CS551: Introduction to Deep Learning

End Semester, Spring 2017

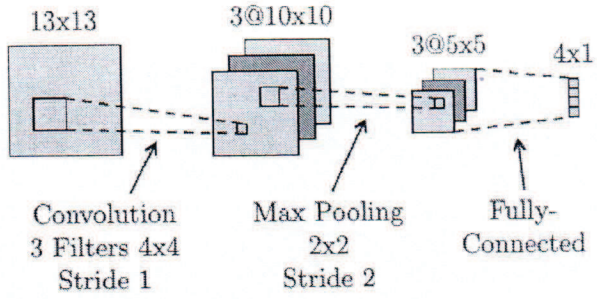
IIT Patna

Attempt all questions. Do not write anything on the question paper.

Time: 3 Hrs | Full marks: 50

1. What is early stopping? Show that it acts as regularizer. (3+5)

2. Below is a diagram of a small convolutional neural network that converts a 13x13 image into 4 output values. The network has the following layers/operations from input to output: convolution with 3 filters, max pooling, ReLu, and finally a fully-connected layer. For this network we will not be using any bias/offset parameters (b). Please answer the following questions about this network.



   (a) How many weights in the convolutional layer do we need to learn?

   (b) How many ReLu operations are performed on the forward pass?

   (c) How many weights do we need to learn for the entire network?

   (d) State True or false: A fully-connected neural network with the same size layers as the above network ($13 \times 13 \rightarrow 3 \times 10 \times 10 \rightarrow 3 \times 5 \times 5 \rightarrow 4 \times 1$) can represent any classifier that the above convolutional network can represent.

   (e) What is the disadvantage of a fully-connected neural network compared to a convolutional neural network with the same size layers? (2+2+2+1+1)

3. Explain backpropagation methodology in the context of recurrent neural network. What are the issues in computation of the gradients here? (6+2)

4. Given a set of m points $\{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(m)}\}$ in $\mathbb{R}^n$ and we want represent these points in k dimension where $k < n$. Propose a suitable methodology for it. (10)

5. Let's consider a simple 3 state Markovian Decision Process with two actions - L and R. The transition probabilities are:

Action L	Outcomes		
	State 1	State 2	State 3
In state 1	0	1/4	3/4
In state 2	3/4	0	1/4
In state 3	1/4	3/4	0

Action R	Outcomes		
	State 1	State 2	State 3
In state 1	0	3/4	1/4
In state 2	1/4	0	3/4
In state 3	3/4	1/4	0

The reward in state 2 is 1 and 0 elsewhere. The discount factor is 0.5.

(a) Show the utility (value) estimates for the first three iterations of the value iteration algorithm. To make things simpler, assume that you keep a copy of the utility estimates from the previous iteration and use those in the new iteration. You may start with $U(1) = U(2) = U(3) = 0$

(b) Using the approximately optimal values of the utility $U(1) = U(3) = 0.5$ and $U(2) = 1.25$, compute the best action for state 1 from these values numerically. Explain how you did it. (7+3)

6. Write a short note on batch normalization. (6)