National University of Computer and Emerging Sciences
Lahore Campus

Machine Learning For Robotics (CS 4090)

Sessional-II Practice

Date: Due Before Mid Exam

Course Instructor(s)

Dr. Mirza Mubasher Baig

Total Time (Hrs):

Total Marks: **3%**

Total Questions: **3**

Instructions: Attempt all questions

Question 1: Multilayer Perceptron Learning

i) A neuron receives three inputs x = [1, 2, 3] with weights w = [0.2, −0.4, 0.6]and bias b = 0.5. Compute the neuron output if the activation function is sigmoid.

ii) Given that :
Input x = [1, 2], target output t =1, weights w = [0.3, 0.2], bias b = 0.1, activation is sigmoid

a. Create a computation graph of this neuron and update the weights assuming cross-entropy loss is the loss function.
b. Perform the forward pass and propagate gradient back to update weights using a learning rate of 0.1. Show the updated weights

iii) Design a formula to compute trainable parameters of a feed-forward neural with M inputs, N neurons in a single hidden layer and K output neurons

iv) We know that a feed-forward neural network can be represented as a series of matrix into a vector multiplications, for example the following expression represents a two layer neural network

$$f = W_2 \max(0, W_1 x + b_1) + b_2$$

a. What are dimensions of $\boldsymbol{W_1}$, $\boldsymbol{W_2}$, $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$ is input vector *x* is a 1024 dimensional and the network has 100 neurons in the hidden layer and 10 neurons in the output layer
b. Prove that the above network is equivalent to a single layer network (i.e. no hidden layers) with 10 neurons if the activation function **max** is replaced with linear activation

c. For the input x = $[1,0]^T$, compute the network output if the following weights and Biases are used

$$W_1 = \begin{bmatrix} 0.1 & -0.2 \\ 0.4 & 0.5 \end{bmatrix}, \quad W_2 = \begin{bmatrix} 0.3 & -0.1 \end{bmatrix}$$

$$b_1 = [0.1, -0.1]^T \quad , \quad b_2 = 0.05$$

Question 2: CNN

i) Compute the output dimensions of the volume obtained after passing a volume of dimensions 128 x 128 x3 from a convolution layer having 32 filters each of size 3x3 followed by passing it from a 2x2 max pooling layer with a stride of 2. Assume that zero padding is used in the convolution layer

ii) A CNN architecture uses the following layers to process a 32x32 image:

Conv1: 3×3 kernel, stride = 1, padding = 1
MaxPool1: 2×2, stride = 2, padding = 0
Conv2: 5×5 kernel, stride = 1, padding = 2
MaxPool2: 2×2, stride = 2, padding = 0
Fully Connected Layer with 10 output neurons

a. Compute the size of the receptive field of a single neuron in the conv1 layer
b. Compute the size of the receptive field of a single neuron in the conv2 layer
c. Compute the size of the receptive field of a single neuron in the output layer

iii) Compare the size of your networks (# of weights and memory needed) used in the assignment comparing the performance of LeNet and simple three layer feed0forward network for the MNIST dataset.
{Hint: Mid might contain a question about this assignment}

Question 3: RNN

i) Given the weights and state of a recurrent neural network and time ***t-1***

$$W_{xh} = \begin{bmatrix} 0.1 & 0.3 \\ 0.2 & 0.4 \end{bmatrix}, \quad W_{hh} = \begin{bmatrix} 0.5 & 0.1 \\ 0.3 & 0.2 \end{bmatrix}$$

$$h_{t-1} = [0.2, 0.1], \quad b_h = [0.1, 0.1]$$

Compute the updated state for this network if it uses **tanh** as activation and $\mathbf{x_t}$ = [1, 2] is the input at time instance *t*.

ii) Device a formula to compute number of trainable parameters for a recurrent neural network having a **D**-dimensional vector as input, a **H**-dimensional state and a **K**-dimensional output

iii) While training a RNN one might have to deal with a vanishing gradient problem, Explain what do we mean by vanishing gradient and describe why this might be a common case in case of recurrent neural networks.

iv) Give one example problems suitable for each for the following RNN structures

a. Many to Many
b. One to Many
c. Many to One