Only Pytorch can be used as a DL library. There is no gamification this time. This is going to be a long assignment. Please start early!

Real world data like images are generated from several independent and interpretable underlying factors. The goal in disentangled representation learning [2] is to separate out these factors to improve performance on downstream tasks like classification, detection and style transfer.

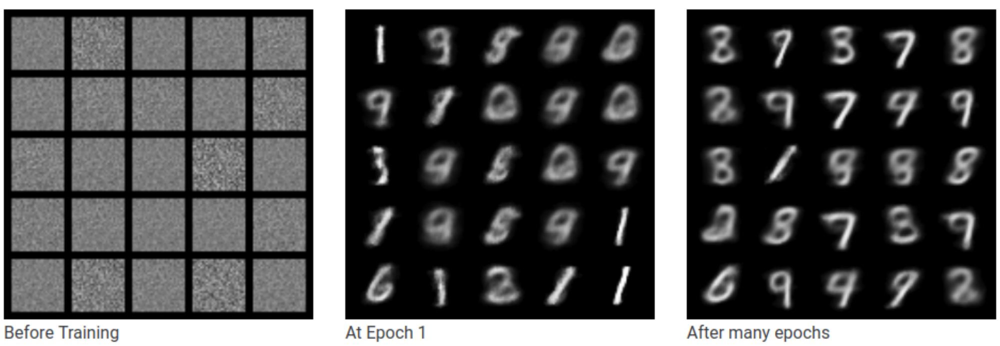
Ques 1. [**70 marks**] In this question, you will implement a simple generative model - The Variational Autoencoder. You have to implement the VAE on the MNIST dataset. You can use the inbuilt dataloader from pytorch. Use the same train/test splits as used in the default setting of the pytorch dataloader.

Deliverables:

(1) Implement the VAE and save the model weights to be used for evaluation. You are free to use a bottleneck of any arbitrary size and any architecture for the encoder and the decoder. [**10 marks**]

(2) Visualize a grid of generated images from the VAE on the MNIST dataset. [**20 marks**]

(3) Include a plot similar to the one shown below, to show the improvement in reconstruction quality of MNIST images. [**10 marks**]



(4) Visualize t-SNE plots of the VAE latent space, color coded by the MNIST identity labels. [**10 marks**]

(5) Train an SVM classifier based on the latent features of the trained VAE. Report the test set accuracy, confusion matrix, precision, recall and F1-score for the trained classifier. [**20 marks**]

Ques 2. [**130 marks**] For this question, you need to implement one of the state-of-the-art variational autoencoder techniques of disentanglement Cycle-consistent VAE on the synthetic caricatures dataset - 2D Sprites. You can download the data from here.

Note: All your results, quantitative / qualitative should be in a format similar to the one depicted in the paper.

Deliverables:

(1) Implement the approach as indicated in the paper. Save the weights of your best model to be used for evaluation. [**10 marks**]

(2) Show style-transfer grids depicting qualitative disentanglement of the specified features, which should be the identity of the caricature for this question. [**25 marks**]

(3) Select a pair of images from the test set. Show linear interpolations in the specified as well the unspecified space. Repeat this for 3 such pairs. Report your observations along with the qualitative results. [**15 marks**]

(4) Train a classifier for the specified partition of the latent space and similarly one for the unspecified one too. Report the accuracies obtained.[**40 marks**]

(5) On the trained model, train a prediction network, with input as the unspecified partition and output as the specified partition, and another one with input and output reversed. Decode the predictions obtained. Qualitatively, depict a batch of original images along with the ones obtained from the prediction network. Mark the misclassifications, if any. Report your analysis. [**40 marks**]

Ques 3. [**80 marks**] For this question, you need to implement a DANN (Domain Adversarial Neural Network, reference paper). There are two main tasks you will have to perform: a toy 2D classification problem, and an image classification problem.

Note: All your results, quantitative / qualitative should be in a format similar to the one depicted in the paper.

Deliverables:

(1) Refer Fig 2 (Page 14) and Algorithm 1 (Page 10) given in the paper. Generate the twinning moon 2D data exactly as mentioned in the paper (Section 5.1.1, Experiments on a Toy Problem, Page 3-14). Show plots of the generated data distribution. Then, implement both the standard NN and the shallow DANN (exactly as described in the paper). Finally, show the "LABEL CLASSIFICATION" and "DOMAIN CLASSIFICATION" plots (as shown in in Fig 2), for both the standard NN as well as the shallow DANN. Also report the accuracies obtained by both the models on the source and target distributions independently. [**30 marks**]

(2) For this part, you have to perform a source and target image classification. The source dataset is MNIST (can be directly used from Pytorch's torchvision dataloader) and the target dataset is MNIST-M (as described in the paper in section 5.2.4 on page 23). The MNIST-M dataset can directly be downloaded from here. You have to implement both the source-only model and the DANN for a comparative analysis. The architecture that you have to use is the same as the one given in Fig 4.a on page 21. Report the label classification accuracies (similar to Table 2 on page 24). Also, plot the t-SNE representations of the data points across source and target distributions for both the source-only model and the DANN (as shown in Fig 5, page 22). [**50 marks**]

References

[1] Higgins, Irina, et al. beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework

[2] Bengio, Yoshua. Deep learning of representations: Looking forward