ECS795P Deep Learning and Computer Vision, 2020

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)**

Ans: In an image classification the images are categorized into different classes. There classes are either given (supervised) or extracted from the training dataset (unsupervised).

In supervised learning, the model requires a huge amount of labeled training data. These labels are the classes the model has to categorize the images in. The labels are either generated manually (annotated by humans) or by using a generative model. The model then learns a decision boundary to classify the images to its respective classes based on the labeled training dataset. Supervised learning is expensive and may suffer from bias.

On the other hand, in unsupervised learning labeled training data is not needed. Unsupervised learning helps in understanding the underlying similarities between the images and then clustering it into different classes. These models are naturally fit for handling large and unstructured image repositories.

Given below is the block diagram of supervised learning and unsupervised learning from [1].

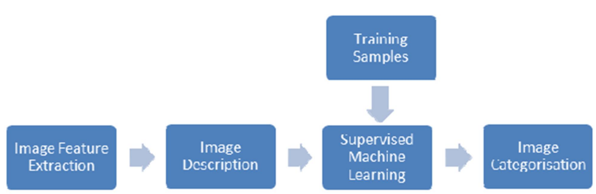


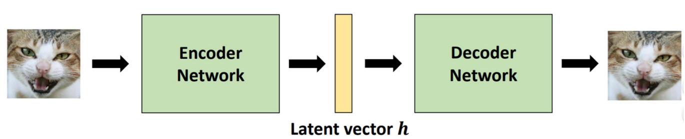
FIGURE 1: The Block diagram of a typical supervised Image categorisation process.

FIGURE 2: The Block diagram of an unsupervised Image categorisation process.

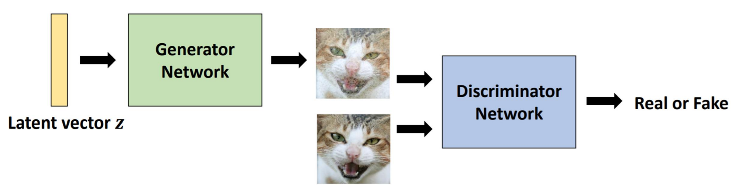
2. **What is the difference between an auto-encoder and a generative adversarial network considering (1) model structure; (2) optimized objective function; (3) training procedure on different components? (10% of CW2)**

Ans: An autoencoder learns to represent the input in an efficient way while also learning how to reconstruct the input from its compressed form. Whereas, a generative adversarial network learns how to generate data which is indistinguishable from the real data. The difference between the two network is given below:

(1) Model Structure: Autoencoders are networks which learns to generate an optout "r" as close as possible to the input image. It consists of an encoder network and a decoder network. The encoder network produces a latent vector "h" which is a compressed representation of the input image. The latent vector is then passed to the decoder network which then tries to reconstruct the original image. The model structure is given below.



A generative adversarial network consists of two network as well but these networks are a generator and a discriminator network which are in a two player minimax game. GANs try to generate samples from a simple distribution. The generator tries to fool the distriminator by generating real looking images and discriminator tries to decide if the generated image is real or fake.



(2) Optimised Objective Function: The objective function for the autoencoder is a square of the pixel wise difference between the predicted image and the groundtruth image as given below.

$$L(x, y; \theta) = -\frac{1}{M}\sum_{i=1}^{M} ||x_i - r_i||^2$$

The objective function for a GAN is given below. The objective of the GAN is a two player minimax game where the discriminator tries to maximise D(x) that is the probability when given the real image and it also tries to minimise D(G(z)) that is probability when given a fake image. The generator tries to maximise the discriminator's chances of making a mistake by maximising D(G(z)).

$$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{x\sim p_{data}}\log \underbrace{D_{\theta_d}(x)}_{\substack{\text{Discriminator output}\\ \text{for real data x}}} + \mathbb{E}_{z\sim p(z)}\log(1-\underbrace{D_{\theta_d}(G_{\theta_g}(z))}_{\substack{\text{Discriminator output for}\\ \text{generated fake data G(z)}}})\right]$$

- Remark: Discriminator outputs a value between 0 and 1 to denote the likelihood of being real

(3) Training procedure on different components: For autoencoders, the error is back propagated and both the networks i.e. encoder and decoder are updated simulatenouly. The training procedure includes minimising the difference in the generated image and the input image. But in the case of GANs the training procedure for G is to maximise the probability of D making a mistake. Both the networks in GANs are not updated together. The discriminator is updated by ascending its stochastic gradient where as the generator is updated by descending its stochastic gradient.

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)**

Ans: When the discriminator cannot tell the difference between $p_g(x)$ and p(x) then it means the generator has learnt the distribution of the real data and p_g=p_{data}. At that point, the probability of the discriminator making a right decision is 50% as $D^*_G(x) = \frac{p_{data(x)}}{p_{data(x)} + p_{g(x)}}$. According to the theorem 1 in [2], the global minimum of the virtual training criterion C(G) is achieved if and only if p_g = p_{data}. At that point, C(G) achieves the value -log 4. The theorem proves that the C(G) is given by C(G) = -log(4) + 2* JSD(p_{data} || p_g) and that the Jensen–Shannon divergence between two distributions is always non-negative and zero only when they are equal. Therefore, the global minimum of C(G) given by C* = - log(4) and this is reached when p_g = p_{data}.

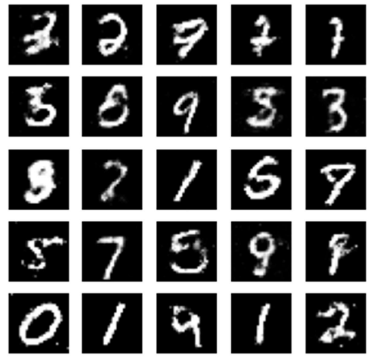
4. **Show the generated images at 10 epochs, 20 epochs,50 epochs,100 epochs by using the architecture required in Guidance. (15% of CW2)**

Ans: The results of the network with dropout layers are given below:

Epoch 10

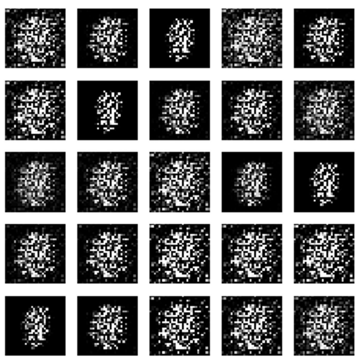
Epoch 20

Epoch 50

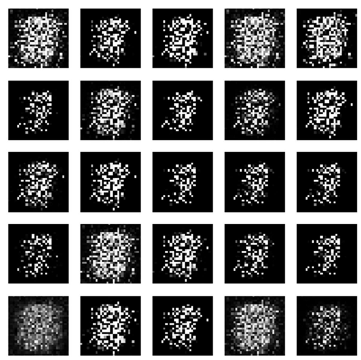


Epoch 100

The results without the dropout layers are given below:



Epoch 10



Epoch 20

Epoch 50

Epoch 100

References:

[1] Olaode, A., Naghdy, G. and Todd, C., 2014. Unsupervised classification of images: A review. International Journal of Image Processing, 8(5), pp.325-342.

[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).