AMCS 215
Mathematical Foundations of Machine Learning
Assignment 9, Due Nov 12, 2023

Reading material:

- Section 20.3 from reference 1 (Autoencoders)
- Chapter 15 from Reference 1 (RNNs)
- Chapter 9 from `https://d2l.ai` (this is a computationally-oriented discussion of RNNs)
- RNN tutorial in pytorch at `https://pytorch.org/tutorials/intermediate/char_rnn_classification_tutori`

1. Embedding via feedforward network.
In this exercise, we would like to build an autoencoder for the MNIST training dataset. The encoder consists of two feedforward layers for encoding and two layers for decoding, with the first hidden layer of dimension 100 and the latent dimension (second hidden layer) of dimension 30. We will use the Scaled Exponential Linear Unit nonlinear activation (selu) of all layers except the output layer[1].

- define a network for the autoencoder
- define a reconstruction loss function for training the network
- train the network for a few epochs using appropriate hyperparameters (optimizer, batch size, learning rate, momentum, etc.). Plot the loss history.
- test the quality of the representation by reconstructing a few randomly chosen examples

2. Embedding via CNN network.
In this exercise, we will build another autoencoder neural network for the MNIST data. The encoder part will have the following architecture: Conv2D(16, 3×3, same, selu), MaxPool2D(2x2), Conv2D(32, 3×3, same, selu), MaxPool2D(2×2), Conv2D(64, 3×3, same, selu), MaxPool2D(2×2). The decoder is the mirror image of this, using transposed convolution and without the max pooling layers.

- repeat the steps of Exercise 1 above to build and train the CNN network
- compare the quality of the representation obtained by this CNN autoencoder to those obtained in the previous two exercises (by reconstructing a few randomly chosen examples)

3. Variational autoencoder.
In order to improve the quality of the latent space generated, we can regularize the problem to penalize latent features that deviate too much from a standard normal distribution.

- Eq 6.32 from the textbook (Reference 1) has an expression for the KL-divergence between two general multi-dimensional Gaussians. In the variational autoencoder we discussed, the regularization term was the divergence between Gaussians with diagonal covariance. Show that Eq 6.32 reduces to the regularization expression we used.
- In *pseudo-code*, write a python routine that performs a forward pass through a VAE. No actual code is necessary.

4. Recurrent network.
In this exercise, we will redo the sentiment classification task of Asst 7 with a simple recurrent neural network. You may use a trainable word embedding layer with an embedding dimension of 100 and a hidden RNN layer dimension of 256.

- describe in clean mathematical notation the expressions that are evaluated in a forward pass through the network.
- write the model in `pytorch`
- train the network using a batch size of 64 and appropriate hyperparameters for ADAM
- evaluate the resulting performance

[1] The output layer would have a sigmoid activation in order to produce values between 0.0 and 1.0, which can be mapped to a grayscale image, but we will incorporate this sigmoid in the loss function for numerical stability and for saving computational effort, as we have generally done.

5. 1D CNNs for sequence classification.
Section 15.3.1 in the textbook (Reference 1) describes a "1D" convolutional network for sentiment analysis.

- write down in clean mathematical notation the expressions that are evaluated in a forward pass through the network. Identify the sizes of all intermediate tensors involved for the example of Figure 15.14.
- how many trainable parameters are there in this network?

6. Project idea.
In the remainder of this term, you will work on an ML problem of your choice.

- describe in a *short paragraph* an idea that you would like to work on. It could involve building a model for a problem of interest to you, reproducing the results of a high-quality research paper, implementing a training algorithm, etc.
- you are encouraged to work in a small group (suggested size of 2)
- please submit a title and your one paragraph description separately form the rest of the assignment. There is a separate submission area on blackboard.
- you will develop your idea into a 1-2 page description by Nov 19.
- you will give a 5-minute presentation of your project during the last week of class (Dec 3). Final deliverables are due on Dec 13.