RQ#3

Q1.

Forward propagation is the essential flow of the NN, while back propagation is important as it finetunes the weights and biases so that NN can produce better predictions

Q2.

Yes.

Q3.

In backprop, stochastic gradient descent is used instead of gradient descent to nudge the weights and biases, because gradient descent requires the nudging to be done on all training data, which would cost too much time, while stochastic gradient descent first divides the training data into batches, then does the nudging on only one batch at each layer.

Q4.

They are differentiable.

Q5.

ReLU works faster but only on values greater than zero. Sigmoid outputs a probability distributions but causes vanishing gradients.