**DL theory : Assingments-16**

1. Activation functions are used in the artificial neural networks to introduce non-linearity into the network. They are applied to the output of each neuron before it is passed to the next layer. a) Sigmoid function compresses all the input values between 0 and 1, it is mostly used in binary classification problems. b) Tanh function is similar to sigmoid but compresses all the input values between -1 and 1. c) ReLU (Rectified Linear Unit) function threshold all the input values at zero, it is used to introduce non-linearity into the network. d) ELU (Exponential Linear Unit) function is similar to ReLU but it tends to produce a more smooth output. e) LeakyReLU function is similar to ReLU but it allows a small, non-zero gradient when the unit is not active. f) Swish function is a self-gated version of ReLU, it is used to introduce non-linearity into the network.
2. The learning rate is a hyperparameter that controls how much the weights of the network are updated during the training process. If the learning rate is high, the network may overshoot the optimal solution and may not converge. If the learning rate is low, the network may take too long to converge. Increasing the learning rate will cause the network to learn faster but may overshoot the optimal solution. Decreasing the learning rate will cause the network to learn slower but may converge more accurately.
3. Increasing the number of internal hidden neurons in a neural network can increase the capacity of the network to learn from the data. However, it can also increase the chances of overfitting, which is when the network performs well on the training data but poorly on new data.
4. Batch computation is a way to process the data in groups or batches, instead of one sample at a time. Increasing the size of the batch computation can speed up the training process, but it can also consume more memory and computational resources.
5. Overfitting occurs when a model is trained too well on the training data and performs poorly on new data. Regularization is a technique that is used to prevent overfitting by adding a penalty term to the cost function. This term discourages the model from assigning too much weight to any one feature.
6. Loss and cost functions are used to measure the performance of a neural network. Loss functions are used to measure the difference between the predicted output and the actual output. Cost functions are used to measure the overall performance of the network.
7. Underfitting occurs when a model is too simple and is unable to capture the underlying patterns in the data. This results in poor performance on both the training and test data.
8. Dropout is a regularization technique that is used to prevent overfitting by randomly dropping out neurons during the training process. This makes the network more robust and less reliant on any one neuron. This forces the network to learn multiple independent representations of the data which makes the model more generalizable.