**DL theory : Assingments-13**

1. Logistic Regression classifiers are generally preferable to Perceptrons because they can model a wider range of problems, such as those with non-linearly separable classes. Logistic Regression classifiers can be thought of as Perceptrons with a sigmoid activation function, which maps the input to a probability value between 0 and 1. This allows for the modeling of more complex decision boundaries.
2. The logistic activation function was a key ingredient in training the first MLPs because it allows for the modeling of more complex decision boundaries, and it is differentiable, which is necessary for backpropagation.
3. Three popular activation functions are sigmoid, ReLU, and tanh.
4. The shape of the input matrix X would be (m,10) where m is the number of examples. The shape of the hidden layer's weight vector Wh would be (10,50) and the shape of its bias vector bh would be (1,50). The shape of the output layer's weight vector Wo would be (50,3) and its bias vector bo would be (1,3). The shape of the network's output matrix Y would be (m,3).
5. To classify email into spam or ham, you would need one neuron in the output layer, and you would use a sigmoid activation function in the output layer. To classify MNIST images, you would need 10 neurons in the output layer, one for each digit, and you would use a softmax activation function in the output layer.
6. Backpropagation is an algorithm used to train artificial neural networks. It works by calculating the gradient of the loss function with respect to the network's weights and biases, and then updating the weights and biases in the opposite direction of the gradient. Reverse-mode autodiff is a method for efficiently calculating gradients for large neural networks by working backwards through the computation graph.
7. Some of the hyperparameters that can be tweaked in an MLP include the number of hidden layers, the number of neurons in each layer, the learning rate, the batch size, and the regularization parameter. If the MLP overfits the training data, you can try reducing the number of neurons, increasing the regularization parameter, or using techniques such as dropout or early stopping.
8. Training a deep MLP on the MNIST dataset and achieving over 98% precision would be a non-trivial task. It would require a significant amount of hyperparameter tuning and regularization techniques. Some of the items you would need to add to your training pipeline could include: saving checkpoint, restoring the last checkpoint, adding summaries, and plotting learning curves using TensorBoard.