DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Indian Institute of Technology Jodhpur

Pattern Analysis and Machine Intelligence

March 28, 2022

Time: 4 hours — *Ph.D. Comprehensive Examination (Mar 2022)* — Maximum Marks: 100

PART-A

1. (2 points) True/False: The matrices $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$ are inverses of one another. Justify your answer.

2. (4 points) Design an objective function defined on $\mathbb{R}^2$ where gradient descent will be exceedingly slow. Also justify your answer.

3. (4 points) True/False: "We cannot remove the bias parameter from a convolution layer or a fully-connected layer before applying batch norm." Justify your answer.

4. (2 points) True/False: "Autoencoder with linear activation function and single hidden layer is same as PCA." Justify your answer.

5. (4 points) True/False: "In the basic formulation of GAN, both the generator and the discriminator perform a binary logistic regression with the cross-entropy loss." Justify your answer.

6. (2 points) True/False: "Interpretability and explainability of a neural network are the same." Justify your answer.

7. (5 points) Discuss, along with the necessary mathematical steps wherever needed, how you would need to modify the AdaGrad algorithm to achieve a less aggressive decay in learning rate.

8. (6 points) Let $f(x) = x^2(1 + sin(x))$. Answer the following (2 marks each):

    1. Plot the function $f(x)$.
    2. Draw the computational graph of $f(x)$.
    3. Calculate the value of $f'(0)$.

9. (10 points) The problem of imbalanced domain learning has become very important with deep learning algorithms, more importantly with multilable-multiclass problems. You have been given a multi-label and multi-class classification problem where there are n-classes and m-labels (hint: to visualize multilable-multiclass (ML-MC) problem, consider face recognition application where class labels are identities and labels are gender, age, race, etc).

    Design a deep learning classifier for ML-MC problem with n-classes and m-labels with imbalanced domain learning. Clearly show your architecture details (including any assumptions) and argue why your approach will be able to handle imbalanced data? Also, show how back-propagation would work - both forward pass and backward pass as well as weight updates. Draw a clear block diagram to highlight your pipeline.

10. (6 points) Distributed deep learning with sequential data is another very important area which works with model parallelism and data parallelism concepts.

1. Explain model parallelism and data parallelism with respect to how gradients can be calculated and weights can be updated.
2. How distributed DL is different compared to Federated deep learning, and how gradient calculation and weight updates are done in Fed DL?

11. (5 points) Run two passes of learning for the following autoencoder with g(x) = x/4, $\eta = 0.5$.

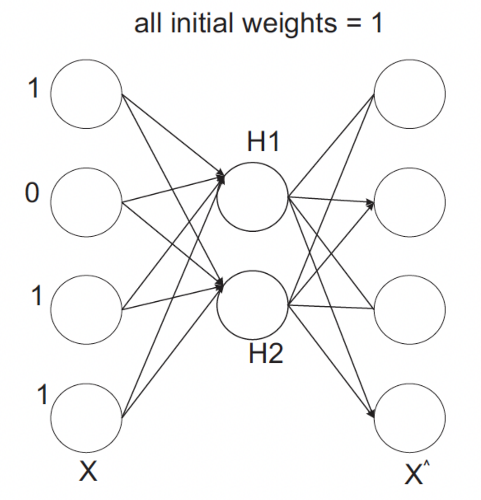

Figure 1: Figure for autoencoder numerical.

PART-B1

Paper: "Masked Autoencoders Are Scalable Vision Learners"

1. (5 points) "Our MAE pre-training can be implemented efficiently, and importantly, does not require any specialized sparse operations."
    1. What are the specialized sparse operations generally required in other models?
    2. In your opinion, why they are not required in this model?
2. (5 points) Can this concept of masking be applied to other architectures like CNN? Explain your answer.
3. (8 points) What kind of masking techniques have been used in the paper? Explain the choice of the technique. Summarize other kinds of masking techniques available in the literature with their advantages and disadvantages. Compare and contrast at least 4 of them wrt applications of semantic segmentation, object recognition, face recognition, and x-ray classification.
4. (7 points) In the MAE paper, a self-supervised pre-training has been performed. Explain how the self-supervised pre-training is done with MAE and show an alternate pre-training mechanism without self-supervision.

PART-B2

Paper: "Free Lunch for Few-shot Learning: Distribution Calibration"

1. (5 points) "The α in Eq.6 determines the degree of dispersion of features sampled from the calibrated distribution." Elaborate.

2. (5 points) In Section-3.2.1, the Tukey's Ladder of Powers Transformation is not applied while calculating the statistics of the base classes (Eq.1 and Eq.2). Briefly discuss why?

3. (5 points) While calculating the calibrated mean in Eq.6, the numerator is given by $\sum_{i \in \mathbb{S}_N} \mu_i + \tilde{\mathbf{x}}$. Which of the following is the correct form of the numerator and why (justify your answer): $\sum_{i \in \mathbb{S}_N} (\mu_i + \tilde{\mathbf{x}})$ or $(\sum_{i \in \mathbb{S}_N} \mu_i) + \tilde{\mathbf{x}}$

4. (5 points) What can we infer from the empirical results in Table-4 other than what is mentioned in the paper? Briefly discuss.

5. (5 points) The proposed algorithm transfers the statistics from the base classes to the novel classes. Briefly discuss one experimental set-up (other than those discussed in the paper) that would allow a more thorough testing of this assumption, and also justify your answer.