Advanced Machine Learning 2023

Final Exam

Chennai Mathematical Institute

04 December 2023, 09:30–12:30 (3 hours) Marks: 100

Instructions:

- This question paper consists of two sides, with a total of 14 questions worth 100 marks. All questions must be answered; there are no optional ones.
- Before you begin writing your answers, carefully review all the questions.
- Make sure to write you name and roll number on the main booklet *and* on all the additional sheets that you use.
- If you need to use the restroom, ask for a toilet pass from the invigilators.
- Attendance will be taken during the exam. An invigilator will approach you with the attendance sheet. Please sign next to your name.
- **WARNING:** CMI's academic policy regarding cheating applies to this exam. Any form of copying or use of electronic devices is strictly prohibited. Violation will result in an automatic F grade, immediate exclusion from placement activities, and potential expulsion without exception.

1. Consider a neural network with 1 input, 1 output and k hidden layers, each with d neurons with ReLU activation. Calculate the total number parameters of this network. *(5 marks)*
2. Describe the activation functions Sigmoid, ReLU and Leaky-ReLU with leak α. Plot the curves of their gradients (approx values) in the range $[-10, 10]$. From these values what can you say about their effects on the gradient descent algorithm in a neural network. *(5 marks)*
3. Consider a fully-connected neural network for classifying images (say MNIST). What happens if we initialize all parameters of a neural network to zero at the start of training? What happens if we set them all to one instead? *(5 marks)*
4. Recall that we studied 2D convolutions in class. Can you describe a 1D convolution of kernel size 3 with 5-channel input. Write down the mathematical equation for this convolution. *(5 marks)*
5. Suppose you have a simple neural network, that given a number x as input, computes a number y as described below. Write down the equivalent residual network for this neural network. *(5 marks)*

$$x \longrightarrow z_1 = h_1(x) \longrightarrow z_2 = h_2(z_1) \longrightarrow y = h_3(z_2)$$

6. Consider the collection of triplets of intervals of the real-line $\{[a, b] \cup [c, d] \cup [e, f]\}$. What is the VC dimension of this collection? *(5 marks)*

7. Describe the regularizer term in the loss function for Variational Autoencoder. Explain how it affects the latent space, e.g. when decoding a random point or interpolating between two points corresponding to two different images in the training set. *(5 marks)*

8. Describe how can compute the Action-value function from the state-value function of an MDP, and vice-versa. *(5 marks)*

9. What is "Experience Replay" in Deep Reinforce Learning and why is it necessary? What happens if we do not incorporate Experience Replay in the training process? *(5 marks)*

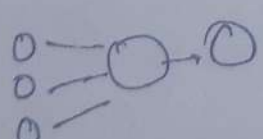
10. What is the key difference between Monte-Carlo methods and Temporal Difference Learning methods in Reinforcement Learning? *(5 marks)*

11. Describe a small neural network with ReLU activations that computes (1) the average of four numbers and (2) the maximum of four numbers? Can you use this to implement 2x2 MaxPooling and 2x2 average pooling? *(10 marks)*

12. Describe an LSTM unit in full details including the functions of it's various gates, inputs and outputs. *(10 marks)*

13. Suppose you are training a GAN to generate images of cats and dogs. After training for sometime you take a look at the output of the Generator and observe that it is producing very good quality images of both cats and dogs. However, the discriminator still is able to successfully identify fake images at least 50% of the time. Can you give an explanation for this scenario. *(10 marks)*

14. Suppose that you are developing a deep learning model for a client (e.g. an airline) that is given a short social media posts (like a tweet) as input and you must classify it is positive, negative or irrelevant to the client. The input will be short english sentence (at most 50 words) and the output must be one {positive, negative, irrelevant}. You will be given a training-set of 10000 labeled sentences. Please describe a complete deep learning model for this task, using 1D Convolutions. Also describe how you will train and test this network using the given data. *(20 marks)*



$\frac{1}{1+e^{-z}}$ Relu(avg.).

$1 - D(x)$