ij} ||x_i - G(z_j)||_2^2 \gamma_{ij}$$

$$W(x, G(z)) = \inf_{\gamma \in \pi} \sum_{ij} (||x_i - G(z_j)||_2^2 + ||x_i - G(z_j)||_1)\gamma_{ij}$$

$$W(Enc(x), G(z)) = \inf_{\gamma \in \pi} \sum_{ij} ||Enc(x_i) - G(z_j)||_2^2 \gamma_{ij}$$

$$\begin{aligned} L_d &= -E_{x \sim P_r}[logD(x)] - E_{x \sim P_g}[log(1 - D(x))] \\ L_g &= E_{x \sim P_r}[logD(x)] + E_{x \sim P_g}[log(1 - D(x))] \\ L(G, D) &= \int_x \bigg(p_r(x) \log(D(x)) + p_g(x) \log(1 - D(x)) \bigg) dx \end{aligned}$$

$$\tilde{x} = D(x), A = p_r(x), B = p_g(x)$$

$$
\begin{aligned}
f(\tilde{x}) &= Alog\tilde{x} + Blog(1-\tilde{x}) \\
\frac{df(\tilde{x})}{d\tilde{x}} &= A\frac{1}{ln10}\frac{1}{\tilde{x}} - B\frac{1}{ln10}\frac{1}{1-\tilde{x}} \\
&= \frac{1}{ln10}(\frac{A}{\tilde{x}} - \frac{B}{1-\tilde{x}}) \\
&= \frac{1}{ln10}\frac{A-(A+B)\tilde{x}}{\tilde{x}(1-\tilde{x})}
\end{aligned}
$$

$$D^*(x) = \tilde{x}^* = \frac{A}{A+B} = \frac{p_r(x)}{p_r(x)+p_g(x)} \in [0,1]$$

$$
\begin{aligned}
D_{JS}(p_r\|p_g) =& \frac{1}{2}D_{KL}(p_r||\frac{p_r+p_g}{2}) + \frac{1}{2}D_{KL}(p_g||\frac{p_r+p_g}{2}) \\
=& \frac{1}{2}\Bigg(\log 2 + \int_x p_r(x)\log\frac{p_r(x)}{p_r+p_g(x)}dx\Bigg) + \\
& \frac{1}{2}\Bigg(\log 2 + \int_x p_g(x)\log\frac{p_g(x)}{p_r+p_g(x)}dx\Bigg) \\
=& \frac{1}{2}\Bigg(\log 4 + L(G, D^*)\Bigg)
\end{aligned}
$$

thus:

$$L(G,D^*) = 2D_{JS}(p_r\|p_g) - 2\log 2$$

cost function:

$$
\begin{aligned}
||x||_1 &= \big(\sum_{i=1}^{n}|x_i|\big) \\
||x||_2 &= \big(\sum_{i=1}^{n}x_i^2\big)^{1/2} \\
||x||_{L^p} &= \big(\sum_{i=1}^{n}x_i^p\big)^{1/p}
\end{aligned}
$$

Banach Space: ________

Gradient Penalty

$$\lambda\big(||\nabla_{\tilde{x}}f(\tilde{x})|| - 1\big)$$

consistency term:

$$CT|_{x_1,x_2} = E_{x_1,x_2}\Bigg[max\Bigg(0, \frac{d(f(x_1),f(x_2))}{d(x_1,x_2)} - M^{'}\Bigg)\Bigg]$$

$$CT|x', x'' = E_{x\sim P_r}[max(0, d(D(x'), D(x'')) + 0.1 \cdot d(D\_(x'), D\_(x'')) - M')]$$

orthonormal regularization

$$\lambda||W^TW - I||_F^2$$

spectual normalization

$$||f||_{lip} \leq \prod_{l=1}^{L+1}\sigma(W^l)$$

$$W_{SN} = \frac{W}{\sigma(W)}$$

sinkhorn distance:

$$\pi^\lambda = argmin < \pi, c > -\frac{1}{\lambda}h(\pi)$$

Lipschitz:

$$||f||_{Lip} = sup\frac{d_Y(f(x),f(y))}{d_X(x,y)}$$