Fisher GAN

Youssef Mroueh, TomSercu

汇报人：吴凡

Original GAN

- The original GAN formulation optimizes the Jensen-Shannon divergence

$$V = E_{x\sim P_{data}}[logD(x)] + E_{x\sim P_G}\left[log\left(1 - D(x)\right)\right] = E_{x\sim P_{data}}\left[log\frac{P_{data}(x)}{P_{data}(x) + P_G(x)}\right] + E_{x\sim P_G}\left[log\frac{P_G(x)}{P_{data}(x) + P_G(x)}\right]$$

$$= -2log2 + 2JSD\left(P_{data}(x)||P_G(x)\right)$$

- Later work generalized this to optimize f-divergences, KL, the Least Squares objective

Fisher GAN

- We build in this work on the Integral probability Metrics(IPM) framework for learning GAN
- the IPM defines a critic function f belonging to a function class $\mathcal{F}$, that maximally discriminates between the real and fake distributions

$$d_{\mathscr{F}}(\mathbb{P},\mathbb{Q}) = \sup_{f \in \mathscr{F}} \left\{ \mathop{\mathbb{E}}_{x \sim \mathbb{P}} f(x) - \mathop{\mathbb{E}}_{x \sim \mathbb{Q}} f(x) \right\}$$

$*$ Describe the discrepancy between the means of a function under two different distributions

First formulation

- Rayleigh Quotient（瑞利商）

$$R(A,x)=\frac{x^H A x}{x^H x}$$

* x为非零向量，A为$n \times n$的Hermitian（埃尔米特）矩阵。 Hermitian矩阵是指矩阵的对称位置上的值相互为共轭复数，如果A是实数矩阵，则只需要A为对称矩阵即可

- 有界

$$\lambda_{min} \leq \frac{x^H A x}{x^H x} \leq \lambda_{max}$$

- Purpose

Ensures the stability of the training while maintaining the capacity of the critic

过于苛刻的约束会使训练变得困难，例如wGAN

First formulation: Rayleigh Quotient form

- the Fisher IPM for a function space $\mathcal{F}$, $\mathcal{F}$ can be any symmetric function class（在对称函数中，函数的输出值不随输入变量的排列而改变）

$$d_{\mathscr{F}}(\mathbb{P},\mathbb{Q}) = \sup_{f\in\mathscr{F}} \frac{\underset{x\sim\mathbb{P}}{\mathbb{E}}[f(x)] - \underset{x\sim\mathbb{Q}}{\mathbb{E}}[f(x)]}{\sqrt{1/2\mathbb{E}_{x\sim\mathbb{P}}f^2(x) + 1/2\mathbb{E}_{x\sim\mathbb{Q}}f^2(x)}}$$

Second formulation: Constrained form

$$d_{\mathscr{F}}(\mathbb{P},\mathbb{Q}) = \sup_{f\in\mathscr{F},\, \frac{1}{2}\mathbb{E}_{x\sim\mathbb{P}}f^2(x)+\frac{1}{2}\mathbb{E}_{x\sim\mathbb{Q}}f^2(x)=1} \mathscr{E}(f) := \mathop{\mathbb{E}}_{x\sim\mathbb{P}}[f(x)] - \mathop{\mathbb{E}}_{x\sim\mathbb{Q}}[f(x)]$$

Learning GAN with Fisher IPM

- Objective

$$\min_{g_\theta} \sup_{f_p \in \mathscr{F}_p} \hat{\mathscr{E}}(f_p, g_\theta) := \frac{1}{N}\sum_{i=1}^{N} f_p(x_i) - \frac{1}{M}\sum_{j=1}^{M} f_p(g_\theta(z_j))$$

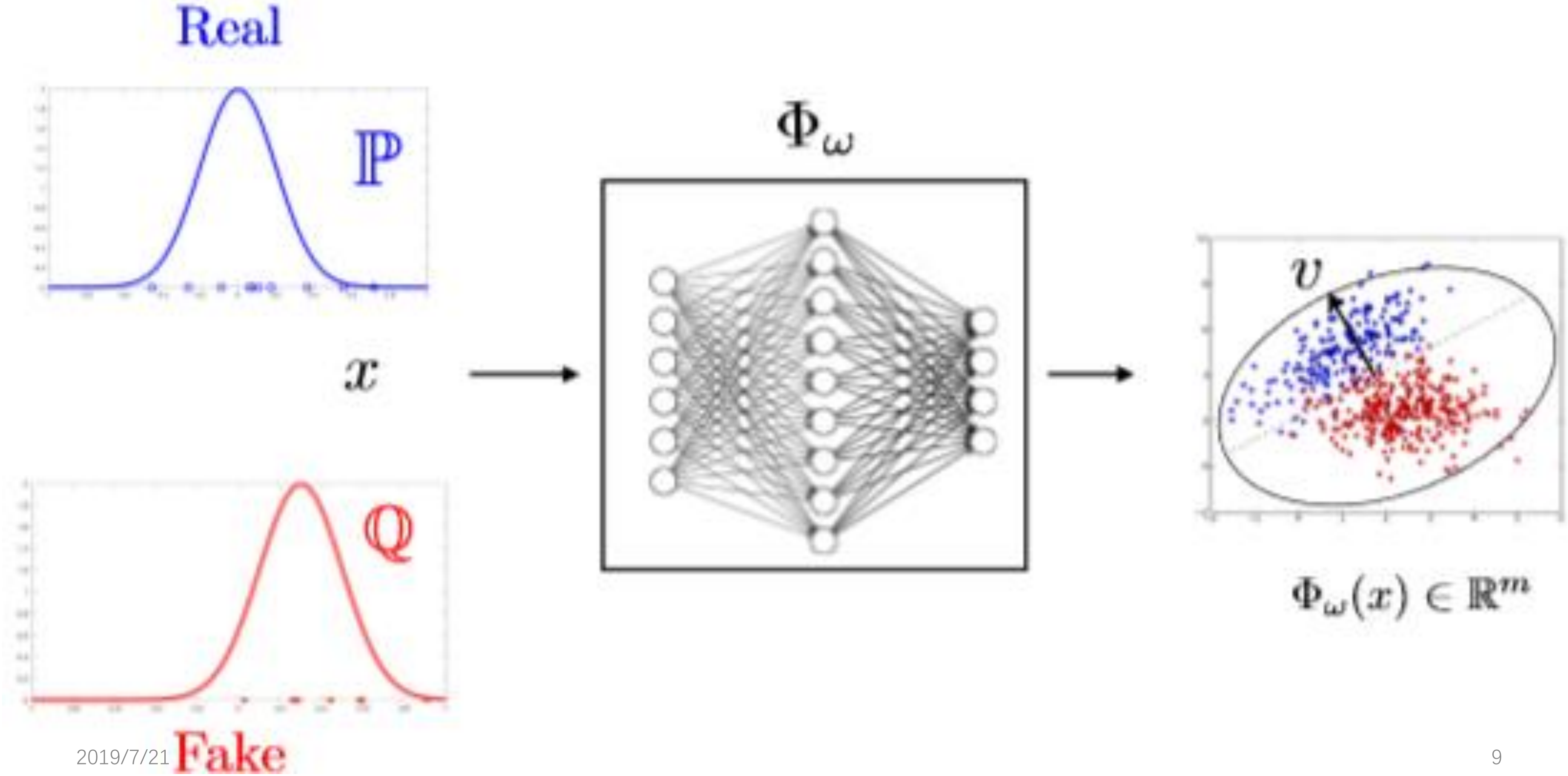
- Constraint

$$\hat{\Omega}(f_p, g_\theta) = \frac{1}{2N}\sum_{i=1}^{N} f_p^2(x_i) + \frac{1}{2M}\sum_{i=1}^{M} f_p^2(g_\theta(z_j)) = 1$$

Fisher IPM with Neural Networks

- $\mathcal{F}$ is a finite dimensional Hilbert space induced by a neural network Φ_ω

Rayleigh Quotient form



Rayleigh Quotient form

- Fisher IPM defined on $\mathcal{F}_{v,\omega}$:

$$d_{\mathscr{F}_{v,\omega}}(\mathbb{P},\mathbb{Q}) = \max_{\omega} \max_{v} \frac{\langle v, \mu_\omega(\mathbb{P}) - \mu_\omega(\mathbb{Q})\rangle}{\sqrt{v^\top(\frac{1}{2}\Sigma_\omega(\mathbb{P}) + \frac{1}{2}\Sigma_\omega(\mathbb{Q}) + \gamma I_m)v}}$$

$$\Sigma_\omega(\mathbb{P}) = \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left(\Phi_\omega(x)\Phi_\omega(x)^\top\right) \qquad \mu_\omega(\mathbb{P}) = \mathop{\mathbb{E}}_{x\sim\mathbb{P}}(\Phi_\omega(x))$$

$*$ We add a regularization term to avoid singularity of the covariance.

Constrained Form

- 原始形式

$$d_{\mathscr{F}_{v,\omega}}(\mathbb{P},\mathbb{Q}) = \max_{\omega, v, v^\top(\frac{1}{2}\Sigma_\omega(\mathbb{P})+\frac{1}{2}\Sigma_\omega(\mathbb{Q})+\gamma I_m)v=1} \langle v, \mu_\omega(\mathbb{P}) - \mu_\omega(\mathbb{Q})\rangle$$

- 变形

记协方差为 $\Sigma_\omega(\mathbb{P};\mathbb{Q}) = \frac{1}{2}\Sigma_\omega(\mathbb{P}) + \frac{1}{2}\Sigma_\omega(\mathbb{Q}) + \gamma I_m$

令 $u = (\Sigma_\omega(\mathbb{P};\mathbb{Q}))^{\frac{1}{2}} v$

推出

$$\begin{aligned} d_{\mathscr{F}_{u,\omega}}(\mathbb{P},\mathbb{Q}) &= \max_{\omega} \max_{u, \|u\|=1} \left\langle u, (\Sigma_\omega(\mathbb{P};\mathbb{Q}))^{-\frac{1}{2}} (\mu_\omega(\mathbb{P}) - \mu_\omega(\mathbb{Q}))\right\rangle \\ &= \max_{\omega} \left\|(\Sigma_\omega(\mathbb{P};\mathbb{Q}))^{-\frac{1}{2}} (\mu_\omega(\mathbb{P}) - \mu_\omega(\mathbb{Q}))\right\| \end{aligned}$$

另一种形式 $d_{\mathscr{F}_{v,\omega}}(\mathbb{P},\mathbb{Q}) = \max_{\omega} \sqrt{(\mu_\omega(\mathbb{P}) - \mu_\omega(\mathbb{Q}))^\top \Sigma_\omega^{-1}(\mathbb{P};\mathbb{Q})(\mu_\omega(\mathbb{P}) - \mu_\omega(\mathbb{Q}))}$

Conclusion

- The Rayleigh Quotient is not amenable to optimization, we will consider Fisher IPM as a constrained optimization problem

$$\min_{g_\theta} \max_{\omega} \max_{v, v^\top\left(\frac{1}{2}\Sigma_\omega(\mathbb{P}_r)+\frac{1}{2}\Sigma_\omega(\mathbb{P}_\theta)+\gamma I_m\right)v=1} \langle v, \mu_w(\mathbb{P}_r) - \mu_\omega(\mathbb{P}_\theta)\rangle$$

- Learning GAN with Fisher IPM: define the Augmented Lagrangian corresponding to Fisher GAN objective and constraint

$$\mathcal{L}_F(p, \theta, \lambda) = \hat{\mathscr{E}}(f_p, g_\theta) + \lambda(1 - \hat{\Omega}(f_p, g_\theta)) - \frac{\rho}{2}(\hat{\Omega}(f_p, g_\theta) - 1)^2$$

$$\hat{\mathscr{E}}(f_p, g_\theta) := \frac{1}{N}\sum_{i=1}^{N} f_p(x_i) - \frac{1}{M}\sum_{j=1}^{M} f_p(g_\theta(z_j))$$

$$\hat{\Omega}(f_p, g_\theta) = \frac{1}{2N}\sum_{i=1}^{N} f_p^2(x_i) + \frac{1}{2M}\sum_{j=1}^{M} f_p^2(g_\theta(z_j))$$

Conclusion

- Algorithm

Algorithm 1 Fisher GAN

Input: ρ penalty weight, η Learning rate, n_c number of iterations for training the critic, N batch size
Initialize $p, \theta, \lambda = 0$
repeat
 for $j = 1$ **to** n_c **do**
  Sample a minibatch $x_i, i = 1 \dots N, x_i \sim \mathbb{P}_r$
  Sample a minibatch $z_i, i = 1 \dots N, z_i \sim p_z$
  $(g_p, g_\lambda) \leftarrow (\nabla_p \mathcal{L}_F, \nabla_\lambda \mathcal{L}_F)(p, \theta, \lambda)$
  $p \leftarrow p + \eta$ ADAM (p, g_p)
  $\lambda \leftarrow \lambda - \rho g_\lambda$ {SGD rule on λ with learning rate ρ}
 end for
 Sample $z_i, i = 1 \dots N, z_i \sim p_z$
 $d_\theta \leftarrow \nabla_\theta \hat{\mathscr{E}}(f_p, g_\theta) = -\nabla_\theta \frac{1}{N} \sum_{i=1}^{N} f_p(g_\theta(z_i))$
 $\theta \leftarrow \theta - \eta$ ADAM (θ, d_θ)
until θ converges

Thanks