E0294: Systems for Machine Learning, Jan'25
Indian Institute of Science, Bangalore
Assignment #2

Deadline: 16th February, 11:59 PM

(The numbers in the braces denote the points)

Please write the solution in latex. Put all your codes and the PDF into a zip file and submit. The submitted code should be properly commented and the name of the variables should be intuitive. The code should be implemented by yourself. Marks will be deducted due to codes without comments, with non-intuitive variable names, with high similarity.

Problem 1 (10)

a) If we have an input of size 8 × 8, with 5 channels, and the output is of dimensions 5×5×3, give two possible configurations of the weights and the layer. Can weights of dimension 3× 3 be used to get the output of the aforementioned dimensions? What about weights of dimension 2 × 2? **(3)**

b) Let us say we have an input of size 13 × 15 with 3 channels. We first apply convolution with 7 filters of dimension 4 × 6 (1 stride and no zero padding). Then we apply convolution using 8 filters of dimension 3 × 4 (2 stride and one padding). Finally, we apply max-pooling using a filter of dimension 2 × 2. What is the final dimension of the output? **(7)**

Problem 2 (25)

Consider LeNet-5 network proposed in https://ieeexplore.ieee.org/document/726791

a) Implement LeNet-5 using PyTorch library in Python. Train the network using MNSIT dataset. Plot accuracy vs epochs and loss vs epochs. Consider a number of epochs till the training procedure converges. **(7)**
b) Replace all convolution layers with depth-wise separable convolution. Plot accuracy vs epochs and loss vs epochs. Consider a number of epochs till the training procedure converges. **(7)**
c) What is the total number of multiply operations and addition operation with regular convolution (network obtain in a) and depth wise separable convolution (network obtain in b)? **(4)**
d) Prepare a mathematical function (f) which maps number of operations in a regular convolutional layer to depth-wise separable convolutional layer. Specifically, derive f(x, other parameters) = y, where x is the number of operations in a regular convolution and y is the number operation in depth-wise operation resulting in same output size. **(7)**

Problem 3 (15)

Consider a Multi-head Transformer Algorithm as shown in the figure below. The algorithm consists of one encoder and one decoder (similar to the one discussed in the class). Assume 8 attention heads, embedding size of 256. Assume 16-bit floating point numbers for inputs, outputs and all intermediate operations. Also notice that some of the blocks have been skipped.

a) Implement the transformer algorithm. **(10)**
b) Comment on the total number of bits to be communicated between two consecutive computing blocks with increasing embedding size. Provide proper explanations and visuals. **(5)**