

# RI-GAN

Kosta Grujić, Momčilo Knežević

Matematički fakultet

## Sažetak

## 1 Uvod

## 2 *Related Work*

U ovom poglavlju ćemo opisati princip rada GAN-ova i metrika koje ćemo nadalje koristiti.

### 2.1 Generativne suparničke mreže

Prvo dajemo neformalnu ideju. Osnovni cilj je naučiti raspodelu podataka tako da je moguće vršiti uzorkovanje, čime se dobija mogućnost generisanja podataka koji zapravo ne postoje. GAN arhitektura podrazumeva upotrebu dva modela. Prvi je zadužen za učenje pomenute raspodele, dok drugi treba da razlikuje stvarne i generisane podatke. Suparničkim učenjem ova dva modela se stvarna raspodela podataka sve bolje aproksimira.

Neka je  $\mathcal{X} = \mathbb{R}^{d \times d}$  prostor slika dimenzija  $d \times d$ . Bilo koji skup slika možemo predstaviti nekom raspodelom  $p_r$  nad  $\mathcal{X}$ . Skup nad kojim obučavamo model možemo definisati kao prost slučajan uzorak iz raspodele  $p_r$ . Cilj je naučiti parametrizovanu raspodelu  $p_g$  koja aproksimira  $p_r$ .

GAN se sastoji od *generatora* i *diskriminatora*. Diskriminator posmatramo kao preslikavanje  $D : \mathcal{X} \rightarrow [0, 1]$ , dok generator kao  $G : \mathcal{Z} \rightarrow \mathcal{X}$ , gde je  $\mathcal{Z}$  latentni prostor. Pomenuti latentni prostor služi za uzorkovanje šuma na osnovu kojeg generator pokušava da generiše elemente raspodele  $p_r$ . Ako takvo uzorkovanje šuma vršimo iz raspodele  $p_z$  definisane nad  $\mathcal{Z}$ , tada raspodelu  $p_g$  možemo definisati kao  $G(p_z)$ .

Napomenimo male vrednosti diskriminatora odgovaraju generisanim podacima, dok veće stvarnim. Prema tome, generator nastoji generisanju podataka za koje diskriminator daje vrednosti bliske 0, dok diskriminator za takve ulaze nastoji da se sve viši približi vrednosti 1.

Obučavanje diskriminatora i generatora se vrši uporedo, usled čega se može videti kao min-max igra. Takav proces se može formalizovati:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_r} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log(1 - D(G(z)))]. \quad (1)$$

Primetimo da ne postoji ograničenje za izbor modela  $D$  i  $G$ . Međutim, pretpostavljajući njihovu diferencijabilnost i parametrizovanost, obučavanje se može vršiti propagacijom unazad i time u velikoj meri automatizovati.

U početnim iteracijama obučavanja može doći do zasićenja gradijenata [1]. Usled lošeg generatora, diskriminator je u stanju vrlo precizno da razlikuje lažne i prave podatke. S tim u vezi, optimizacija iz formule 1 se može zapisati u numerički pogodnijem obliku:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_r} [\log D(x)] - \mathbb{E}_{z \sim p_z} [\log D(G(z))]. \quad (2)$$

## 2.2 DCGAN

## 2.3 IS

Aдекватno obučen GAN poseduje sigurnost i varijabilnost. To znači da generator može dati svojstvene podatke, kao i da ravnomerno pokriva čitav kombinatorni prostor.

Prepostavimo da svakoj instanci skupa podataka  $x$  možemo pridružiti oznaku  $y$ . Skup podataka takve oznake ne mora da sadrži, ali ih je moguće uvesti dodatnim modelom  $\mathcal{M}$  (kao što je Inception [10]) koji je treniran kao klasifikacioni model na nekom drugom skupu podataka (kao što je ImageNet [11]). Zato možemo posmatrati uslovnu raspodelu  $p_{\mathcal{M}}(y|x)$ . Marginalna raspodela  $p_{\mathcal{M}}(y)$  se dobija kao  $\int_x p_{\mathcal{M}}(y|x) dp_g(x)$ .

Prema tome,  $p_g(y|x)$  treba da odgovara Dirakovoj  $\delta$  funkciji, a  $p_g(y)$  ravnomernoj raspodeli. Konačno, IS<sup>1</sup> definišemo:

$$\text{IS}(p_g) = \exp(\mathbb{E}_{x \sim p_g} [KL(p_g(y|x) || p_g(y))]), \quad (3)$$

gde je  $KL$  Kulbek-Lajblerova divergencija.

Ispostavlja se da ova metrika dobro korelira sa ljudskom odlukom [7]. Međutim, IS ni na koji način ne koristi  $p_r$ . Takođe, skup podataka nad kojim je  $\mathcal{M}$  treniran možda nema ni približno sličnu raspodelu kao  $p_r$  usled čega GAN model biva proglašen lošim.

---

<sup>1</sup>eng. *Inception score*

## 2.4 FID

Kao što smo već rekli, ideja obučavanja GAN-a je što bolja aproksimacija  $p_r$  raspodelom  $p_g$ . Ukoliko  $p_r$  predstavlja raspodelu podataka iz realnog sveta (poput slika ljudi), tada možemo pretpostaviti ograničenost njenog nosača. U tom slučaju važi:

$$p_r \stackrel{s.s.}{=} p_g \iff (\forall k \in \mathbb{N}) \int_{\mathcal{X}} p_r(x) x^k dx = \int_{\mathcal{X}} p_g(x) x^k dx. \quad (4)$$

Drugim rečima, jednakost momenata povlači skoro sigurnu jednakost raspodela. U praksi je desnu stranu ekvivalencije 4 skoro nemoguće pokazati. Zato se ona aproksimira za vrednosti do nekog  $k$ . Mi ćemo aproksimaciju vršiti za  $k = 2$ .

Kako raspodele  $p_r$  i  $p_g$  nisu eksplicitno poznate, uvodimo operator  $\phi$  tako da jednakost  $\phi(p_r)$  i  $\phi(p_g)$  odgovara jednakosti odgovarajućih raspodela. Kao i kod IS, koristimo Inception model čiji će neki unutrašnji sloj predstavljati operator  $\phi$ . Inception model obučen nad ImageNet skupu podataka poseduje kvalitetnu aproksimaciju raspodele slika iz stvarnog sveta čiji duboki slojevi poseduju vizuelne karakteristike visokog nivoa. Iz tog razloga, za očekivati je da forsiranjem sličnosti  $\phi(p_r)$  i  $\phi(p_g)$  postizemo sličnost  $p_r$  i  $p_g$ .

Međutim, raspodele  $\phi(p_r)$  i  $\phi(p_g)$  takođe nisu eksplicitno poznate. Poznato je da za fiksirane momente prva dva reda normalna raspodela ima maksimalnu entropiju. Zato pretpostavljamo:  $\phi(p_r) = \mathcal{N}(\mu_r, \Sigma_r)$  i  $\phi(p_g) = \mathcal{N}(\mu_g, \Sigma_g)$ . Razliku ovih raspodela merimo Frešeovim rastojanjem:

$$F(\phi(p_r), \phi(p_g)) = \|\mu_r - \mu_g\| + \text{tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}), \quad (5)$$

gde je  $\text{tr}$  operator traga matrice.

Konačno, definišemo  $\text{FID}^2(p_r, p_g)$  kao  $F(\phi(p_r), \phi(p_g))$  [8]. Primetimo da za razliku od IS, FID koristi  $p_r$ .

---

<sup>2</sup>eng. *Fréchet Inception distance*

### 3 Kvantifikovanje obučenosti

### 4 Eksperimenti

### 5 Zaključak

## Literatura

- [1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. Generative Adversarial Nets. *arXiv preprint arXiv:1406.2661*, 2014.
- [2] Qiantong Xu, Gao Huang, Yang Yuan, Chuan Guo, Yu Sun, Felix Wu, Kilian Weinberger. An empirical study on evaluation metrics of generative adversarial networks. *arXiv preprint arXiv:1806.07755*, 2018.
- [3] Alec Radford, Luke Metz, Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv preprint arXiv:1511.06434*, 2016.
- [4] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, Yuichi Yoshida. Spectral Normalization for Generative Adversarial Networks. *arXiv preprint arXiv:1802.05957*, 2018.
- [5] Martin Arjovsky, Soumith Chintala, Léon Bottou. Wasserstein GAN. *arXiv preprint arXiv:1701.07875*, 2017.
- [6] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, Aaron Courville. Improved Training of Wasserstein GANs. *arXiv preprint arXiv:1704.00028*, 2017.
- [7] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, Xi Chen. Improved Techniques for Training GANs. *arXiv preprint arXiv:1606.03498*, 2016.
- [8] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernard Nessler. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *NIPS*, 2017.
- [9] Ziwei Liu, Ping Luo, Xiaogang Wang, Xiaoou Tang. Deep Learning Face Attributes in the Wild. *Proceedings of International Conference on Computer Vision (ICCV)*, 2015.

- [10] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich. Going Deeper with Convolutions. *arXiv preprint arXiv:1409.4842*, 2014.
- [11] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *arXiv preprint arXiv:1409.0575*, 2014.