# Deep Learning
Mid-I
Important Questions

1. Explain the need for nonlinear activation functions in neural networks. How the ReLU activation function solves the XOR problem?
2. What is the choice of output layer units in the task of multiclass classification and why?
3. List out the activation functions used in deep neural networks with suitable equations and determine their derivatives.
4. What is the role played by cost/loss functions in training deep neural networks? Explain which cost functions are used in the task of linear regression and why?
5. Define the cost functions required in binary classification and multiclass classification tasks. How are their derivatives computed with respect to the parameters of the model?
6. Explain the following terms and the relationships existing among them (i) Capacity of the model, (ii) Complexity of the data, (iii) Underfitting, and (iv) Overfitting.
7. Consider 4 samples in which 2 of them belong to one category C1 and the remaining 2 belong to another category C2. Determine the solution region in terms of a single decision rule which separates these two classes assuming that they are linearly separable. Also, explain how the gradient descent approach finds the solution to this classification problem.
8. Suppose that while using an adaptive learning rate optimizer, the learning rate is aggressively and monotonically decreasing. What modifications are necessary to the adaptive learning rate to tackle such scenarios?
9. What is meant by the regularization of a deep learning model? What different strategies are used for regularization?
10. Explain in detail the process of ridge regularization and how it reduces the generalization error.
11. Explain in detail the process of lasso regularization and how it acts as a feature selector.
12. Define the term momentum and explain how it accelerates the SGD.
13. Explain and distinguish between Vanilla gradient descent, Stochastic gradient descent, and mini-batch gradient descent optimization algorithms. What are the common challenges they face?
14. Given the network below with corresponding initial weights, biases, training inputs, and the target values, determine the total error that we get after one forward pass. Assume that all the neurons are of a sigmoid type and the task is a binary classification task with a binary cross-entropy loss function.
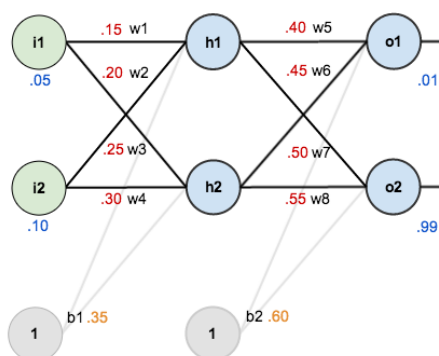
15. Distinguish Adagrad optimization from the basic gradient descent optimization.
16. What is the chain rule of calculus and how is it useful in a backpropagation algorithm?
17. Compute the Hessian of $f(x, y) = x^3 - 2xy - y^6$ at the point (1, 2).
18. Almost all neural networks are trained using derivatives. Why do we use a derivative for deciding how to change the weights?
19. What is the difference between statistical optimization and training of machine/deep learning models?