AI3603: Computer Vision, Spring 2025

Indian Institute of Technology Hyderabad
Homework 1, Review of Deep Learning
20 points. Assigned 11.02.2025, Due **11:59 pm on 17.02.2025**

The greatest teacher, failure is. – Jedi Master Yoda

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-hw1.ipynb`.
- Do not turn in images. Please use the same names for images in your code as in the database. The TAs will use these images to test your code.

1. **MNIST Image Classification:** Do the following for an artificial neural network (ANN), convolutional neural network (CNN), and a vision transformer (ViT).
   (a) Train a classifier using a 70:10:20 data split for training, validation, and testing respectively. Plot the training and validation loss over training epochs. Assume CE loss. Clearly describe the choice of your optimizer and the hyperparameters. (6)
   (b) Experiment with different architectures and suggest one that is a good tradeoff between accuracy and parameter size. (1)
   (c) Display the tSNE plots for test data for the best and worst architectures from the previous experiment. For the ANN, plot the logit output. For the CNN, plot the features at the bottleneck layer. For the ViT, plot the CLS token. (1)
   (d) For the trained CNN and the ViT models, display the feature maps at each of the convolutional layer and the encoder layer outputs respectively. Based on these feature maps, comment on what you think the models are learning at various layers. (1)
   (e) Explain the relatively poor performance of the ViT compared to the two other models. (1)
2. **MNIST Autoencoder:**
   (a) Train an convolutional autoencoder (CAE) on the MNIST dataset using the MSE loss function. Experiment with different architectures and suggest one that is a good tradeoff between reconstruction loss and parameter size. Verify that the CAE has indeed learnt to encode an image and decode it by displaying the input output image pair. (4)
   (b) Demonstrate the separability of the digit classes in the latent space (at the encoder output) using a tSNE plot. (1)
   (c) Now train a denoising CAE at different noise levels ($\sigma = 1, 5, 10, 15$). In other words, generate noisy images at these different noise levels and use the clean image as the ground truth. Again, use MSE loss. Verify qualitatively (by displaying the noisy and denoised image pair) and quantitatively (using PSNR and SSIM) that the denoising CAE has learnt to denoise its input. (4)
   (d) Comment on the latent representations using tSNE plots (as a function of noise strength). Compare these with the vanilla CAE in the 2 (a). (1)