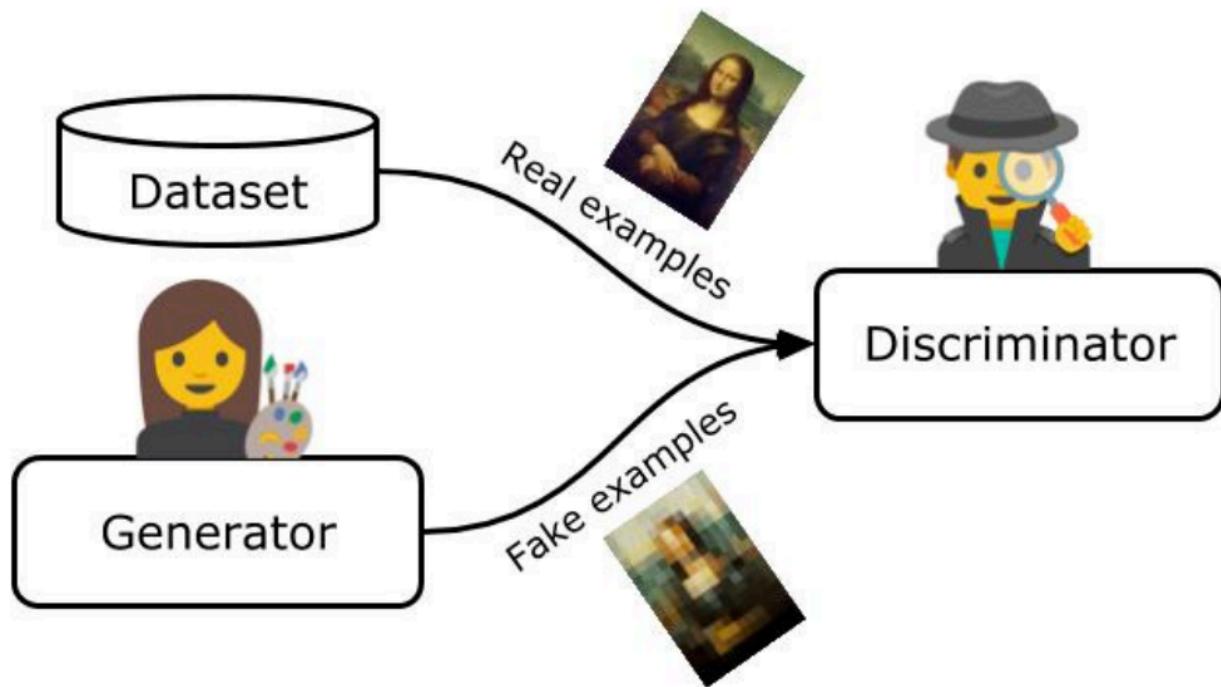


به نام خالق زیبایی ها

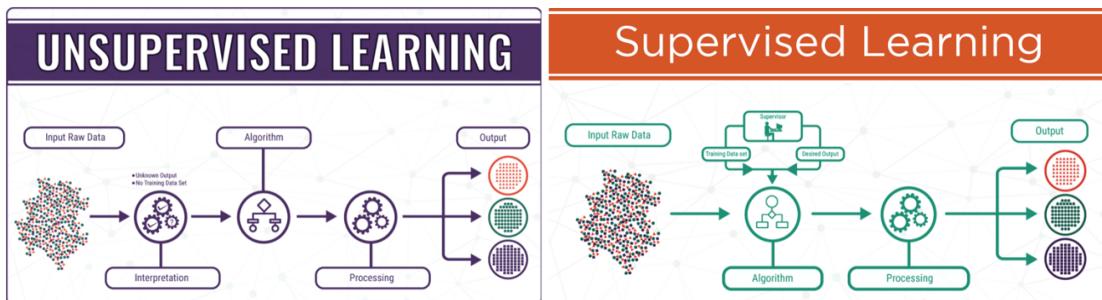
دستور کار آزمایشگاه یادگیری عمیق(بینایی ماشین)



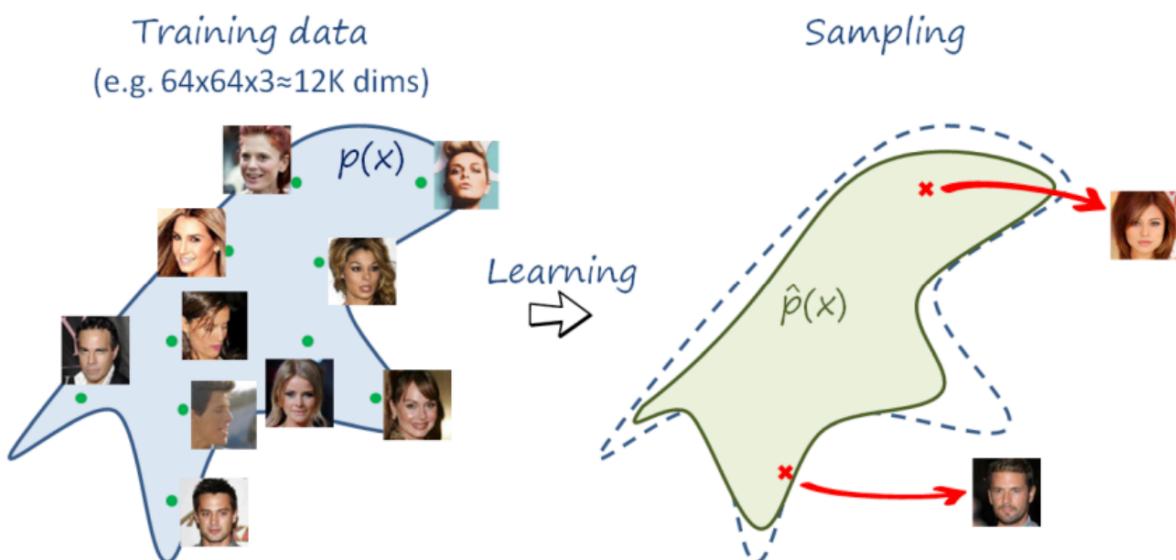
GAN

آزمایش : شبکه مولد تخصصی (GAN)

روش های مورد بررسی در یادگیری عمیق تا به اینجای کار به دو دسته عمده تقسیم بندی میشود. یادگیری با نظارت^۱ که در آن سعی بر یادگیری تابعی از فضای فرضیه به منظور مپ کردن داده های آموزش بر مقدار یا لیبل هر نمونه، داشتیم. از سوی دیگر یادگیری بدون نظارت^۲ سعی در یادگیری ساختار نهان داده هایی که برچسب ندارند، دارد.



مدل های مولد^۳ سعی در یادگیری توزیع داده به منظور تولید نمونه هایی با خصوصیات نمونه های آموزش است به عبارتی یادگیری تابع چگالی توزیع داده به شرط داده های آموزش را مورد بررسی قرار میدهد.



Generative modeling and sampling

مدل های مولد معمولا دوراهکار را برای تخمین توزیع داده ها در نظر میگیرند. در یک روش به صورت صریح با توزیع داده کار میشود در روش دیگر خیر که دیاگرام آن مطابق شکل زیر است:

دو رویکرد متفاوت در مدل های مولد را به همراه مزایا و معایب در صورت وجود توضیح دهید.



¹ Supervised learning

² Unsupervised learning

³ Generative

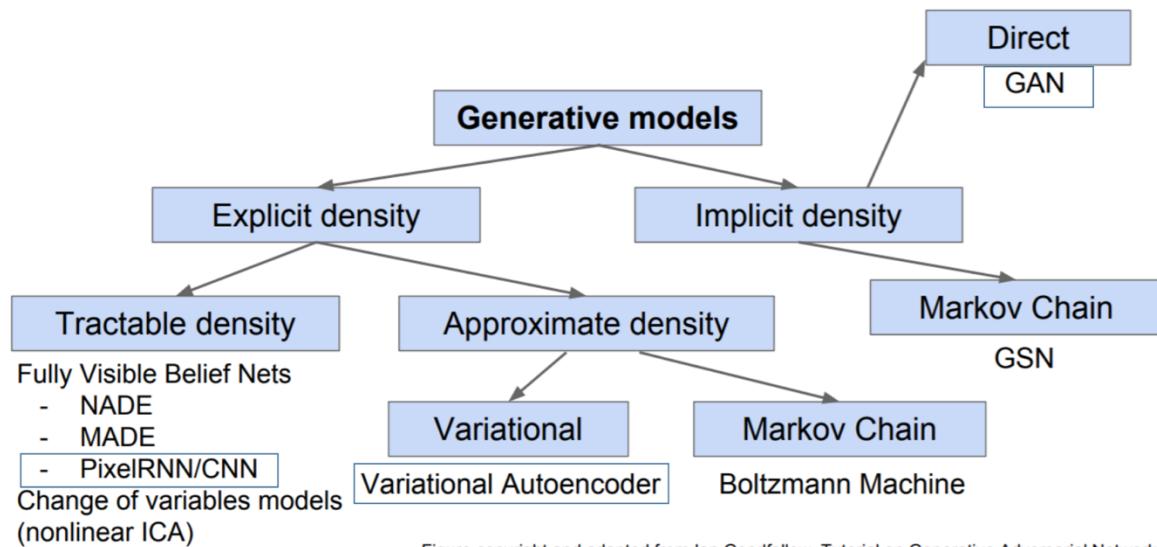


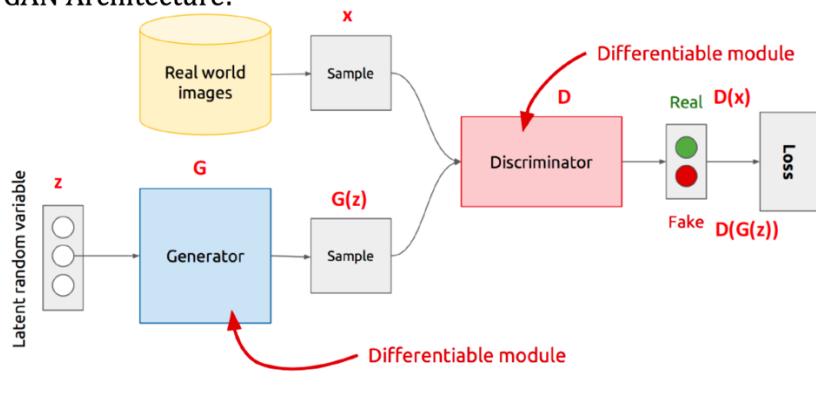
Figure copyright and adapted from Ian Goodfellow, Tutorial on Generative Adversarial Networks, 2017.



در این آزمایش شبکه های مولد تخاصمی به طور خاص مورد بررسی قرار میگیرد و در آخر چند نمونه پیاده سازی میگردد.ایده شبکه های مولد تخاصمی برای اولین بار توسط Ian J. Goodfellow در مقاله Generative Adversarial Nets مطرح شد.در ادامه به بررسی دقیق تر این شبکه ها میپردازیم.

شبکه های مولد تخاصمی از دو بخش(شبکه) مولد^۴(G) و جداکننده^۵(D) تشکیل شده است به این نحو که شبکه مولد رودی نویز را که میتواند توزیع استاندارد یا یکنواخت دارد دریافت کرده و سعی در تولید داده از این توزیع دارد. در این حین شبکه جداکننده سعی در تشخیص داده ی تقلیبی(ساخته شبکه مولد) از داده های واقعی از طریق مشاهده این دو داده دارد به عبارت دیگر این شبکه یک کلاسبرند است که خروجی آن احتمال واقعی بودن داده است.

• GAN Architecture!



⁴ Generator

⁵ Discriminator

به عبارت دیگر G و D در فرآیند آموزش شبکه باید در مقابل یکدیگر قرار گرفته و رقابت کنند به این صورت که G سعی در فریب D دارد و همچنین D میخواهد حتی الامکان فریب نخورد و داده واقعی از تقلیلی را با دقت خوبی تشخیص دهد. مسأله بهینه سازی در این شبکه به صورت زیر است:

$$J^{(D)} = \mathbb{E}_{x \sim p_{data}} [\log D(x)] + \mathbb{E}_{z \sim p(z)} [\log(1 - D(G(z)))]$$

$$J^{(G)} = \mathbb{E}_{z \sim p(z)} [\log D(G(z))]$$

$$\theta_D^* = \max_{\theta_D} \mathbb{E}_{x \sim p_{data}} [\log D_{\theta_D}(x)] + \mathbb{E}_{z \sim p(z)} [\log(1 - D_{\theta_D}(G_{\theta_G}(z)))]$$

$$\theta_G^* = \max_{\theta_G} \mathbb{E}_{z \sim p(z)} [\log(D_{\theta_D}(G_{\theta_G}(z)))]$$

در نهایت مسئله بهینه سازی تابع هدف به صورت زیر درمی آید:

Minimax objective function:

$$\min_{\theta_g} \max_{\theta_d} \left[\mathbb{E}_{x \sim p_{data}} \log D_{\theta_d}(x) + \mathbb{E}_{z \sim p(z)} \log(1 - D_{\theta_d}(G_{\theta_g}(z))) \right]$$

Alternate between:

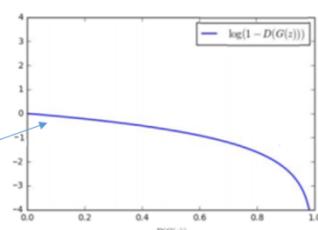
1. Gradient ascent on discriminator

$$\max_{\theta_d} \left[\mathbb{E}_{x \sim p_{data}} \log D_{\theta_d}(x) + \mathbb{E}_{z \sim p(z)} \log(1 - D_{\theta_d}(G_{\theta_g}(z))) \right]$$

2. Gradient descent on generator

$$\min_{\theta_g} \mathbb{E}_{z \sim p(z)} \log(1 - D_{\theta_d}(G_{\theta_g}(z)))$$

Gradient is relatively flat for likely fake samples while we intend to improve generator from them.



تابع هزینه بالا را تحلیل کنید. بدین صورت که در ابتدای فرآیند یادگیری و در انتهای آن خروجی D را تحلیل کنید. مشکل این تابع هزینه چیست؟ تابع هزینه پیشنهادی؟

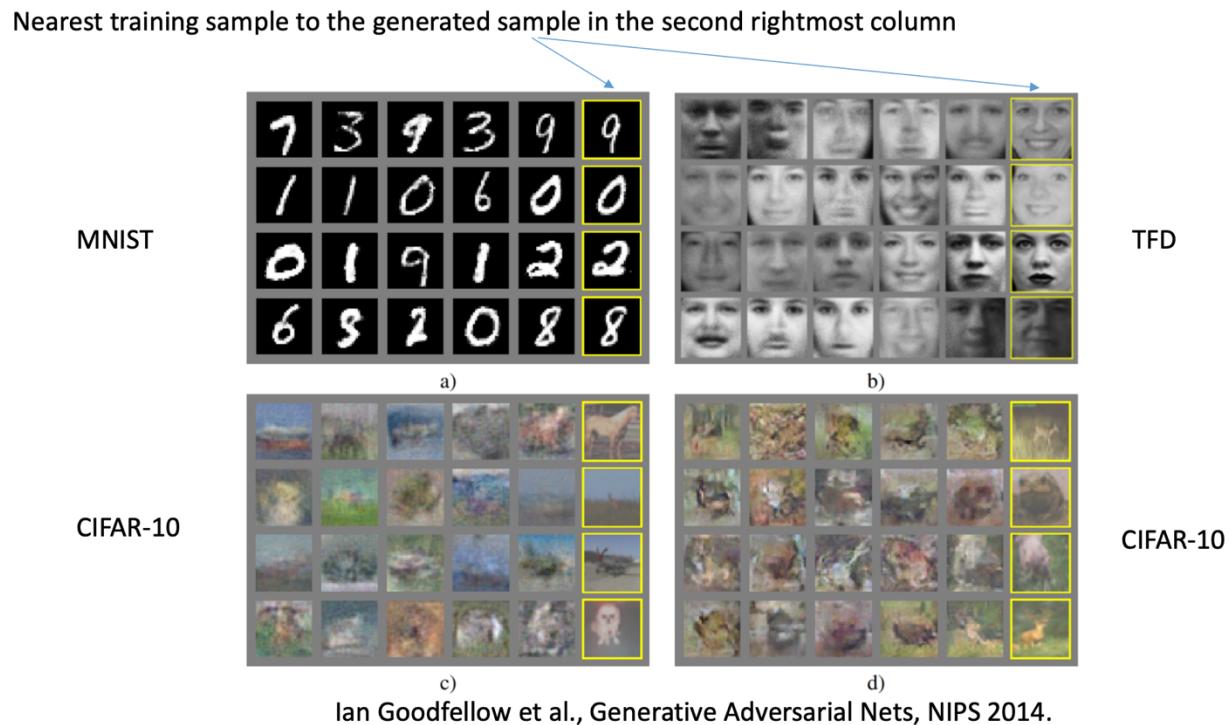
گرادیان در نقاط ابتدایی و انتهایی نمودار بالا را بررسی کنید.



(سوال امتیازی) برای G فیکس شده، D بهینه را محاسبه کنید. در این حالت مقدار تابع تلف D در این نقطه چه مقداری دارد؟



در شکل زیر یک نمونه از نمونه های تولیدی توسط شبکه GAN را مشاهده میکنید:

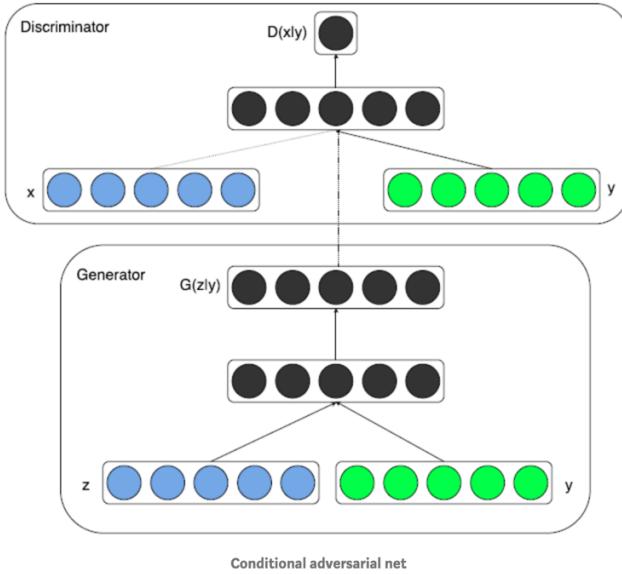


تعدادی از مشکلاتی که در مسئله بهینه سازی GAN با آن مواجهیم را بیان کنید و راه حل آن را نیز ارائه دهید.



Conditional GAN

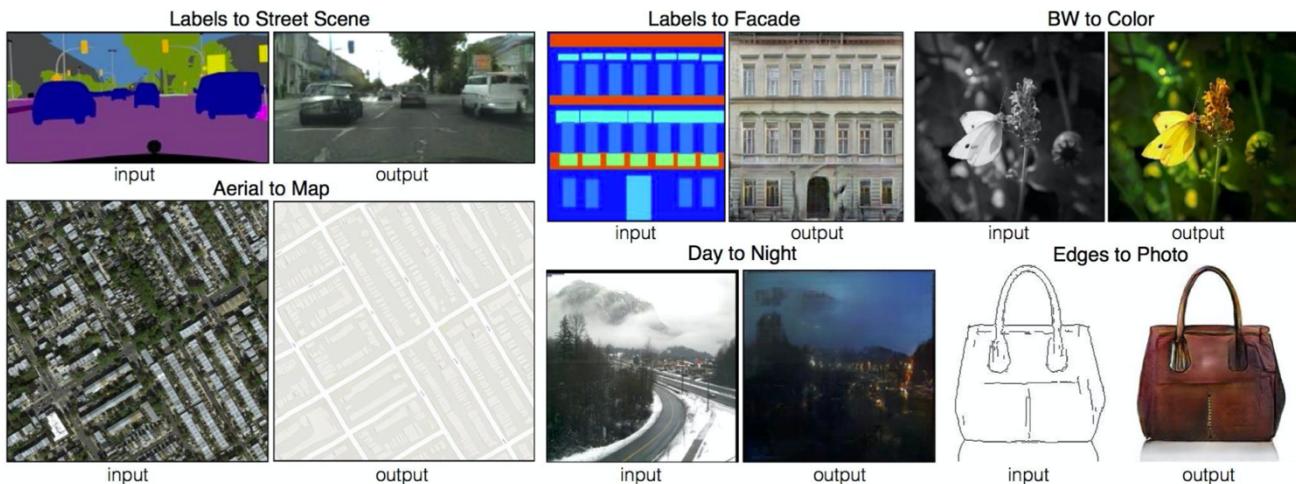
در صورتیکه G و D هر دو بر روی اطلاعاتی اضافه مانند y مشروط شوند. برای نمونه اگر بخواهیم بر روی خروجی G کنترل داشته باشیم که چه داده ای تولید کند باید داده (برچسب) لیبل مورد نظر به عنوان ورودی به G و D داده شود.



مسئله بهینه سازی آن به صورت زیر است :

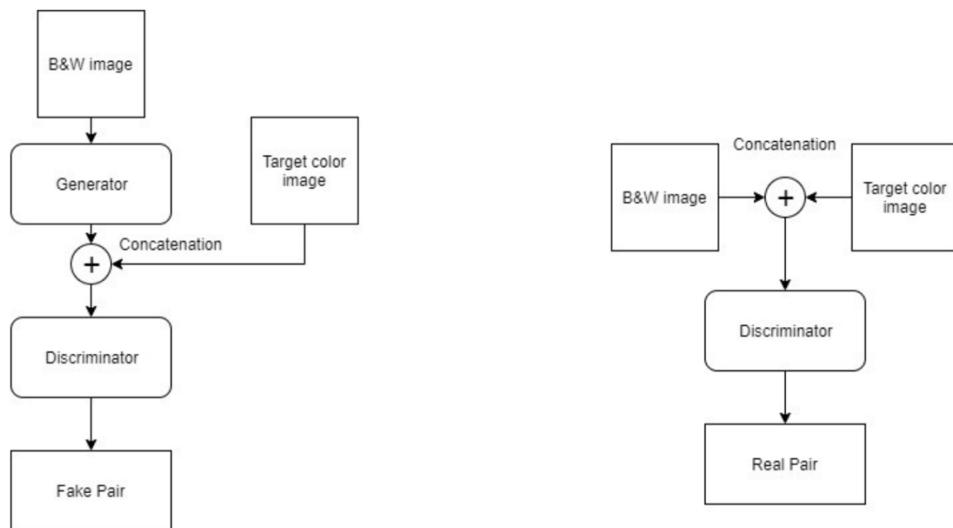
$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x}|\mathbf{y})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z}|\mathbf{y})))]$$

از کاربردهای CGAN میتوان به Image-to-Image Translation یا اصطلاحاً pix2pix نام برد که در آن برای مثال ورودی و خروجی به شکل زیر است:



: pix2pix GAN

معماری بخش مولد بر پایه U-net است و جداگتنده هم یک طبقه بند کانولوشنی PatchGAN است. برای نمونه برای تبدیل عکس های سیاه سفید به رنگی از ساختار زیر استفاده میشود:



Pix2Pix architecture for B&W to colour image generation

ساختار U-net را به صورت اجمالی توضیح دهد



تابع تلف را برای این نوع GAN بیان کنید.(به این لینک مراجعه کنید (<https://arxiv.org/abs/1611.07004>)

