Generative Adversarial Networks

Davi Barreira

FGV - Escola de Matemática Aplicada

Table of contents

Introdução

Generative Adversarial Networks (GAN) foram originalmente introduzidas por Goodfellow et al. (2014). Essas redes são utilizadas com o objetivo de gerar dados sintéticos realísticos a partir de dados reais.

Figure 1: Faces geradas por GAN [1].

[1]https://towardsdatascience.com/generating-modern-arts-using-generative-adversarial-network-gan-on-spell-39f67f83c7b4

Introdução

A geração de novas amostras sintéticas tem diferentes utilidades, como aprendizado semi-supervisionado, geração de exemplos adversariais, *style transfer*, entre outros.

Figure 2: Style transfer utilizando CycleGan [2].

[2]https://towardsdatascience.com/style-transfer-with-gans-on-hd-images-88e8efcf3716

A ideia geral por trás das GANs é utilizar duas redes neurais competindo uma com a outra, sendo uma rede responsável por gerar amostras parecidas com os dados reais, enquanto a outra busca identificar quando o dado é real ou sintético.

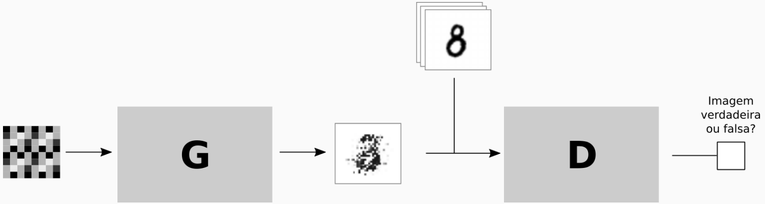

Figure 3: Desenho esquemático de uma GAN "convencional".

Entendendo a Função Objetivo

$$V(D,G) = \mathbb{E}_{x \;\, p_{data}(x)}\left[\log\left(D(x)\right)\right] + \mathbb{E}_{z \;\, p_z(x)}\left[1 - \log\left(D(D(z))\right)\right]$$

Entendendo a Função Objetivo

$$V(D, G) = \mathbb{E}_{x \quad p_{data}(x)} \left[\log\left(D(x)\right)\right] + \mathbb{E}_{z \quad p_z(x)} \left[1 - \log\left(D(D(z))\right)\right]$$

$$= \int_x p_{data}(x) \log\left(D(x)\right) dx + \int_z p_z(z) \log\left(1 - D(G(z))\right) dz$$

Entendendo a Função Objetivo

$$V(D, G) = \mathbb{E}_{x\ \ p_{data}(x)}\left[\log\left(D(x)\right)\right] + \mathbb{E}_{z\ \ p_{z}(x)}\left[1 - \log\left(D(D(z))\right)\right]$$

$$= \int_x p_{data}(x)\log\left(D(x)\right)dx + \int_z p_z(z)\log\left(1 - D(G(z))\right)dz$$

$$x = G(z) \rightarrow z = G^{-1}(x) \rightarrow dz = (G^{-1}(x))dx$$

$$\rightarrow p_g(x) = p_z(G^{-1})(G^{-1})'(x)dx$$

Entendendo a Função Objetivo

$$V(D, G) = \mathbb{E}_{x \;\, p_{data}(x)} [\log (D(x))] + \mathbb{E}_{z \;\, p_z(x)} [1 - \log (D(D(z)))]$$

$$= \int_x p_{data}(x) \log (D(x)) dx + \int_z p_z(z) \log (1 - D(G(z))) dz$$

$$x = G(z) \rightarrow z = G^{-1}(x) \rightarrow dz = (G^{-1}(x))dx$$
$$\rightarrow p_g(x) = p_z(G^{-1})(G^{-1})'(x)dx$$

$$= \int_x p_{data}(x) \log (D(x)) dx + \int_x p_z(G^{-1}(x)) \log (1 - D(x))(G^{-1})'(x)dx$$

Entendendo a Função Objetivo

$$V(D, G) = \mathbb{E}_{x \;\; p_{data}(x)} \left[\log\left(D(x)\right)\right] + \mathbb{E}_{z \;\; p_z(x)} \left[1 - \log\left(D(D(z))\right)\right]$$

$$= \int_x p_{data}(x) \log\left(D(x)\right) dx + \int_z p_z(z) \log\left(1 - D(G(z))\right) dz$$

$$x = G(z) \rightarrow z = G^{-1}(x) \rightarrow dz = (G^{-1}(x)) dx$$
$$\rightarrow p_g(x) = p_z(G^{-1})(G^{-1})'(x) dx$$

$$= \int_x p_{data}(x) \log\left(D(x)\right) dx + \int_x p_z(G^{-1}(x)) \log\left(1 - D(x)\right)(G^{-1})'(x) dx$$

$$= \int_x p_{data}(x) \log\left(D(x)\right) dx + \int_x p_g(x) \log\left(1 - D(x)\right) dx$$

Entendendo a Função Objetivo

$$V(D,G) = \mathbb{E}_{x \quad p_{data}(x)}\left[\log\left(D(x)\right)\right] + \mathbb{E}_{z \quad p_z(x)}\left[1 - \log\left(D(D(z))\right)\right]$$

$$= \int_x p_{data}(x)\log\left(D(x)\right)dx + \int_z p_z(z)\log\left(1 - D(G(z))\right)dz$$

$$x = G(z) \rightarrow z = G^{-1}(x) \rightarrow dz = (G^{-1}(x))dx$$
$$\rightarrow p_g(x) = p_z(G^{-1})(G^{-1})'(x)dx$$

$$= \int_x p_{data}(x)\log\left(D(x)\right)dx + \int_x p_z(G^{-1}(x))\log\left(1 - D(x)\right)(G^{-1})'(x)dx$$

$$= \int_x p_{data}(x)\log\left(D(x)\right)dx + \int_x p_g(x)\log\left(1 - D(x)\right)dx$$

$$= \int_x p_{data}(x)\log\left(D(x)\right) + p_g(x)\log\left(1 - D(x)\right)dx$$

Entendendo a Função Objetivo

$$\max_D V(D, G) = \max_D \int_x p_{data}(x) \log\left(D(x)\right) + p_g(x) \log\left(1 - D(x)\right) dx$$

Entendendo a Função Objetivo

$$\max_D V(D, G) = \max_D \int_x p_{data}(x) \log\left(D(x)\right) + p_g(x) \log\left(1 - D(x)\right) dx$$

$$\frac{\partial}{\partial D(x)} \left(p_{data}(x) \log\left(D(x)\right) + p_g(x) \log\left(1 - D(x)\right)\right) = 0$$

Entendendo a Função Objetivo

$$\max_D V(D, G) = \max_D \int_x p_{data}(x) \log\left(D(x)\right) + p_g(x) \log\left(1 - D(x)\right) dx$$

$$\frac{\partial}{\partial D(x)} \left(p_{data}(x) \log\left(D(x)\right) + p_g(x) \log\left(1 - D(x)\right)\right) = 0$$

$$\rightarrow \frac{p_{data}(x)}{D(x)} - \frac{p_g(x)}{1 - D(x)} = 0$$

$$\rightarrow D(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}$$

Entendendo a Função Objetivo

Suponha que o discriminador $D_G^*(x)$ seja ótimo
Então o gerador ótimo produz $p_g(x) = p_{data}(x)$

$$\rightarrow D_G^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}$$

Entendendo a Função Objetivo

$$C(G) = \max_D V(G, D)$$

Entendendo a Função Objetivo

$$C(G) = \max_D V(G, D)$$

$$= \max_D \int_x p_{data}(x) \log D(x) + p_g \log 1 - D(x) dx$$

Entendendo a Função Objetivo

$$C(G) = \max_D V(G, D)$$

$$= \max_D \int_x p_{data}(x) \log D(x) + p_g \log 1 - D(x) dx$$

$$= \int_x p_{data}(x) \log \left(D_G^*(x)\right) + p_g(x) \log \left(1 - D_g^*(x)\right) dx$$

Entendendo a Função Objetivo

$$C(G) = \max_D V(G, D)$$

$$= \max_D \int_x p_{data}(x) \log D(x) + p_g \log 1 - D(x) dx$$

$$= \int_x p_{data}(x) \log \left(D_G^*(x)\right) + p_g(x) \log \left(1 - D_g^*(x)\right) dx$$

$$= \int_x p_{data}(x) \log \left(\frac{p_{data}(x)}{p_{data}(x) + p_g(x)} \right) + p_g(x) \log \left(\frac{p_{data}(x)}{p_{data}(x) + p_g(x)} \right) dx$$

Entendendo a Função Objetivo

$$C(G) = \max_D V(G, D)$$

$$= \max_D \int_x p_{data}(x) \log D(x) + p_g \log 1 - D(x) dx$$

$$= \int_x p_{data}(x) \log \left(D_G^*(x)\right) + p_g(x) \log \left(1 - D_g^*(x)\right) dx$$

$$= \int_x p_{data}(x) \log \left(\frac{p_{data}(x)}{p_{data}(x) + p_g(x)}\right) + p_g(x) \log \left(\frac{p_{data}(x)}{p_{data}(x) + p_g(x)}\right) dx$$

$$= \int_x p_{data}(x) \log \left(\frac{p_{data}(x)}{\frac{p_{data}(x) + p_g(x)}{2}}\right) + p_g(x) \log \left(\frac{p_{data}(x)}{\frac{p_{data}(x) + p_g(x)}{2}}\right) dx$$

Entendendo a Função Objetivo

$$C(G) = \max_D V(G, D)$$

$$= \max_D \int_x p_{data}(x) \log D(x) + p_g \log 1 - D(x) dx$$

$$= \int_x p_{data}(x) \log \left(D_G^*(x)\right) + p_g(x) \log \left(1 - D_g^*(x)\right) dx$$

$$= \int_x p_{data}(x) \log \left(\frac{p_{data}(x)}{p_{data}(x) + p_g(x)}\right) + p_g(x) \log \left(\frac{p_{data}(x)}{p_{data}(x) + p_g(x)}\right) dx$$

$$= \int_x p_{data}(x) \log \left(\frac{p_{data}(x)}{\frac{p_{data}(x) + p_g(x)}{2}}\right) + p_g(x) \log \left(\frac{p_{data}(x)}{\frac{p_{data}(x) + p_g(x)}{2}}\right) dx$$

$$= KL\left[p_{data}(x) || \frac{p_{data}(x) + p_g(x)}{2}\right] + KL\left[p_g(x) || \frac{p_{data}(x) + p_g(x)}{2}\right] - \log 4$$

Entendendo a Função Objetivo

$$C(G) = KL\left[p_{data}(x)||\frac{p_{data}(x)+p_g(x)}{2}\right] + KL\left[p_g(x)||\frac{p_{data}(x)+p_g(x)}{2}\right] - \log 4$$

Entendendo a Função Objetivo

$$C(G) = KL\left[p_{data}(x)||\frac{p_{data}(x)+p_g(x)}{2}\right] + KL\left[p_g(x)||\frac{p_{data}(x)+p_g(x)}{2}\right] - \log 4$$

$$\min_G C(G) = 0 + 0 - \log 4 = -\log 4$$

$$KL\left[p_{data}(x)||\frac{p_{data}(x)+p_g(x)}{2}\right] = 0$$

Entendendo a Função Objetivo

$$C(G) = KL\left[p_{data}(x)||\frac{p_{data}(x)+p_g(x)}{2}\right] + KL\left[p_g(x)||\frac{p_{data}(x)+p_g(x)}{2}\right] - \log 4$$

$$\min_G C(G) = 0 + 0 - \log 4 = -\log 4$$

$$KL\left[p_{data}(x)||\frac{p_{data}(x)+p_g(x)}{2}\right] = 0$$

quando, $p_{data}(x) = \dfrac{p_{data}(x)+p_g(x)}{2}$

Entendendo a Função Objetivo

$$C(G) = KL\left[p_{data}(x) || \frac{p_{data}(x) + p_g(x)}{2}\right] + KL\left[p_g(x) || \frac{p_{data}(x) + p_g(x)}{2}\right] - \log 4$$

$$\min_G C(G) = 0 + 0 - \log 4 = -\log 4$$

$$KL\left[p_{data}(x) || \frac{p_{data}(x) + p_g(x)}{2}\right] = 0$$

quando, $p_{data}(x) = \dfrac{p_{data}(x) + p_g(x)}{2}$

$$\rightarrow p_{data}(x) = p_g(x)$$

KL (Kullback-Leibler) divergence

Jensen-Shannon Divergency (symetric KL):

$$JSD(P||Q) = \frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_K L(Q||M),$$

$$M = \frac{1}{2}(P + Q)$$

Resumo:

$$V(G,D) = \mathbb{E}_{x \ p_{data}(x)}\left[\log\left(D(x)\right)\right] + \mathbb{E}_{z \ p_z(x)}\left[1 - \log\left(D(D(z))\right)\right]$$

Gerador G, Discriminador D

Encontro G^*, tal que

$$G^* = \arg\min_{G}\max_{D} V(G,D)$$

Dado $G, \max_D V(G,D)$

$$= -2\log 2 + 2JSD(P_{data}(x)||P_g(x))$$

Algorithm 1: GAN descrita em Goodfellow et al. (2014)

for *número de iterações de treino* **do**

  for *k passos* **do**

    Amostre m valores $\{\boldsymbol{z}^{(1)}, ..., \boldsymbol{z}^{(m)}\}$ da priori $p_z(\boldsymbol{z})$;

    Amostre m exemplos $\{\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(m)}\}$ da fução dos dados $p_{data}(\boldsymbol{x})$;

    Atualize o *discriminator* utilizando *stochastic gradient descent*:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[log D(\boldsymbol{x}^{(i)}) + log(1 - D(G(\boldsymbol{z}^{(i)}))) \right]$$

  end

  Amostre m valores $\{\boldsymbol{z}^{(1)}, ..., \boldsymbol{z}^{(m)}\}$ da priori $p_z(\boldsymbol{z})$;

  Atualize o *generator* utilizando *stochastic gradient descent*:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} log(1 - D(G(\boldsymbol{z}^{(i)})))$$

end

References i

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc.