**Question 1:**

**1. Name 3 assumptions required for linear regression.**

**Linear Regression:**

It is a helpful statistical strategy we can use to comprehend the connection between two factors, x, and y. The primary base to see of Linear Regression is that there is a straight connection between the independent variable, x, and the independent variable y. In any case, before we conduct linear regression, we should initially ensure that three assumptions are met:

* Linear relationship: There exists a direct connection between the independent variable, x, and the dependent variable, y.
* Independence: The residuals are independent. Specifically, there is no connection between successive residuals in time series data.
* Normality: The residuals of the model are normally distributed.
* **Linear Regression Classification:**

The simplest method to distinguish if this supposition is met is to make a scatter plot of x versus y. This permits you to outwardly check whether there is a direct connection between the two factors. On the off chance that it would seem that the focuses in the plot could fall along a straight line, at that point there exists some kind of direct connection between the two factors and this classifies that it’s a Linear Regression. In case of not meeting the criteria, we can apply a nonlinear transformation to the independent and/or dependent variable. Methods like square root, taking the log. Moreover, we can add another independent variable to the model.

* **Independence:**

In this assumption of Linear Regression, the residuals are independent. This is generally important when working with time-series data. Preferably, we don't need there to be an example among continuous residuals. For instance, residuals shouldn't consistently become bigger over a long period. The easiest method to test if this supposition is met is to take a see at a residual time series plot, which is a plot of residuals versus time. Preferably, a large portion of the residual autocorrelations should fall inside the 95% certainty groups around zero, which are situated at about +/ - 2-over the square base of n, where n is the example size.

* **Normality:**

In Normality the residuals are normally distributed. To check for this assumption, we create a visual Q-Q plot. A Q-Q plot is a plot that represents the structure of residuals showing its pattern of a normal distribution or not. If points follow a rough straight diagonal line, then the normality assumption is met. To avoid violation of Normality, it should be seen that outliers are worked upon.

**Reference:**

[1] Zach, The Four Assumptions of Linear Regression, Retrieved from https://www.statology.org/linear-regression-assumptions/

**2. Name 3 activation functions used in neural networks.**

**Activation Functions:**

It’s just a function that we use to get the output of the node. It is also known as Transfer Function. It is used to determine the output of a neural network like yes or no. It maps the resulting values in between 0 to 1 or -1 to 1 etc.

**Linear Activation Function:**

It takes the inputs, multiplied by the weights for each neuron, and creates an output signal proportional to the input. In one sense, a linear function is better than a step function because it allows multiple outputs, not just yes and no. However, a linear activation function has two major problems:

* Not possible to use
* All layers of the neural network collapse into one

So, a linear activation function turns the neural network into just one layer. A neural network with a linear activation function is simply a linear regression model

**Sigmoid or Logistic Activation Function:**

The Sigmoid Function curve looks like an S-shape. The main reason why we use the sigmoid function is that it exists between (0 to 1). Therefore, it is especially used for models where we have to predict the probability as an output. Since the probability of anything exists only between the range of 0 and 1, sigmoid is the right choice. The function is differentiable. That means, we can find the slope of the sigmoid curve at any two points.

**ReLU (Rectified Linear Unit) Activation Function:**

The ReLU is the most used activation function in the world right now. Since it is used in almost all the convolutional neural networks or deep learning. Range: [ 0 to infinity). The function and its derivative both are monotonic. But the issue is that all the negative values become zero immediately which decreases the ability of the model to fit or train from the data properly. That means any negative input given to the ReLU activation function turns the value into zero immediately in the graph, which in turn affects the resulting graph by not mapping the negative values appropriately

**References:**

[1] Avinash Sharma V (Mar 30, 2017), Understanding Activation Functions in Neural Networks was retrieved from <https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0>

[2] Jason Brownlee on (Jan 9, 2019), A Gentle Introduction to the Rectified Linear Unit (ReLU) was retrieved from <https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/>