ECS767P Emerging Topics in Machine Learning and Computer Vision, 2019

Course Work 2:
Unsupervised Learning by Generative Adversarial Network

1. **What is the difference between supervised learning & unsupervised learning in image classification task? (10% of CW2)**

Answer:

Basic difference between supervised & unsupervised is the 'label'. While in supervised image classification task, we could have class labels for each categories(e.g the number 0-9 in MNIST dataset). We can use loss(cost) function to directly correct our GD steps.

But in unsupervised task, we just have the distribution of the data with none labels. So we need to use algorithms to learn the distribution.

Interpret in statistical field, it could be explained by Bayes Model that supervised based on conditional probability(find the decision boundary) while unsupervised based on probability density(or similarity for different clusters).

2. **What is the difference between an auto-encoder and a generative adversarial network considering (1) model structure; (2) optimized objective function; (3) training procedure on different components. (10% of CW2)**

Answer:

(1)The encoder in VAE(Encoder + Decoder) maps the sample ***x*** from real-image dataset to latent code ***h***, then use ***h*** as the input of decoder to reconstruct the input ***r***. While in GAN (Generator +Discriminator), the input of the generator is Gaussian Random Noise(latent vector ***z***), the input of discriminator is ***G(z)*** and ***x*** in real set. The structure of VAE is more similar to a 'stream' and the structure of GAN is like a loop of battle game between G and D.

(2)The goal of VAE is to maximize the variational lower bound.

VAE	$L(x,y;\theta) = -\frac{1}{M}\sum_{i=1}^{M} \| x_i - r_i \|^2$	Here in VAE minimizes the L2 loss function to reconstruct. ***x*** is the input and ***r*** is the output. While in GAN, we use ***Minimax*** to do different Gradient Descent steps on G & D. Outputs of D is a value between 0-1 by using log that just denote to the likelihood of being real.
GAN	$\min_G \max_D V(D,G)$	

(3)In VAE, using maximizing likelihood based on approximate inference. While GAN just use backpropagation. In each epoch, firstly train the Discriminator(with

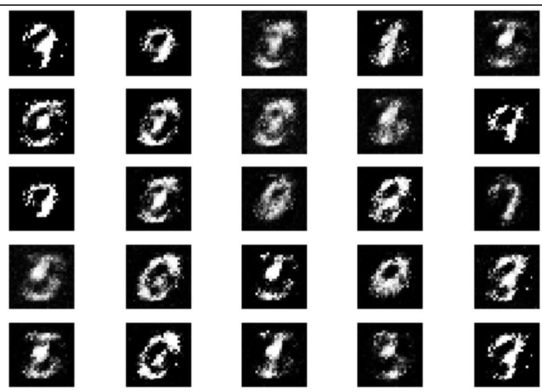
BP in G stopped), then train the Generator(with BP in D stopped). Again and again to fit the distribution of G, until close to the Nash Equilibrium, which is hard to train.

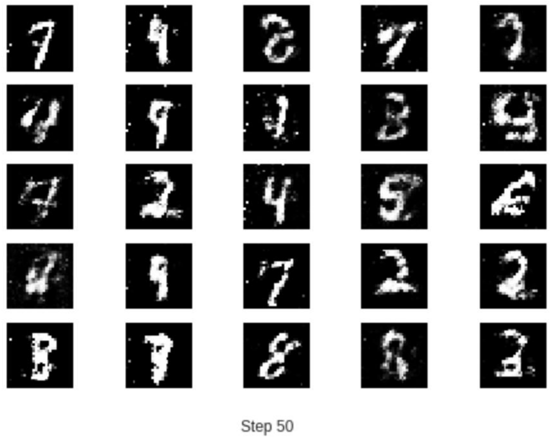
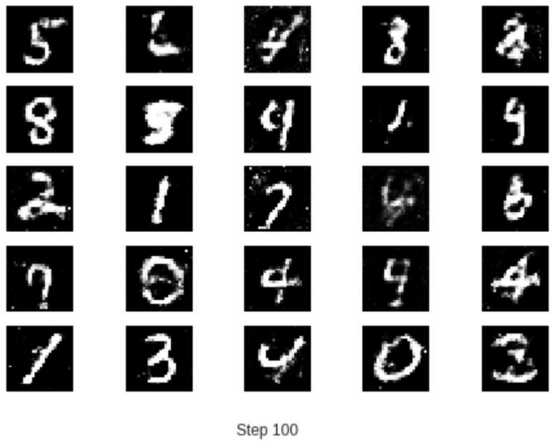
3. **How is the distribution $p_g(x)$ learned by the generator compared to the real data distribution $p(x)$ when the discriminator cannot tell the difference between these two distributions? (15% of CW2)**

Answer:

$p_g(x) = p(x)$, which means the generator's distribution is very similar to the real one. After several steps of training, if ***G*** and ***D*** have enough capacity, they will reach the Nash Equilibrium Point which both cannot improve. The discriminator is unable to differentiate between the two distributions, ***(i.e. D(x) = 0.5)*** .

4. **Show the generated images at 10 epochs, 20 epochs,50 epochs,100 epochs by using the architecture required in Guidance. (15% of CW2)**

Epochs	**Figures**
10	 Step 10
20	 Step 20

<table>
<tr><td>50</td><td>
</td></tr>
<tr><td>100</td><td>
</td></tr>
<tr><td colspan="2">Answer:
At the beginning of 10 epochs, we cannot figure out what the generated images are. During the training epochs, the $\boldsymbol{G(z)}$ will become more and more similar to a real image, which means the distribution of $\boldsymbol{p_g(z)}$ is similar to $\boldsymbol{p_{data}(x)}$. Though still some latent noise could not be generated well.</td></tr>
</table>