AI3603: Computer Vision, Spring 2025

Indian Institute of Technology Hyderabad
Homework 3, Generative Modelling and Contrastive Learning
20 points. Assigned 23.03.2025, Due **11:59 pm on 30.03.2025**

You must never be fearful of what you are doing when it is right. – Marie Curie

Instructions:

- It is **strongly recommended** that you work on your homework on an *individual* basis. If you have any questions or concerns, feel free to talk to the instructor or the TAs.
- You are free to use Copilot. Please turn in your prompts.
- Use the MNIST dataset.
- Please turn in Python Notebooks with the following notation for the file name: `your-roll-number-`hw3.ipynb.

1. **Distance between PDFs:** In this question you will explore the other "distances" between PDFs discussed in class. To verify the implementation of these distances, use the normalized histogram of the stereo image pair (`left.png, right.png`). Recall from class that a normalized histogram is a valid probability mass function (PMF).
   (a) **Kullback-Leibler (KL) Divergence:** The KL divergence between two PDFs (PMFs) p and q is given by: $D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$. Write a function that accepts two PDFs (PMFs) p, q and outputs the KL divergence between them. Verify that the KL divergence is not symmetric. (1)
   (b) **Jensen Shannon (JS) Divergence:** The definition of JS divergence between two PDFs p and q is given by: $J(p,q) = 0.5 * (D(p||m) + D(q||m))$ where $m = \frac{p+q}{2}$ and $D(p||q)$ is the KL divergence between p and q. Write a function that accepts two PDFs (PMFs) p, q and outputs the JS divergence between them. Verify that the $JS(p,q)$ is symmetric indeed while $D(p||q)$ is not. Again, use the normalized histograms of the stereo image pair. (2)
   (c) **Wasserstein Distance:** The Wasserstein-1 distance between two PDFs r and s is given by: $W_1(r,s) = \inf_{\pi \in \Pi(r,s)} \mathbb{E}_{(x,y)\sim\pi}|x-y|$. The set $\Pi(r,s)$ is composed of all bivariate joint PDFs whose marginals equal r and s. Given a tuple $(p_{(X,Y)}, r_X, s_Y)$ of a joint histogram $p_{(X,Y)}$, and marginals r_X, s_Y, write a function that accepts this tuple and checks if $p_{X,Y} \in \Pi(r,s)$. Verify your function with a positive example and a negative example. (2)

2. **Generative Adversarial Network (GAN) (10)**
   (a) Train a GAN on the MNIST dataset. Use the Deep Convolutional GAN (DCGAN) architecture for the generator.
   (b) Experiment with the number of times the discriminator is trained per generator training step.
   (c) Verify that the model has indeed learned to generate *diverse samples of good quality*.
   (d) Compare the samples generated by the GAN with those generated by VAE and DDPM qualitatively and quantitatively.
   (e) Mode collapse is a problem in GANs when the generator learns to generate only a few categories. Intentionally introduce mode collapse in your GAN and explain how you did it.

3. **Zero-shot Learning** Implement a zero-shot learning model using the CLIP model and test in on the MNIST test dataset. Specifically, use text embeddings of the class label strings (`the handwritten digit zero, the handwritten digit one`, etc.). How do you explain the accuracy of the model? Visualize the embeddings of the "best matched" images using tSNE plots. (5)