1. CDT402-Deep Learning for Data Science

2. **Module 2**
3.

1. List any three methods to prevent overfitting in neural networks.
2. With the equation explain the drawback of momentum-based gradient descent. How it is rectified in Nesterov accelerated Gradient Descent?
3. Differentiate gradient descent with and without momentum. Give equations for weight updation in GD with and without momentum. Illustrate plateaus, saddle points and slowly varying gradients.
4. Describe the effect in bias and variance when a neural network is modified with more number of hidden units followed with dropout regularization.
5. Explain how L2 regularization improves the performance of deep feed forward neural networks.
6. Initializing the weights of a neural network with very small or large random numbers is not advisable. Justify.
7. Suppose you are working on developing a deep learning model to predict customer churn for a telecommunications company. The dataset you are working with contains millions of records and hundreds of features.
   (i) How would you choose between using batch, stochastic, or mini-batch gradient descent for training your model?
   (ii) What factors would you consider in making this decision?
   (iii) How might your choice of optimization algorithm impact the performance of your model in terms of both accuracy and training time?
8. What is a Perceptron? Explain the disadvantages of a Perceptron.
9. Explain how L1 and L2 regularization techniques help in preventing overfitting in neural networks.
10. Differentiate between Gradient Descent (GD) and Stochastic Gradient
11. What is Gradient Descent? Explain its working principle. How would you modify Gradient Descent to ensure faster convergence when dealing with highly non-convex loss functions? Discuss with suitable equations and diagrams.
12. Descent (SGD) in terms of their working principles.
13. Write weight updating rule in gradient descent when the error function is a) mean squared error b) cross entropy
14. How does dataset augmentation help when working with limited training data?
15. How does hyperparameter tuning influence the training process and performance of a neural network?
16. How does the choice of learning rate affect the convergence of a neural network?
17. Distinguish between AdaGrad, RMSProp and Adam in the context of deep neural network training?
18. Explain the concept of early stopping in neural network training.
19. A given function is of the form $J(\Theta)= \Theta^3 -2\Theta +3$. What is the weight update rule for the gradient descent optimization at step t+1? Consider $\alpha=0.2$ to be the learning rate.
20. A neural network was trained to classify images into different categories. After training the classifier, observed a large gap between the training accuracy (100%) and the test accuracy

(40%). What could be the problem with the training of such a classifier. Discuss any four methods to reduce this gap.

21. You have a deep model with millions of parameters, and it starts memorizing the training data instead of learning meaningful patterns. What could be the possible reasons for this issue? Explain any four methods to resolve it.
22. Detail on vanishing and exploding gradient problem.
23. How L1 regularization method leads to weight sparsity.
24. A large feedforward neural network designed for a binary classification task is trained using the sigmoid activation function. However, after the first epoch, the weights in the earlier layers stop updating, even though the network has not yet converged. Explain the underlying reason for this issue.
25. Explain the advantages and limitations of using Gradient Descent for training deep neural networks. How does Gradient Descent with Momentum help in overcoming some of these limitations?
26. Why does increasing the number of hidden layers not always improve a neural network's performance?