Special Topics: Deep Learning

Sheet 1 — COMP 499/691

Assignment 3

Due Date: April 28

Submission: Submit a pdf of your written answers and code and output from programming portions. You will also submit a runnable ipynb (single file) including all answers to programming portions of the questions.

Note: The assignments are to be done individually. You may discuss ideas regarding questions with others but should not share answers. List the names of anyone you have extensively discussed a question with at the top of your final submission.

1. (a) We will use the MNIST and SVHN datasets provided by pytorch torchvision modules. A base CNN model architecture to use in this exercise is provided in the starter package:
`https://colab.research.google.com/drive/13OPzfcPI9K9sskoGBetFokC4WnOIcmX3`.
For training in these excercises the suggested batch size is 128, optimizer Adam with learning rate 1e-3 and 20 epochs. However you may adjust these as appropriate or desired. In questions (b)-(d) show the training curves and final test accuracies for both datasets. For (b)-(d) convert the SVHN dataset to grayscale to allow using the same base network, while in (e) use the full color version of the dataset.

   (b) (10 points) Train each of these datasets individually using 20 epochs.

   (c) (10 points) Now train the model in the multi-task setting. You should add for each task a "head" which consists of a 1 hidden layer network: meaning one linear layer + non-linearity+ final project to 10 classes. During training to construct the batch of size 128 sample sample half from SVHN and half from MNIST. You can use a separate dataloader with batchsize 64 for each to accomplish this.

   (d) (10 points) Repeat the multi-task training but now with each dataset being only of size 500. Refer to code from the final project testbed for constructing the appropriate dataloaders.

   (e) (30 points + 10 extra credit) We will use a triplet loss to train models that allows learning a meaningful distance metric between two domains. In this question we will use this to map MNIST and SVHN digits to a joint representation. The rest of this question is described in the starter code. Note: This replaces part (e) and (f) in v1 of the assignment. If you have already successfully done (e) and (f) you do not need to redo the modified version of this question.

Eugene Belilovsky: eugene.belilovsky@concordia.ca

2. (a) We will study the training dynamics of GANs for a simple case. Consider the following function

$$V(d, g) = dg$$

with $g \in \mathcal{R}$ and $d \in \mathcal{R}$.

(b) (10 points) Consider the gradient descent/ascent with learning rate α as the optimization procedure to iteratively minimize $V(d, g)$ w.r.t g and maximize V(d,g) w.r.t d. We will apply the gradient descent/ascent to update g and d simultaneously. Write explicitly the update rule of g and d. Write your answer in the following form:

$$[d_{k+1}, g_{k+1}]^T = A[d_k, g_k]^T$$

where A is a 2x2 matrix; i.e. specify the value of A.

(c) (10 points) The optimization procedure you describe above characterizes a map which has a a stationary point, what are the coordinates (d,g) of the stationary points?

(d) (10 points) Analyze the eigenvalues of A and predict what will happen to d and g as you update them jointly. In other words, predict the behaviour of d_k and g_k as k $\longrightarrow \infty$.

3. (a) (10 points) Show your interpolations from the answer to Lab 9 part 2 (interpolations in the space of BigGAN model). The only requirement here is that the images are different than the example images that come with the lab and those of your classmates. If you have done the lab already you only need to include the images from your lab answer to receive credit.

Eugene Belilovsky: eugene.belilovsky@concordia.ca