CS551: Introduction to Deep Learning

End Semester, Spring 2022

IIT Patna

Attempt all questions.

Time: 3 Hrs

Total marks: 50

1. (a) Describe continuous-bag-of-words and skip-gram models in the context of word2vec.
(b) Describe neural network architectures that are used for those two model.
(c) Word2Vec represents a family of embedding algorithms that are commonly used in a variety of contexts. Suppose in a recommender system for online shopping, we have information about co-purchase records for items $x_1, x_2, \ldots, x_n$ (for example, item x_i is commonly bought together with item x_j). Explain how you would use ideas similar to Word2Vec to recommend similar items to users who have shown interest in any one of the items. (3+3+2)

2. (a) Consider a neural network with three layers including an input layer. The first (input) layer has four inputs x_1, x_2, x_3, and x_4. The second layer has six hidden units corresponding to all pairwise multiplications. The output node o simply adds the values in the six hidden units. Let L be the loss at the output node. Suppose that you know $\frac{\partial L}{\partial o} = 2$ and $x_1 = 1$, $x_2 = 2$, $x_3 = 3$, and $x_4 = 4$. Compute $\frac{\partial L}{\partial x_i}$ for each i.
(b) Describe back-propagation through time algorithm for recurrent neural network. (4+6)

3. (a) What is early-stopping approach?
(b) Prove that it can act as L_2 regularizer.
(c) Prove that adding noise to samples can act as regularizer. (2+4+3)

4. (a) How does splitting a dataset into train, validation and test sets help identify overfitting?
(b) You are designing a neural network model to detect driver fatigue in cars. It is crucial that that your NN detects fatigue, to prevent any accidents. Which of the following is the most appropriate evaluation metric: Accuracy, Precision, Recall, Loss Value. Explain your choice.
(c) Explain why dropout in a neural network acts as a regularizer.
(d) How is dropout being used in prediction of a test sample? (2+2+2+1)

5. Consider the MDP given in Figure 1 Assume the discount factor $\gamma = 0.9$. The r-values are rewards, while the numbers next to arrows are probabilities of outcomes. Note that only state S_1 has two actions. The other states have only one action for each state.
(a) Write down the numerical value of $V(S_1)$ after the first and the second iterations of Value Iteration. Initial value function: $V^0(S_0) = 0$, $V^0(S_1) = 0$,$V^0(S_2) = 0$,$V^0(S_3) = 0$.
(b) Find the optimal value for state S_1 that is $V^*(S_1)$. [Hints: Try to avoid detailed iterative method!].
(c) State true/false: Any MDP converges after the 1st value iteration for $\gamma = 1$.
(d) State true/false: Any MDP converges after the 1st value iteration for a discount factor γ (3+3+1+1)

6. (a) A convolutional neural network has 4 consecutive layers as follows: 3x3 conv (stride 2) - 2x2 Pool - 3x3 conv (stride 2) - 2x2 Pool. How large is the support (the set of image pixels which activate) of a neuron in the 4th non-image layer of this network?
(b) Determine the number of computations (multiplications only) 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 $A \times A$. Assume that it uses appropriate number of zero padding to keep the output size 28×28.
(c) Suppose three blocks of '1×1 Conv, 64' are removed then determine number of computations and the output size. (2+3+3)

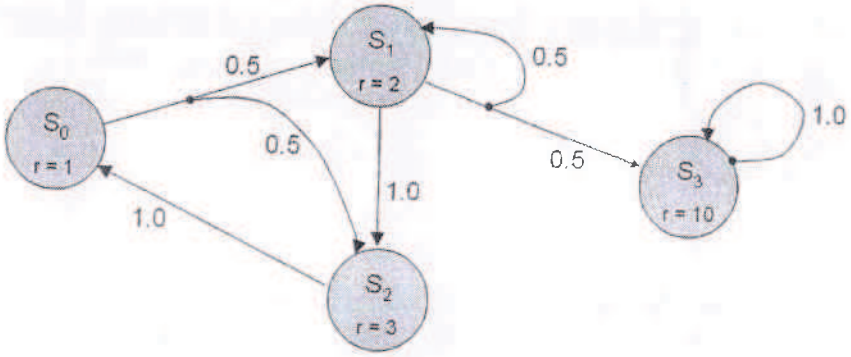

Figure 1

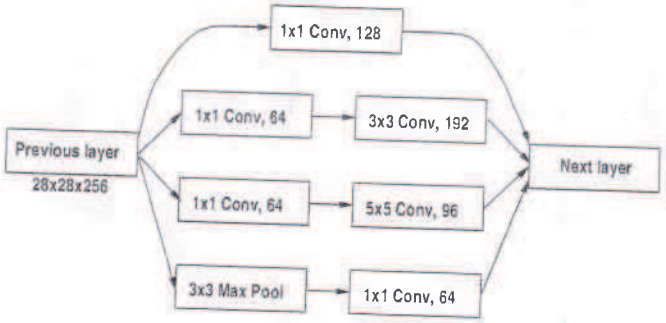

Figure 2