

# Neural Network and Deep Learning

Final Exam

**Total Points: 100**

**Deadline: April 29, 2022**

Delayed submission will be graded as zero.

No deadline extension allowed.

Copying/cheating will be graded as zero.

1. **[Points 30]** Design a word-level language model using the LSTM network and show network architecture in detail. Discuss details of input-output their formats, shapes, network blocks, vocabulary, etc. How would you train your model and how would you test your model? Is it possible to improve your initial language model, how? Present your reasons behind the improvement.
2. **[Points 10]** What is the context bottleneck problem in an encoder-decoder-based machine-translation network? How would you solve this problem? Discuss the details of solutions including the suggested solution model.
3. **[Points 15]** Show differences among LSTM based encoder-decoder and transformer networks? Mention the application scenarios (at least two) where you will be applying LSTM based encoder-decoder and where you apply transformer network.
4. **[Points 45]** Suppose you are classifying cars, buses, and trucks images of shape 256x256, using a convolutional neural network and each image has three channels. Your convolutional neural network contains two convolutional layers, one fully connected layer, and then a SoftMax classifier. Your fully connected layer has 96 neurons. You are designing such convolutional layers so that you can get better performances. Your filter size for these layers can be any size between 2x2 to 8x8. You also apply the max-pooling layer after the second and third convolutional layer of any size between 2x2 to 7x7 that fits with your architecture. You are free to choose any number of filters (between 16 to 128) for each convolutional layer as well.
  - a. Draw this convolutional neural network diagram with details parameters. **Note: Show the configuration or dimensions of each layer as you see in the lecture slides. [Points 10]**
  - b. Compute the number of parameters in each layer and showcase the total number of parameters of your network. Compute the number of parameters, for both with and without bias in each layer. **[Points 10]**
  - c. Is it possible to reduce the total number of parameters of your previous CNN (4a) by adding additional filters (any size of 2x2 to 8x8)? Please show your detailed calculation to showcase the lower number of parameters of such a CNN network. Also, draw the overall diagram with each layer dimension of your reduced parameter-based CNN network. Note that you need to add additional filters to reduce the number of parameters of your initial CNN model (4a). You cannot increase the size of the filter and/or change (decrease) the number of filters to reduce the total number of parameters of your previous CNN(4a) model. **[Points 25]**

### Submission Instructions:

**Important.** You can submit your text question answer in a PDF/Doc. Please don't copy your homework. **In case of copying from others will be directly graded as zero without showing any reasons.** Based on the questions, everyone may have a different architecture with different filters size, numbers of filters, etc. Therefore, you would end up with different configurations and parameters, and operations. **Delayed submission will be graded zero automatically.**