AMCS 215
Mathematical Foundations of Machine Learning
Assignment 8, Due October 31, 2023

Reading material:

- Sections 14-1 and 14.2 from Ref 1 and/or chapter 7, section 8.5 from `https://d2l.ai`
- https://arxiv.org/abs/1603.07285 has a guide to convolution arithmetic that you may find helpful.

Notes:

- The midterm exam will be on **Wed Nov 1, 10:00am–12:00pm.**
- No late submissions will be accepted for this assignment.

1. Topic classification.
In this exercise, we will develop an MLP classifier that can identify the topic of a short news article. The training data consists of a corpus of 120,000 news articles from Reuters, AP, and other international news agencies. Each training article is fixed to have a length of 50 words (either truncated or padded to make it of length 50) and is labeled by the topic that it is most related to: World, Sports, Business, or Sci/Tech.

Your task is to develop an MLP architecture, train it with an appropriate loss function, and evaluate it on a test data set that is also provided. A starter file can be found on Blackboard.

- write a class that defines a neural network architecture consisting of an embedding layer with embedding dimension 20, followed by averaging the word embeddings for every example (this is also known as global average pooling), followed by a dense layer of 64 neurons with ReLU activations, and finally a dense layer of output dimension 4 (we have 4 unique classes in the classification).
- write an appropriate loss function, and train your model with ADAM as the optimizer and a batch size of 1024. Run the optimization for at least 15 epochs. Plot the loss history.
- generate a 4×4 confusion matrix to evaluate the quality of the trained classifier on the test data.
- do you think it would be better to use global max-pooling instead of average pooling in the model? Explain your reasoning.

2. CNN for MNIST data.
One way to improve the quality of the classifier for the MNIST data is to use a convolutional neural network operating directly on 28×28 2D input, instead of the dense layers used in a previous assignment.

- write a class that implements the network architecture below. The network consists of 5 layers with trainable parameters (2 convolutional layers and 3 dense layers), with max-pooling operations on the convolution layers to downscale the data to a smaller spatial dimensions.
- Train the network on the MNIST data for two epochs
- Compare the quality of the resulting classifier with the one you obtained previously.