Name – Animesh Devendra Chourey
Student ID - 210765551

ECS795P Deep Learning and Computer Vision, 2022

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)**

   To perform image classification using supervised learning, one requires large amount of labelled dataset (images). The algorithms learn to map the images(x) to its associated labels(y) and then the learnt model classifies the new inserted images into one of the pre-defined classes. The biggest challenge while performing supervised learning is the labeling of the images as large dataset is required and it is not always feasible to correctly tag the dataset properly. Annotating the data is labor-intensive and inefficient.

   However, when it comes to unsupervised learning, the data has no label(y). The main objective is to learn some underlying structure of the data(x). These are used to analyze and cluster unlabeled datasets to discover the hidden patterns in them without human intervention. These algorithms still require human intervention to validate the outputs generated. Generative Adversarial Networks (GANs) and Auto Encoders are the most popular algorithms based on unsupervised learning.

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)**

   Model Structure:

   Auto encoder comprises of two networks; encoder and decoder. Encoder's task is to map input *x* (representing the image) to latent space *h*. This latent space (vector) *h* is the compressed representation of the input image. The Decoder's task is to reconstruct the realistic image *r* that the encoder saw before from the latent space *h*. As they learn to represent high dimension data to lower dimension, they are more suitable for dimensionality reduction.

   Generative Adversarial Networks also comprises of two networks; Generator and Discriminator. The generator's task is to generate realistic images and the discriminator's task is to differentiate between the generated image as real or fake. Generator take random noise *z* as input to generate realistic data *G(z)*. This generated data is authenticated by the discriminator *D(G(z))* to qualify as real or not. Thus, they are more suitable for generating new data.

Optimized objective function:

The objective of auto encoder is to minimize the reconstruction errors.

$$l(f(\mathbf{x})) = \frac{1}{2}\sum_k(\widehat{x}_k - x_k)^2$$

In case of GANs, the objective of generator is to minimize the loss and the objective of discriminator is to maximize the loss.

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

Training Procedure:

For auto encoders, both the encoder network and decoder network are updated simultaneously through backpropagation. For most cases, constructing a loss function where one term encourages our model to be sensitive to the inputs (i.e. reconstruction loss) and the second term discourages overfitting. The task is to minimize the difference between the input image and the reconstructed image.

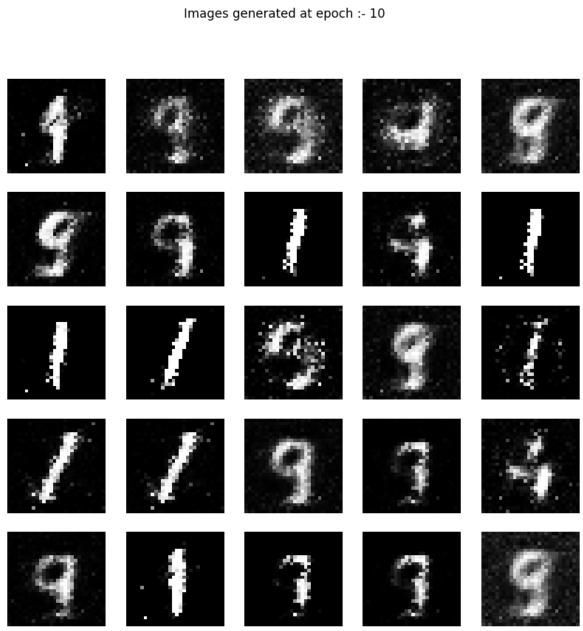
For Generative Adversarial Networks, the training is done parallelly for both the generator and discriminator. Generator is trained to generate realistic data by ascending its stochastic gradient, while the discriminator is trained to distinguish between real and fake data by descending it's stochastic gradient.

3. **How is the distribution $p_g(x)$ learned by the generator compared to the real data distribution $p_{data}(x)$ when the discriminator cannot tell the difference between these two distributions? (10% of CW2)**

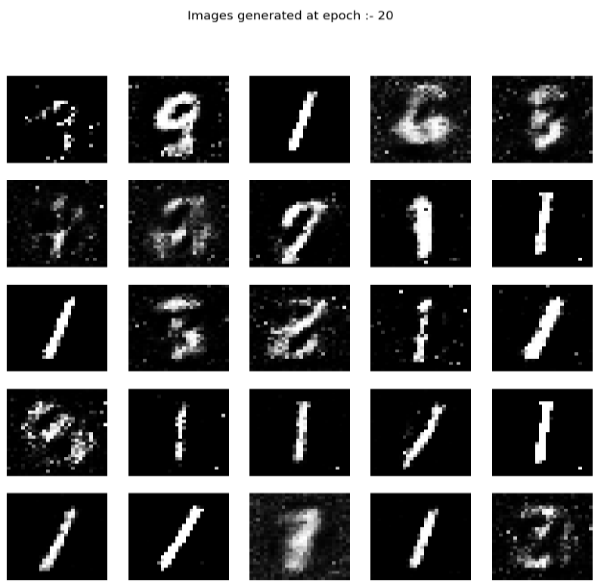
When the discriminator cannot tell the difference between the real data distribution p(x) and the generated data $p_g(x)$, then what it means is that the generator has learnt the underlying structure required to generate the similar realistic images needed i.e. $p_g = p_{data}$. The generative model is now perfectly replicating the data generating process. What this basically means is *the global minimum of the virtual training criterion C(G) is achieved.* At that point C(G) attains the value of -log (4). At this point both the generator and the discriminator potentially cannot improve, and the convergence is guaranteed with the probability of the discriminator making the right decision being $D^*_G(x) = ½$.

4. **Show the generated images at the 10th epoch, the 20th epoch, the 50th epoch, the 100th epoch by using the architecture required in Guideline. (10% of CW2)**

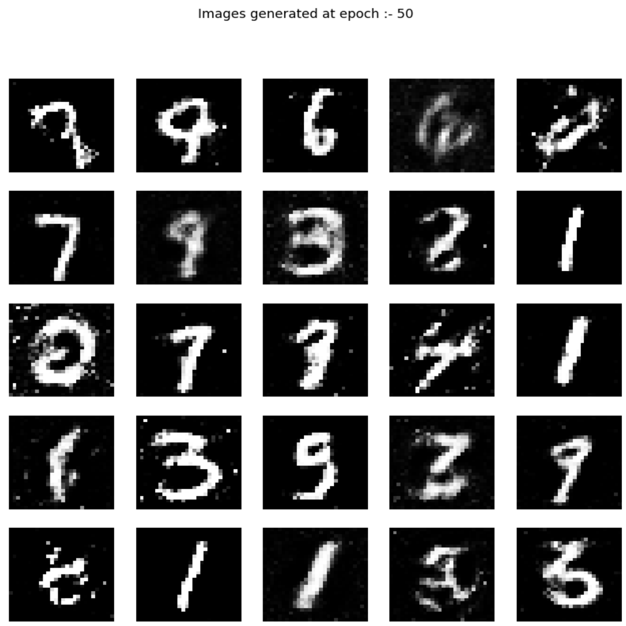
Images generated at 10^{th} epoch –



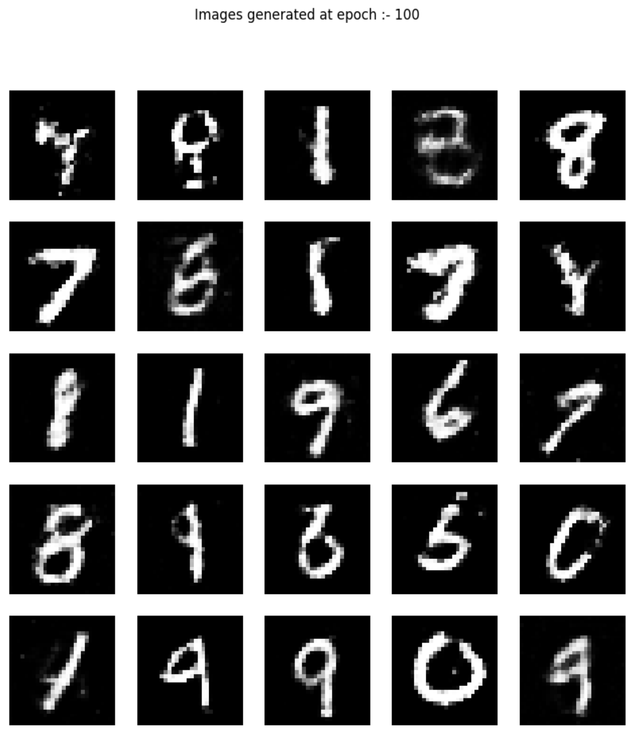
Images generated at 20^{th} epoch –



Images generated at 50^{th} epoch –



Images generated at 100^{th} epoch –



5. Plot the loss curve during training. (10% of CW2)