DL ASSIGNMENT 16

Q1) sigmoid is of the form 1/(1 + negative exponential). It ranges from o to 1. It is smooth. Tanh ranges from -1 to 1. It is also smooth. Relu is negligibly constant upto a threshold input and then linear after threshold. It is based on threshold where sharp change takes place. Elu is similar to relu but smoother till threshold and is even negative till threshold. Leaky relu is also like relu but negative before threshold with some linear increase. Swish function is similar to relu but smooth. It has slight smooth dip just before threshold for smoothness.

Q2) increasing results in faster learning but can also result in overfitting or exploding gradient. Decreasing learning rate could result in undercutting or vanishing gradient problem.

Q3) learning and accuracy could improve depending on use case. However, number of parameters increases increasing computational difficulty and slowing down training. It may also result in vanishing gradients with more back propagation.

Q4) regularisation may decrease. Computation and hardware hurdles may come to due increased computations with increased batch size.

Q5) we adopt regularisation so that our model actually understands relationships and this understanding is not limited to just training set. Then we would not have highly varying accuracies with different datasets.

Q6) the functions calculate the loss between prediction and actual values via various formulas to enable correction and updation in training.

Q7) undercutting means model is unable to understand relationships in data resulting in huge losses and poor accuracy.

Q8) we use dropout to regularise and to strengthen each neuron in its understanding.