Mathematics behind GAN

Li Jun

Outline

What is GAN

CAN Abetract

GAN Abstract

Previous Generative Models

Math For Basic

Wassertein GAN

Improved Training

of Wassertein GAN

What is KL

Mathematics behind GAN

Li Jun

 $2018~\mathrm{c}~1~\mathrm{ce}~29~\mathrm{F}$

Contents

Mathematics behind GAN

Li Jun

Outline

What is GAN?

O/ (IV / (DStract

Previous Generative Models

Math For Bas GAN

Wassertein GAN

Improved Training of Wassertein GAN

What is KL divergence?

- What is GAN?
- Q GAN Abstract
- Previous Generative Models
- Math For Basic GAN
- Wassertein GAN
- 6 Improved Training of Wassertein GAN
- What is KL divergence?

Math behind GAN

Mathematics behind GAN

Li Jur

Outline

What is GAN?

GAN Abstract

G/ (IV / (D)C) acc

Previous Generative

Math For Basic GAN

Wassertein GAN

Improved Training of Wassertein GAN

What is KL

Definition

GAN is composed of two networks: Descrimitive Network, and Generative Network.

GAN abstract

Mathematics behind GAN

Outline

What is GAN?

Wildt is GAIN:

GAN Abstract

Previous Generative

Math For Basic GAN

Wassertein GAN

Improved Training of Wassertein GAN

What is KL

 GAN is a framework for estimating generative models via an adversarial process

GAN abstract

Mathematics behind GAN

Outline

What is GAN

GAN Abstract

Previous Generative

Math For Basi

Wassertein GAN

Improved Training of Wassertein GAN

What is KI

- GAN is a framework for estimating generative models via an adversarial process
- simultaneously train two models: A generative model G and A discriminative model D.

GAN abstract

Mathematics behind GAN

GAN Abstract

GAN is a framework for estimating generative models via an adversarial process

- simultaneously train two models: A generative model G and A discriminative model D.
- This framework corresponds to a minimax two-player game.



Mathematics behind GAN

Li Ju

Outline

What is GAN?

o/ tre / tbstrac

Previous Generative Models

Math For Basic GAN

Wassertein GAN

Improved Training

What is KL

deep Boltzmann machine

Mathematics behind GAN

Li Jur

Outline

What is GAN?

GAN Abstract

Previous Generative Models

Math For Basic

Wassertein GAN

Improved Training of Wassertein GAN

What is KL

- deep Boltzmann machine
- @ Generative stochastic networks

Mathematics behind GAN

Li Jur

Outline

What is GAN

GAN Abstrac

Previous Generative Models

Math For Basic

Wassertein GAN

Improved Training of Wassertein GAN

What is KL

- deep Boltzmann machine
- @ Generative stochastic networks
- variational autoencoders(VAEs)

Mathematics behind GAN

Li Jur

Outline

What is GAN

GAN Abstrac

Previous Generative Models

Math For Basi GAN

Wassertein GAN

Improved Training of Wassertein GAI

What is KL

- deep Boltzmann machine
- @ Generative stochastic networks
- variational autoencoders(VAEs)
- 4 . .

Mathematics behind GAN

Li Jur

Outline

What is GAN

GAN Abstract

Models

Math For Basic GAN

Wassertein GAN

Improved Training of Wassertein GAN

What is KL divergence?

Generator

- data x
- ② input noise variables $p_z(z)$
- **3** mapping to data space as $G(z; \theta_g)$, where G is a differentiable function represented by a multilayer perceptron with parameter θ_g .

Mathematics behind GAN

Li Jur

Outline

What is GAN

GAN Abstrac

Models

Math For Basic GAN

Wassertein GAN

Improved Training of Wassertein GAN

What is KL

Discriminator

- $D(x; \theta_d)$ which is a multilayer perceptron that outputs a single scalar.
- ② D(x) represents the probability that x came from data rather than p_g

Mathematics behind GAN

Li Jur

Outline

What is GAN?

GAN Abstrac

Previous Generative

Math For Basic GAN

Wassertein GAN

Improved Training of Wassertein GAN

What is KL

minimax playgame

$$\min_{G} \max_{D} V(D, G) = \\ E_{x \sim p_{data}}[\log D(x)] + E_{z \sim p_z}[\log(1 - D(G(z))]$$

Mathematics behind GAN

Math For Basic GAN

Optimium D

$$\max_{D} V(D, G) = E_{x \sim p_{data}} [\log D(x)] + E_{z \sim p_{z}} [\log(1 - D(G(z)))]$$

= $E_{x \sim p_{data}} [\log D(x)] + E_{x \sim p_{z}} [\log(1 - D(x))]$

$$= E_{X \sim p_{data}}[\log D(X)] + E_{X \sim p_g}[\log(1 - D(X))]$$

$$= \int_{x} p_{data}(x) [\log D(x)] dx + \int_{x} p_{g}(x) \log(1 - D(x)) dx$$

Mathematics behind GAN

Li Ju

Outline

What is GAN

CANI Abatua at

GAN Abstract

Previous Generative Models

Math For Basic GAN

Wassertein GAN

Improved Training of Wassertein GAN

Optimium D

- $\max_{D} V(D, G) =$ $\int_{X} p_{data}(x) [\log D(x)] dx + \int_{X} p_{g}(x) \log(1 D(x)) dx$
- ② for given x, $p_{data}(x)$ is constant, marked as a
- **3** for given x, $p_g(x)$ is constant, marked as b
- **3** To find max of f(D), $\frac{\partial f(D)}{\partial D} = 0$
- **1** That is, for given G, $D^* = \frac{p_{data}(x)}{p_g(x) + p_{data}(x)}$

Mathematics behind GAN

Li Jui

Outlin

What is GAN

GAN Abstract

Previous Generative

Math For Basic GAN

Wassertein GAN

Improved Training of Wassertein GAN

What is KL divergence?

Optimium D

for given G,
$$D^* = \frac{p_{data}(x)}{p_g(x) + p_{data}(x)}$$

 $\max_{D} V(D^*, G) = \int_{X} p_{data}(x) [\log D^*(x)] dx + \int_{X} p_g(x) \log(1 - D^*(x)) dx$
 $= \int_{X} p_{data}(x) [\log \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}] dx + \int_{X} p_g(x) [\log \frac{p_g(x)}{p_{data}(x) + p_g(x)}] dx$
 $= \int_{X} p_{data}(x) [\log \frac{\frac{1}{2}p_{data}(x)}{\frac{1}{2}(p_{data}(x) + p_g(x))}] dx + \int_{X} p_g(x) [\log \frac{\frac{1}{2}p_g(x)}{\frac{1}{2}(p_{data}(x) + p_g(x))}] dx$
 $= -\log 4 + KL(p_{data}||\frac{1}{2}(p_{data} + p_g)) + KL(p_g||\frac{1}{2}(p_{data} + p_g))$

Mathematics behind GAN

Li Jui

Outline

What is GAI

GAN Abstrac

GAN Abstract

Previous Generative Models

Math For Basic GAN

Wassertein GAN

Improved Training of Wassertein GAN

What is KL

Optimum Discriminator D^*

$$\max_{D} V(D, G) =$$

$$-\log 4 + KL(p_{data}||\frac{1}{2}(p_{data}+p_g)) + KL(p_g||\frac{1}{2}(p_{data}+p_g))$$

if
$$p_g = p_{data}$$

then
$$\max_{D} V(D, G) = -\log 4$$

Mathematics behind GAN

Li Jur

Outline

What is GAN

CAN Abetraci

GAN Abstrac

Previous Generative Models

Math For Basic GAN

Wassertein GAN

Improved Training of Wassertein GAN

What is KL

Distance at Optimum Discriminator

$$\max V(D,G) =$$

$$-\log 4 + KL(p_{data}||\frac{1}{2}(p_{data}+p_g)) + KL(p_g||\frac{1}{2}(p_{data}+p_g))$$

$$\max_{D} V(D, G) = -\log 4 + 2 * JSD(p_{data}||p_g)$$

Mathematics behind GAN

Li Jur

Outlin

What is GAN

CAN Abetract

Models

Math For Basic GAN

Wassertein GAI

Improved Training of Wassertein GAN

What is KL

Theorem (Optimum Generator $G^st)$

After we get optimum Discriminator D^* , we want to get optimum Generator by gradient descent to get the minimum value of $\max_{G} V(G,G)$, $\rightarrow \min_{G} \max_{D} V(G,D)$ and when $p_g = p_{data}$, it can get the minimum.

$$p_g = p_{data}$$
, it can get the min

$$\theta_{\mathsf{g}} \leftarrow \theta_{\mathsf{g}} - \alpha \frac{\partial L(\mathsf{G})}{\partial \theta_{\mathsf{g}}}$$

Mathematics behind GAN

Li Jur

Outlin

What is GAN7

Wilat is GAIV:

GAN Abstrac

Previous Generativ Models

Math For Basic GAN

Wassertein GAN

Improved Training of Wassertein GAN

What is KL

Total Variation Distance

$$\delta(P_r, P_g) = \sup_{A \in \Sigma} |P_r(A) - P_g(A)|$$

Mathematics behind GAN

Li Jur

Outline

What is GAN?

GAN Abstract

Previous Generative Models

Math For Basic

Wassertein GAN

Improved Training of Wassertein GAN

What is KL divergence? Total Variation Distance

$$\delta(P_r, P_g) = \sup_{A \in \Sigma} |P_r(A) - P_g(A)|$$

Kullback-Leibler (KL) Divergence

$$KL(P_r||P_g) = \int P_r(x) \log \frac{P_r(x)}{P_g(x)} d\mu(x)$$

Mathematics behind GAN

Li Jur

Outline

What is GAN

GAN Abstract

Previous Generative Models

Math For Basi GAN

Wassertein GAN

Improved Training of Wassertein GAM

Total Variation Distance

$$\delta(P_r, P_g) = \sup_{A \in \Sigma} |P_r(A) - P_g(A)|$$

- **③** Kullback-Leibler (KL) Divergence $KL(P_r||P_g) = \int P_r(x) \log \frac{P_r(x)}{P_n(x)} d\mu(x)$
- Jensen-Shannon (JS) Divergence

$$JS(P_r, P_g) = KL(P_r||(P_r + P_g)/2) + KL(P_g||(P_r + P_g)/2)$$

Mathematics behind GAN

Li Jui

Outline

What is GAN

GAN Abstract

Previous Generative
Models

Math For Basi

Wassertein GAN

Improved Training of Wassertein GAN

What is KL divergence?

Total Variation Distance

$$\delta(P_r, P_g) = \sup_{A \in \Sigma} |P_r(A) - P_g(A)|$$

- **③** Kullback-Leibler (KL) Divergence $KL(P_r||P_g) = \int P_r(x) \log \frac{P_r(x)}{P_\sigma(x)} d\mu(x)$
- Jensen-Shannon (JS) Divergence $JS(P_r, P_g) = KL(P_r||(P_r + P_g)/2) + KL(P_g||(P_r + P_g)/2)$
- Earth-Mover(EM) distance or Wasserstein-1

$$W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} E_{(x,y) \sim \gamma}[\|x - y\|]$$

Wassertein GAN

Mathematics behind GAN

Li Ju

Outlin

What is GAN

GAN Abstract

Previous Generativo Models

Math For Basic

Wassertein GAN

Improved Training of Wassertein GAN

What is KL

Distance

$$W(P_r, P_g) = \sup_{\|f\|_L \le 1} E_{x \sim P_r}[f(x)] - E_{x \sim P_{\theta}}[f(x)]$$

$$W(P_r, P_g) = \max_{\|f\|_L \le 1} E_{x \sim P_r}[f(x)] - E_{x \sim P_{\theta}}[f(x)] \text{ We should}$$

find all 1-Lipschitz functions f: $\chi \to \mathbf{R}$

Lipschitz Function

$$\|f(x_1) - f(x_2)\| \le k\|x_1 - x_2\|$$
 when $k = 1$, for 1-Lipschitz

That is: Find Max distance $W(P_r, P_g)$, and then optim by gradi $\nabla_{\theta}W(P_r, P_{\theta}) = -E_{z \sim p(z)}[\nabla_{\theta}f(g_{\theta}(z))]$ Also, we clamp the weights to a fixed box,say W = [-0.01, 0.01].

Wassertein GAN summary

Mathematics behind GAN

Outline

What is GAN

AN Abstrac

revious Generative

Math For Basi

Wassertein GAN

Improved Training

of Wassertein GAI

What is KL

So, Wassertein GAN just clip the weights? and then assume the neural network learn the f function to fullfil the 1-Lipschitz function!

Improved Wassertein GAN

Mathematics behind GAN

Improved Training of Wassertein GAN

o be continued.

KL divergence

Mathematics behind GAN

Li Ju

Outline

What is GAN?

GAN Abstract

GAN Abstract

Models

Math For Basic

Wassertein GAN

Improved Training of Wassertein GAN

What is KL divergence?

Definition

$$\mathit{KL}(p||q) = \sum_{k=1}^{N} p_k \log \frac{p_k}{q_k}$$

What's the mean of KL divergence

the divergence (distance) of two distributions.