CS551: Introduction to Deep Learning

End Semester, Spring 2019

IIT Patna

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

Time: 3 Hrs | Full marks: 50

1. (a) What is momentum and how does it help in improving convergence? (b) Describe briefly the steps for stochastic gradient descent with momentum. (3+3)

2. (a) Prove that adding noise to samples can act as regularizer. (b) What is dropout? How is it being used in prediction of a test sample? (5+3)

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

4. Consider the simple n-state MDP as shown in Figure 1. Starting from state s_1, the agent can move to the right (a_0) or left (a_1) from any state s_i. Actions are deterministic and always succeed. Rewards are given upon taking an action from the state. Taking any action from the goal state G earns a reward of $r = 1$ and the agent stays in state G. Otherwise, each move has zero reward ($r = 0$). Assume a discount factor $\gamma < 1$.

   (a) The optimal action from any state s_i is taking a_0 (right) until the agent reaches the goal state G. Find the optimal value function for all states s_i and the goal state G.

   (b) Does the optimal policy depend on the value of the discount factor γ? Explain your answer.

   (c) Consider adding a constant c to all rewards (i.e. taking any action from states s_i has reward c and any action from the goal state G has reward $1 + c$). Find the new optimal value function for all states s_i and the goal state G. Does adding a constant reward c change the optimal policy? Explain your answer.

   (d) After adding a constant c to all rewards now consider scaling all the rewards by a constant a $(-\infty, \infty)$ (i.e. $r_{new} = a(c + r_{old})$). Find the new optimal value function for all states s_i and the goal state G. Does that change the optimal policy? Explain your answer, If yes, give an example of a and c that changes the optimal policy. (3+1+2+3)

5. (a) Determine the number of computations and the output size for the given inception module in Figure 2. The input size is 28×28 and it has depth of 256. In the figure '$A \times A$ Conv, B' denotes B number of filters of size $A \times A$. (b) Suppose three blocks of '1×1 Conv, 64' are removed then determine number of computations and the output size. [Assume that it uses appropriate number of zero padding to keep the output size 28×28.] (4+4)

6. Suppose the neural network as shown in Figure 3 is initialized with the following weights: $w_{h_1,x_0} = -0.4$, $w_{h_1,x_1} = 0.2$, $w_{h_1,x_2} = 0.1$, $w_{h_2,x_0} = -0.2$, $w_{h_2,x_1} = 0.4$, $w_{h_2,x_2} = -0.1$, $w_{o_1,h_0} = 0.1$, $w_{o_1,h_1} = -0.2$, $w_{o_1,h_2} = 0.1$, $w_{o_2,h_0} = 0.4$, $w_{o_2,h_1} = -0.1$, $w_{o_2,h_2} = 0.1$. Assume that both the hidden and the output layers use sigmoid function. (a) What are the activation of the hidden and output units after forward-propagation of the input $\mathrm{x}^1 = (1, 0)$? (b) Suppose x^1 has target output $t = (0.9, 0.1)$. What will be the new value for w_{h_1,x_1} after single iteration of back-propagation on this example? Use learning rate as 0.1 and momentum as 0.9. Assume MSE as the loss function and initial velocity to be 0. (4+4)

7. Suppose that X is a discrete random variable with the following probability mass function given in the table: where $0 \leq \theta \leq 1$ is a parameter. The following 10 independent observations were taken from such a distribution: $(3, 0, 2, 1, 3, 2, 1, 0, 2, 1)$. What is the maximum likelihood estimate of θ. (3)

X	0	1	2	3
$P(X)$	$\frac{2\theta}{3}$	$\frac{\theta}{3}$	$\frac{2(1-\theta)}{3}$	$\frac{(1-\theta)}{3}$

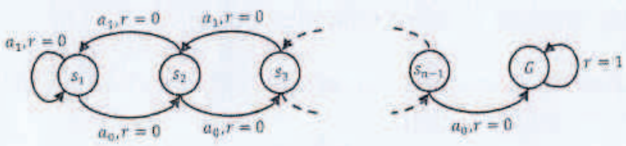

Figure 1

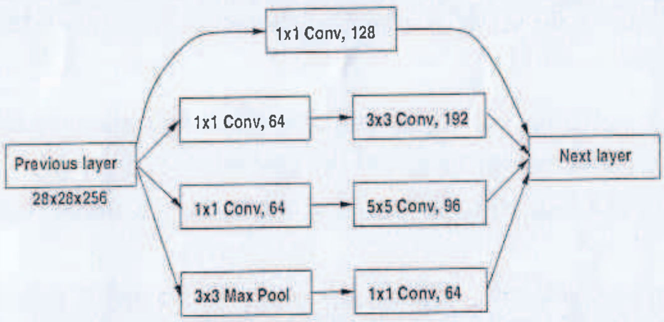

Figure 2

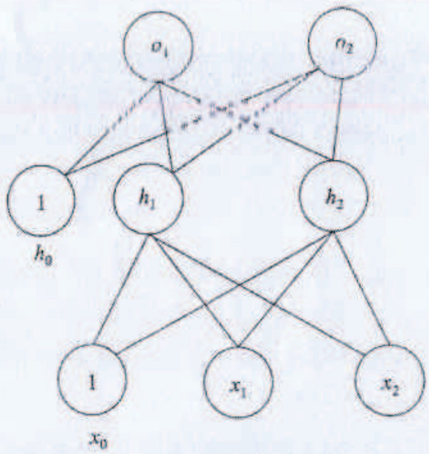

Figure 3